



Article A New Class of Weighted CUSUM Statistics

Xiaoping Shi^{1,*}, Xiang-Sheng Wang² and Nancy Reid³

- ¹ Department of Computer Science, Mathematics, Physics and Statistics, University of British Columbia, Kelowna, BC V1V 1V7, Canada
- ² Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70503, USA
- ³ Department of Statistical Sciences, University of Toronto, Toronto, ON M5S 3G3, Canada
- * Correspondence: xiaoping.shi@ubc.ca

Abstract: A change point is a location or time at which observations or data obey two different models: before and after. In real problems, we may know some prior information about the location of the change point, say at the right or left tail of the sequence. How does one incorporate the prior information into the current cumulative sum (CUSUM) statistics? We propose a new class of weighted CUSUM statistics with three different types of quadratic weights accounting for different prior positions of the change points. One interpretation of the weights is the mean duration in a random walk. Under the normal model with known variance, the exact distributions of these statistics are explicitly expressed in terms of eigenvalues. Theoretical results about the explicit difference of the distributions are valuable. The expansions of asymptotic distributions are compared with the expansion of the limit distributions of the Cramér-von Mises statistic and the Anderson and Darling statistic. We provide some extensions from independent normal responses to more interesting models, such as graphical models, the mixture of normals, Poisson, and weakly dependent models. Simulations suggest that the proposed test statistics have better power than the graph-based statistics. We illustrate their application to a detection problem with video data.

Keywords: asymptotic distribution; Brownian bridge; exact distribution; quadratic weights; weak dependence

1. Introduction

A change point is a location or time at which observations or data obey two different models: before and after. Detecting change points is a nontrivial problem and has been studied by many authors; see a book treatment in [1] and recent advances in CUSUM-based change point tests [2–4]. In real problems, we may know some prior information about the location of the change point, say at the right or left tail of the sequence. How does one incorporate prior information into current CUSUM-based statistics? We consider a new class of weighted CUSUM statistics for a simple model and provide some extensions to more complicated models.

Given a series of univariate random variables Y_1, \ldots, Y_n , we consider the problem of testing whether there is a change in the mean of their distribution. The test statistic we use is:

$$S_n(Y;\tau,\gamma) = \sum_{k=1}^{n-1} w_k^{-1}(\tau) \left\{ \sum_{i=1}^k (Y_i - \bar{Y}) \right\}^{\gamma},$$
(1)

where $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$, $\overline{Y} = n^{-1} \sum_{j=1}^n Y_j$, $\gamma > 0$, and

$$w_k(\tau) = -(k-\tau)^2 + \max\{\tau^2, (n-\tau)^2\} = \begin{cases} (n+k)(n-k), \text{ if } \tau = 0, \\ k(n-k), \text{ if } \tau = n/2, \\ k(2n-k), \text{ if } \tau = n, \end{cases}$$
(2)



Citation: Shi, X.; Wang, X.-S.; Reid, N. A New Class of Weighted CUSUM Statistics. *Entropy* **2022**, *24*, 1652. https://doi.org/10.3390/e24111652

Academic Editor: Carlos Alberto De Bragança Pereira

Received: 20 September 2022 Accepted: 11 November 2022 Published: 14 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). where $\tau = 0, n/2$, and *n* account for three different prior positions of the change point, respectively. We call S_n a weighted CUSUM (WC) statistic.

Inspired by the change point literature, we consider these types of quadratic weights. The term $\max\{\tau^2, (n-\tau)^2\} = \max_{0 \le j \le n} (j-\tau)^2$ is introduced to ensure that the weight $w_k(\tau)$ is positive for any 0 < k < n. Usually, we choose $\gamma = 2$ to capture the change in the mean. When $\tau = n/2$, the weight $w_k(n/2) = k(n-k)$ corresponds to the likelihood ratio test; see Csörgő and Horváth [1] and a related review in Jandhyala et al. [5]. If prior information indicates that the change point more likely occurs in the right or left tail of the sequence, we can set the weight $w_k(0) = (n+k)(n-k)$ (left drifted to the symmetry center point 0) or $w_k(n) = k(2n-k)$ (right drifted to the symmetry center point *n*) to improve the power of the test.

One interpretation of the weights is the mean duration in a random walk $\{X_i, i \ge 0\}$ on N + 1 states, $\{0, 1, ..., N\}$, whose transition probability is given by $P(X_{i+1} = k \pm 1 | X_i = k) = 1/2$ for k = 1, ..., N - 1, $P(X_{i+1} = 0 | X_i = 0) = 1$, and $P(X_{i+1} = N | X_i = N) = 1$. Let T denote the random time at which the process first reaches 0 or N. Then, for k = 1, ..., n - 1, $E(T|X_0 = k) = k(n-k) = w_k(n/2)$ if N = n; $E(T|X_0 = k) = k(2n-k) = w_k(n)$ if N = 2n; and $E(T|X_0 = n-k) = (n+k)(n-k) = w_k(0)$ if N = 2n. Figure 1 depicts four vectors w_k for n = 10. The centers of symmetry of these quadratic weights are at different positions.



Figure 1. Plot of weights: n^2 (uniform), k(n - k) (centered), (n + k)(n - k) (left shifted), and k(2n - k) (right shifted).

The weights in (1) can be thought of as an inverse prior probability on the change point, giving S_n a Bayesian flavor, as in Gardner [6], who used the uniform prior n^{-2} , or Perron [7], who devised a unit-root test for time series. From a frequentist perspective, the weighted sum statistic offers an alternative to the maximum statistic most commonly used Csörgő and Horváth [1], which we show (in small simulations omitted here) has higher power, especially when the change point is at the center of the sequence for any τ , in the right tail of the sequence for $\tau = n$, and in the left tail of the sequence for $\tau = 0$.

For these types of quadratic weights, a couple of questions naturally arise: will different weights lead to different distributions of WC in Equation (1)? If so, how significant will the differences in the distribution be? If two different weights lead to the same distribution, are there any intrinsic reasons? Although one can estimate the distribution of WC by simulation, theoretical results about the explicit differences of the distributions are valuable. Moreover, simulations and computations of eigenvalues for large n are computationally expensive. To answer the aforementioned questions, we shall study the distribution of the WC theoretically; we derive Karhunen–Loève expansions of the exact

and asymptotic distributions of the WC statistics. The calculation of a Karhunen–Loève expansion is a nontrivial task, even under the normal model. Gardner [6] discussed the uniform weight under the normal assumption, but the quadratic weights we consider here increase the difficulty substantially. We present below a unified theory that enables us to establish the distribution of WC using dual Hahn polynomials. The asymptotic distributions for the quadratic weights $w_k(0)$ and $w_k(n)$ are identical, and the expansions of asymptotic distributions between $w_k(0)$ and another quadratic weight $w_k(n/2)$ differ by an odd number of terms. We make a comparison with the expansion of the limit distributions of the Cramér-von Mises statistic and the Anderson and Darling statistics; see also MacNeill [8].

The WC has some variants in other models. For example, in the graphical model, γ can be 1 if we replace Y with a count of edges. Here, the main challenge is to approximate the covariance of edge-count statistics under the null permutations. In the normal mixed model, a variant of WC can be derived by considering a marginal likelihood function. In the Poisson mixed model, however, the calculation of the marginal likelihood function is hindered by an integral without a closed form. To approximate this integral, one may use Laplace, or saddle point approximation [9–13]. Here, we apply the saddle point approximation to the integral and provide a variant of WC related to the log link. For the classical change point Poisson model without latent variables, see [1] (p. 27); for the Poisson process with a change point, we refer readers to Akman and Raftery [14], Loader [15]. Moreover, to adopt the assumption of weak dependence in practice, we avoid the estimation of the variance and provide a variance of WC.

The structure of the paper is outlined as follows. In Section 2, we derive the explicit expansions of the distribution of the WC statistics and explore their connections with the Karhunen–Loève expansion. We derive extended versions of WC by considering the observations as nodes in the graphical model and allowing the observations from a normal or Poisson mixed model to be weakly dependent. In Section 3, we discuss the power of the proposed WC test. In Section 4, we use simulation to compare the performance of this test with that of a graph-based test statistic. In Section 5, we present an application for video data. In Section 6, we discuss the extension to multiple change points and suggest future work on other quadratic weights.

2. Exact and Asymptotic Distributions of the WC Statistics

2.1. Explicit Distribution for a Normal Model

We assume here that $\{Y_i\}$ are independent following a normal distribution with a common known variance σ^2 . The case of unknown σ^2 is addressed in Remark 3, and an extension relaxing the independence assumption is given in Section 2.6.

Following the derivation in Gardner [6], we write (1) as a quadratic form

$$S_n(\mathbf{Y};\tau,2) = \frac{1}{n^2} \sum_{k=1}^{n-1} p_k \left\{ \sum_{i=1}^k (n-k) Y_i - \sum_{i=k+1}^n k Y_i \right\}^2 = \frac{1}{n^2} \mathbf{Y}^\top A A^\top \mathbf{Y} = \mathbf{Y}^\top Q \mathbf{Y}, \quad (3)$$

where $p_k = p_k(\tau) = w_k^{-1}(\tau)$, and $n^2 Q = AA^{\top}$ with $A = (A_1, \dots, A_{n-1})$. Here, $A_k = p_k^{1/2}(n-k, \dots, n-k, -k, \dots, -k)^{\top}$ such that the first *k* entries of A_k are $p_k^{1/2}(n-k)$ and the last n-k entries $-p_k^{1/2}k$.

By using the recurrence identity and the dual Hahn polynomial, we obtain a new exact result in terms of the eigenvalues of Q in (3).

Theorem 1. Assume that $\{Y_i\}$ are independent normally distributed random variables with a common mean and known variance σ^2 . The exact distribution of $S_n(Y;\tau,2)$ is

$$\frac{S_n(\boldsymbol{Y};\tau,2)}{\sigma^2} =_d \sum_{k=1}^n \lambda_k(\tau) Z_k^2, \tag{4}$$

where Z_k^2 are independent and identically distributed normal random variables with mean zero and variance 1, $\lambda_n(\tau) = 0$, and

$$\lambda_k(\tau) = \begin{cases} \frac{1}{k(k+1)}, \ k = 1, \dots, n-1, \text{ if } \tau = n/2\\ \frac{1}{2k(2k+1)}, \ k = 1, \dots, n-1, \text{ if } \tau = 0 \text{ or } n \end{cases}$$

The proof of Theorem 1 is given in Appendix A. We make the following remarks.

Remark 1. It is interesting that $\lambda_{2k}(n/2) = \lambda_k(0)$ for all 0 < k < n/2; namely, the eigenvalues for $w_k(n/2)$ with even indices coincide with the eigenvalues for $w_k(0)$ with indices less than n/2. As the sample size increases from n to n + 1, the n - 1 nonzero eigenvalues are retained and the added nonzero eigenvalue must be $1/\{n(n + 1)\}$ for $w_k(n/2)$ or $1/\{2n(2n + 1)\}$ for $w_k(0)$ or $w_k(n)$. This interesting phenomenon has not been seen in the uniform weights of Gardner [6]. As far as we know, this recursive property of the eigenvalues for the non-uniform weights is new. Figure 2 depicts the pattern of eigenvalues (cross products of rows and columns) illustrated by dots for three weights $w_k(n/2)$ (blue), $w_k(0)$ (green), and $w_k(n)$ (purple) with the increase of n.



Figure 2. The pattern of eigenvalues (cross products of rows and columns) illustrated by dots for three weights $w_k(n/2)$ for $\tau = n/2$ (blue), $w_k(0)$ for $\tau = 0$ (green) and $w_k(n)$ for $\tau = n$ (purple) with the increase of n.

Remark 2. The distribution in (4) can be calculated numerically using Imhof's method [16] or simulated by a Monte Carlo method, but accurate analytical approximations are potentially faster and more stable. A saddle point approximation to the distribution of quadratic forms in normal variates was studied in Kuonen [17], building on Daniels [9,18] and Lugannani and Rice [19].

Remark 3. When the variance σ^2 is unknown, we can replace σ^2 with a consistent estimator

$$\hat{\sigma}^2 = (n-1)^{-1} \sum_{i=1}^n (Y_i - \overline{Y})^2,$$

by using Slutsky's lemma. This also holds in Corollary 1. For dependent data, one issue is to give a valid estimate of the variance; see Section 2.6.

2.2. Karhunen–Loève Expansion

The squared integral of a Brownian bridge arises in the study of tests for goodness-offit. Given a sample of independent and identically distributed random variables with an empirical distribution function $F_n(x)$, the statistic

$$\omega_n^2(\psi) = n \int_{-\infty}^{\infty} \{F_n(t) - F(t)\}^2 \psi\{F(t)\} dF(t)$$

provides a test of the null hypothesis that the observations come from the distribution $F(\cdot)$. The Cramér-von Mises statistic has $\psi(t) \equiv 1$, and the Anderson-Darling statistic has $\psi(t) = 1/\{t(1-t)\}$. Here, we shall discuss two new weights: $\psi(t) = 1/\{t(2-t)\}$ and $\psi(t) = 1/(1-t^2)$.

MacNeill [8] showed that

$$\int_0^1 \{\mathcal{B}(t) - t\mathcal{B}(1)\}^2 dt = \sum_{k=1}^\infty \frac{1}{k^2 \pi^2} Z_k^2,$$

using a Fourier expansion of $\mathcal{B}(t) - t\mathcal{B}(1) = \sum_{k=1}^{\infty} \sqrt{2} \sin(k\pi t) / (k\pi) Z_k$, where $\{\sqrt{2} \sin(k\pi t), k = 1, 2, \dots, \infty\}$ is an orthonormal basis in $L^2(0, 1)$ and $\mathcal{B}(t)$ is a standard Brownian motion and $\mathcal{B}(t) - t\mathcal{B}(1)$ is a Brownian bridge.

Anderson and Darling [20] showed that

$$\int_0^1 \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^2}{t(1-t)} dt = \sum_{k=1}^\infty \frac{1}{k(k+1)} Z_k^2.$$

In Appendix B, we use Jacobi polynomials to derive the Karhunen–Loève expansion for the integrals of the weighted square of the Brownian bridge with two new weights $\psi(t) = 1/\{t(2-t)\}$ and $\psi(t) = 1/(1-t^2)$. The results are stated in the following theorem.

Theorem 2. The two weights $\psi(t) = 1/\{t(2-t)\}$ and $\psi(t) = 1/(1-t^2)$ lead to the same Karhunen–Loève expansions:

$$\int_{0}^{1} \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^{2}}{2t - t^{2}} dt = \sum_{k=1}^{\infty} \frac{1}{2k(2k+1)} Z_{k}^{2}$$
(5)

and

$$\int_0^1 \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^2}{1 - t^2} dt = \sum_{k=1}^\infty \frac{1}{2k(2k+1)} Z_k^2.$$
 (6)

The proof of the above two equalities will be provided in Appendix B. One can see the equivalence of these two equalities by using a change of variable.

Given different probabilities (*p*), Table 1 presents the critical values c_p for which $p = P(\chi_n^2(\tau) \le c_p)$ for different *n*, where $\chi_n^2(\tau) = \sum_{k=1}^n \lambda_k(\tau) Z_k^2$ and calculations of critical values for finite *n* are based on Imhof's method [16] implemented in R package CompQuadForm [21]. A few critical values are tabulated in Anderson and Darling [22] for $\chi_n^2(\tau)$ with $n = \infty$. One can see the critical values converge very quickly as *n* increases to ∞ .

						n					
Weight	р	20	40	60	80	100	200	400	1000	10,000	∞
$w_k(n/2)$	0.90	1.883	1.908	1.916	1.920	1.923	1.928	1.930	1.932	1.933	1.933
	0.925	2.111	2.136	2.145	2.149	2.151	2.156	2.159	2.160	2.161	
	0.95	2.442	2.467	2.476	2.480	2.482	2.487	2.490	2.491	2.492	2.492
	0.975	3.027	3.052	3.061	3.065	3.067	3.072	3.075	3.076	3.077	3.070
	0.99	3.828	3.853	3.861	3.866	3.868	3.873	3.876	3.877	3.878	3.850
$w_k(0)$	0.90	0.599	0.605	0.607	0.608	0.609	0.610	0.611	0.611	0.611	
	0.925	0.675	0.682	0.684	0.685	0.685	0.687	0.687	0.688	0.688	
	0.95	0.786	0.792	0.794	0.795	0.796	0.797	0.798	0.798	0.798	
	0.975	0.981	0.988	0.990	0.991	0.991	0.993	0.993	0.994	0.994	
	0.99	1.249	1.255	1.257	1.258	1.259	1.260	1.261	1.261	1.261	

Table 1. Critical values of $\sum_{k=1}^{n} \lambda_k(\tau) Z_k^2$ for different weights in (4), sizes (*n*), and probabilities (*p*).

In fact, we can connect the limit distribution of WC statistic and its functional limit distribution by the Karhunen–Loève expansion of the integral of the weighted square of Brownian bridge in terms of the Jacobi polynomials. Theorem 1 immediately implies the following asymptotic distribution as $n \rightarrow \infty$.

Corollary 1. Under the assumptions of Theorem 1, when $n \to \infty$,

$$\frac{S_n(\boldsymbol{Y};\tau,2)}{\sigma^2} \to_d \sum_{k=1}^{\infty} \lambda_k(\tau) Z_k^2$$

One can check $\sum_{k=n}^{\infty} \lambda_k(\tau) Z_k^2 \to_p 0$ by Markov's inequality. Hence, $\sum_{k=1}^{n-1} \lambda_k(\tau) Z_k^2$ converges to $\sum_{k=1}^{\infty} \lambda_k(\tau) Z_k^2$ in probability as $n \to \infty$. By the functional limit theorem,

$$\frac{S_n(\boldsymbol{Y};\tau,2)}{\sigma^2} \to_d \begin{cases} \int_0^1 \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^2}{1-t^2} dt, \text{ if } \tau = 0, \\ \int_0^1 \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^2}{t(1-t)} dt, \text{ if } \tau = n/2, \\ \int_0^1 \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^2}{2t-t^2} dt, \text{ if } \tau = n. \end{cases}$$

2.3. Graphical Model

Assume the $\{Y_{i,j}, 1 \le i \le n, 1 \le j \le q\}$ are independent and have common mean $E(Y_{i,j}) = \mu_i$ and variance $Var(Y_{i,j}) = \sigma_i^2$. Consider testing

$$H_{0}: \mu_{i} \equiv \mu, \sigma_{i}^{2} \equiv \sigma^{2} vs H_{a}: \mu_{i} = \begin{cases} \mu_{-}, \text{ for } i \leq k^{*}, \\ \mu_{+}, \text{ for } k^{*} > i, \end{cases} \text{ or } \sigma_{i}^{2} = \begin{cases} \sigma_{-}^{2}, \text{ for } i \leq k^{*}, \\ \sigma_{+}^{2}, \text{ for } k^{*} > i, \end{cases}$$
(7)

where $\mu_{-} \neq \mu_{+}$ or $\sigma_{-}^{2} \neq \sigma_{+}^{2}$, the parameters $\mu, \mu_{-}, \mu_{+}, \sigma^{2}, \sigma_{-}^{2}$, and σ_{+}^{2} are unknown.

A graphical model can be established by treating each *q*-dimensional vector as a node and assigning the Euclidean distance between any two vectors. Here, we consider a path \mathcal{P} with an ordering of nodes (v_1, \ldots, v_n) and edges (v_i, v_{i+1}) for $i = 1, \ldots, n-1$. Associated with the path, the count of edges that connect nodes between arbitrary two disjoint sets $\mathbb{N}_k = \{1, \ldots, k\}$ and $\overline{\mathbb{N}}_k = \{k + 1, \ldots, n\}$ is defined to be:

$$C_{\mathcal{P}}(\mathbb{N}_k, \overline{\mathbb{N}}_k) = \sum_{i=1}^{n-1} I\Big[\Big\{(v_i \in \mathbb{N}_k) \cap (v_{i+1} \in \overline{\mathbb{N}}_k)\} \cup \{(v_{i+1} \in \mathbb{N}_k) \cap (v_i \in \overline{\mathbb{N}}_k)\Big\}\Big], \quad (8)$$

where $I(\cdot)$ is an indicator function that takes 1 if true otherwise 0. The $C_{\mathcal{P}}(\mathbb{N}_k, \overline{\mathbb{N}}_k)$ counts edges between two groups \mathbb{N}_k and $\overline{\mathbb{N}}_k$.

Denote the expectation and variance of $C_{\mathcal{P}}(\mathbb{N}_k, \overline{\mathbb{N}}_k)$ under *n*! permutations of nodes as $E_{\text{perm}}C_{\mathcal{P}}(\mathbb{N}_k, \overline{\mathbb{N}}_k)$ and $\text{Var}_{\text{perm}}C_{\mathcal{P}}(\mathbb{N}_k, \overline{\mathbb{N}}_k)$. By [23],

$$E_{\text{perm}}C_{\mathcal{P}}(\mathbb{N}_k,\overline{\mathbb{N}}_k) = \frac{2k(n-k)}{n} \text{ and } \operatorname{Var}_{\text{perm}}C_{\mathcal{P}}(\mathbb{N}_k,\overline{\mathbb{N}}_k) = \frac{2k(n-k)\{2k(n-k)-n\}}{n^3-n^2}.$$

A WC statistic may be constructed as

$$S_n(\mathcal{P};\tau,\gamma) = \sum_{k=1}^{n-1} w_k^{-1}(\tau) \left\{ -C_{\mathcal{P}}(\mathbb{N}_k,\overline{\mathbb{N}}_k) + \frac{2k(n-k)}{n} \right\}^{\gamma}.$$
(9)

A large value of observed $S_n(\mathcal{P}^*; \tau, \gamma)$ based on the shortest Hamiltonian path (SHP), \mathcal{P}^* , indicates a rejection of the null hypothesis, i.e., there is a change point; see the heuristic algorithm of SHP in Biswas et al. [24] and the analysis of power and change point in Shi, Wu and Rao [25,26] for $\gamma = 2$ and $w_k(\tau) = \operatorname{Var}_{\text{perm}} C_{\mathcal{P}}(\mathbb{N}_k, \overline{\mathbb{N}}_k)$. Here, we will establish the asymptotic distribution of $S_n(\mathcal{P}; \tau, \gamma)$ for $\gamma = 1, 2$. First, we give the following Lemma.

Lemma 1. For $k = \lfloor tn \rfloor$ with 0 < t < 1,

$$\frac{1}{\sqrt{2n}} \left\{ -C_{\mathcal{P}}(\mathbb{N}_k, \overline{\mathbb{N}}_k) + \frac{2k(n-k)}{n} \right\} \to_d \{\mathcal{B}(t) - t\mathcal{B}(1)\}^2 - t(1-t), n \to \infty.$$
(10)

By the functional limit theorem,

$$\sqrt{\frac{n}{2}}S_{n}(\mathcal{P};\tau,1) \rightarrow_{d} \begin{cases} \int_{0}^{1} \frac{\{\mathcal{B}(t)-t\mathcal{B}(1)\}^{2}}{1-t^{2}}dt + \log(2) - 1, \text{ if } \tau = 0, \\ \int_{0}^{1} \frac{\{\mathcal{B}(t)-t\mathcal{B}(1)\}^{2}}{t(1-t)}dt - 1, \text{ if } \tau = n/2, \\ \int_{0}^{1} \frac{\{\mathcal{B}(t)-t\mathcal{B}(1)\}^{2}}{2t-t^{2}}dt + \log(2) - 1, \text{ if } \tau = n, \end{cases}$$
(11)

and

$$\frac{1}{2}S_{n}(\mathcal{P};\tau,2) \rightarrow_{d} \begin{cases} \int_{0}^{1} \frac{\left[\{\mathcal{B}(t)-t\mathcal{B}(1)\}^{2}-t(1-t)\right]^{2}}{1-t^{2}} dt, \text{ if } \tau = 0, \\ \int_{0}^{1} \frac{\left[\{\mathcal{B}(t)-t\mathcal{B}(1)\}^{2}-t(1-t)\right]^{2}}{t(1-t)} dt, \text{ if } \tau = n/2, \\ \int_{0}^{1} \frac{\left[\{\mathcal{B}(t)-t\mathcal{B}(1)\}^{2}-t(1-t)\right]^{2}}{2t-t^{2}} dt, \text{ if } \tau = n, \end{cases}$$
(12)

which solves an open problem in [25,26]. Different values of γ lead to different rates of convergence and different "normings".

2.4. Normal Mixed Model

Assume $Y_{i,j} = \mu_i + U_i + e_{i,j}$, where $1 \le i \le n$, $1 \le j \le q$, $e_{i,j}$ are independent and identically normally distributed with mean zero and variance σ^2 , and U_i are independent latent variables following a normal distribution with mean zero and variance ν^2 .

Consider testing

$$H_{0}: \mu_{i} \equiv \mu \quad vs \quad H_{a}: \mu_{i} = \begin{cases} \mu_{-}, \text{ for } i \leq k^{*}, \\ \mu_{+}, \text{ for } k^{*} > i, \end{cases}$$
(13)

where $\mu_{-} \neq \mu_{+}$, the parameters μ , μ_{-} and μ_{+} are unknown, and we tentatively assume the time k^* , called the change point, and the variances σ^2 and ν^2 to be known.

The marginal log-likelihood function of μ under H_0 is

$$\ell(\mu) = \ell_0 - \frac{\sum_{i=1}^{n} (\overline{Y}_{i.} - \mu)^2}{2\nu^2 + 2\sigma^2/q},$$

where ℓ_0 does not depend on μ and $\overline{Y}_{i} = \sum_{i=1}^{q} Y_{i,i}/q$.

Therefore,

$$\max_{\mu} \ell(\mu) = \ell_0 - \sum_{i=1}^n \frac{(\overline{Y}_{i.} - \hat{\mu}_{1,n})^2}{2\nu^2 + 2\sigma^2/q},$$

where $\hat{\mu}_{t_1,t_2} = \sum_{i=t_1}^{t_2} \overline{Y}_{i} / (t_2 - t_1 + 1).$

In a similar way, the marginal log-likelihood function of μ_{-} and μ_{+} under H_a can be obtained. Then, the marginal log-likelihood ratio is

$$\sum_{i=k^*+1}^{n} \frac{(\overline{Y}_{i\bullet} - \hat{\mu}_{1,n})^2}{\nu^2 + \sigma^2/q} - \sum_{i=1}^{k^*} \frac{(\overline{Y}_{i\bullet} - \hat{\mu}_{1,k^*})^2}{\nu^2 + \sigma^2/q} - \sum_{i=k^*}^{n} \frac{(\overline{Y}_{i\bullet} - \hat{\mu}_{k^*+1,n})^2}{\nu^2 + \sigma^2/q},$$

which is equal to

$$\frac{n\left\{\sum_{i=1}^{k^*} (\overline{Y}_{i.} - \hat{\mu}_{1,n})\right\}^2}{k^*(n-k^*)(\nu^2 + \sigma^2/q)}.$$

As the change point k^* could be unknown in practice, we may sum over $k^* = 1, ..., n - 1$ and consider the average value, which leads to

$$S_n(\overline{Y}_{,i}, n/2, 2) = \sum_{k=1}^{n-1} w_k^{-1}(n/2) \left\{ \sum_{i=1}^k (\overline{Y}_{i, -} - \hat{\mu}_{1, n}) \right\}^2.$$
(14)

where $\overline{\mathbf{Y}}_{\bullet} = (\overline{\mathbf{Y}}_{1,\bullet}, \ldots, \overline{\mathbf{Y}}_{n,\bullet})^{\top}$.

By Theorem 1 and Remark (3) in terms of weighted version for any τ , as $n \to \infty$,

$$\frac{S_n(\overline{\mathbf{Y}}_{\boldsymbol{\cdot}};\tau,2)}{(n-1)^{-1}\left\{\sum_{i=1}^n(\overline{\mathbf{Y}}_{i\boldsymbol{\cdot}}-\hat{\mu}_{1,n})\right\}^2} \to_d \sum_{k=1}^\infty \lambda_k(\tau) Z_k^2$$

2.5. Poisson Mixed Model

Assume $Y_{i,j}$ follows a Poisson distribution with conditional mean $E(Y_{i,j}|U_i) = \exp(\rho_i + U_i)$. Consider testing

$$H_0: \rho_i \equiv \rho \quad vs \quad H_a: \rho_i = \begin{cases} \rho_-, \text{ for } 1 \le i \le k^*, \\ \rho_+, \text{ for } k^* < i \le n, \end{cases}$$
(15)

where $\rho_{-} \neq \rho_{+}$, the parameters ρ , ρ_{-} and ρ_{+} are unknown. Under normal distribution for U_i , the likelihood ratio contains an integral. With the focus on the simple Poisson mixed model without a change point, Hall et al. [27,28] applied the Gaussian variational approximation (GVA) to approximate the integral so as to avoid solving the integral. We provide a saddle point approximation here.

The marginal log-likelihood function of ρ under H_0 is

$$\ell(\rho) = \ell_1 + \sum_{i=1}^n \log \mathcal{I}_i(\rho), \tag{16}$$

where ℓ_1 does not depend on r and $\mathcal{I}_i(\rho) = \int_{-\infty}^{\infty} \exp\left\{-qe^{\rho+u} + q\overline{Y}_i \cdot (\rho+u) - \frac{u^2}{2v^2}\right\} du.$

The calculation of $\ell(\rho)$ is hindered by the lack of a closed form of the integral $\mathcal{I}_i(\rho)$. Here, we apply the saddle point approximation to the integral as shown in Lemma 2.

Lemma 2. For the integral
$$\mathcal{I}(\rho; a, b, v^2) = \int_{-\infty}^{\infty} \exp\left\{-be^u + au - \frac{(u-\rho)^2}{2v^2}\right\} du$$
$$\mathcal{I}(\rho; a, b, v^2) \approx \left(\frac{a}{be}\right)^a \sqrt{\frac{2\pi}{a}} e^{-\frac{(c-\rho)^2}{2v^2}},$$

where the symbol \approx means asymptotic equivalence and the saddle point c solves $\phi'(u) = 0$ with $\phi(u) = au - be^u$, i.e., $c = \log(a/b)$.

In (16), $\mathcal{I}_i(\rho) = \mathcal{I}(\rho; q\overline{Y}_{i,\cdot}, q, \nu^2)$, so Lemma 2 gives the leading term as

$$\ell(\rho) \approx \ell_1 - \sum_{i=1}^n \frac{(\log \bar{Y}_{i \star} - \rho)^2}{2\nu^2},$$

and the leading term approximation to $\max_{\rho} \ell(\rho)$

$$\ell_1 - \sum_{i=1}^n \frac{(\log \bar{Y}_{i \star} - \hat{\rho}_{1,n})^2}{2\nu^2},$$

where $\hat{\rho}_{t_1,t_2} = \sum_{i=t_1}^{t_2} \log \bar{Y}_{i} / (t_2 - t_1 + 1).$

In a similar way, $\max_{\rho_1,\rho_2} \ell(\rho_1,\rho_2)$ under H_a can be approximated, giving the approximate log-likelihood ratio

$$\sum_{i=1}^{n} \frac{(\log \bar{Y}_{i.} - \hat{\rho}_{1,n})^2}{\nu^2} - \sum_{i=1}^{k^*} \frac{(\log \bar{Y}_{i.} - \hat{\rho}_{1,k^*})^2}{\nu^2} - \sum_{i=k^*+1}^{n} \frac{(\log \bar{Y}_{i.} - \hat{\rho}_{k^*+1,n})^2}{\nu^2} = \frac{n\{\sum_{i=1}^{k^*} (\log \bar{Y}_{i.} - \hat{\rho}_{1,n})\}^2}{k^*(n-k^*)\nu^2}.$$
(17)

Considering that the change point k^* is unknown, we may sum (17) over $k^* = 1, ..., n - 1$ as shown in (1) and consider the average value,

$$S_n(\log \overline{Y}_{,i}; n/2, 2) = \sum_{k=1}^{n-1} w_k^{-1}(n/2) \left\{ \sum_{i=1}^k (\log \overline{Y}_{i, \cdot} - \hat{\rho}_{1, n}) \right\}^2.$$
(18)

Note that the term $w_k(n/2)$ is derived from the approximate likelihood ratio statistic, different from the classical Poisson change point statistic in Csörgő and Horváth [1] (p. 27). By Theorem 1 and Remark 3 in terms of weighted version for any τ , as $n, q \to \infty$,

$$\frac{S_n(\log \overline{Y}_{\boldsymbol{\cdot}};\tau,2)}{(n-1)^{-1}\left\{\sum_{i=1}^n(\log \overline{Y}_{i\boldsymbol{\cdot}}-\hat{\rho}_{1,n})\right\}^2} \to_d \sum_{k=1}^\infty \lambda_k(\tau) Z_k^2$$

2.6. Weak Dependence

Now, we consider a space-time model for the distribution of $Y_{i,j}$, where *i* indexes time and *j* indexes space. First, we assume some weak dependence conditions on space by supposing the central limit theorem holds:

$$\frac{1}{\sqrt{q}} \sum_{j=1}^{q} (Y_{i,j} - \overline{Y}_{i,\cdot}) \to_d N(0, \sigma^2),$$
(19)

where $\sigma^2 = \lim_{q \to \infty} q \operatorname{var}(\bar{Y}_{i_{\bullet}})$.

Next, we assume some weak dependence conditions on time by supposing that the following invariance principle or functional central limit theorem holds for any $t \in (0, 1)$ [29,30]:

$$\frac{1}{\sqrt{nq}}\sum_{i=1}^{[nt]} (Y_{i\bullet} - \overline{Y}_{\bullet\bullet}) \Rightarrow \tilde{\sigma}\{\mathcal{B}(t) - t\mathcal{B}(1)\},\tag{20}$$

where $\overline{Y}_{\bullet\bullet} = \sum_{i=1}^{n} Y_{i\bullet} / n = \hat{\mu}_{1,n}$ and $\tilde{\sigma}^2 = \lim_{nq \to \infty} nq \operatorname{var}(\overline{Y}_{\bullet\bullet})$.

The weak dependence conditions in (19) and (20) are satisfied if the series is *m*-dependence, mixing, or linear process. Shao and Zhang [31] proposed a normalized change point statistic

$$M_{n,q}(\boldsymbol{Y}_{\boldsymbol{\cdot}}) = \max_{k} \frac{n}{w_k} \left\{ \sum_{i=1}^{k} (Y_{i\boldsymbol{\cdot}} - \bar{Y}_{\boldsymbol{\cdot}}) \right\}^2,$$
(21)

where $Y_{\bullet} = (Y_{1,\bullet}, \dots, Y_{n,\bullet})^{\top}$ and $w_k = \sum_{i=1}^k \{\sum_{j=1}^i Y_{j\bullet} - (i/k) \sum_{j=1}^i Y_{j\bullet} \}^2 + \sum_{i=k+1}^n \{\sum_{j=i}^n Y_{j\bullet} - (n-i+1)/(n-k) \sum_{j=k+1}^n Y_{j\bullet} \}^2$ is a random weight. They showed that

They showed that

$$\frac{M_{n,q}(\mathbf{Y}_{\boldsymbol{\cdot}})}{q} \rightarrow_{d} \max_{0 < t < 1} \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^{2}}{D_{1,0,t} + D_{2,t,1}},$$

where $D_{1,0,t} = \int_0^t \{\mathcal{B}(s) - (s/t)\mathcal{B}(t)\}^2 ds$ and $D_{2,t,1} = \int_t^1 [\mathcal{B}(1) - \mathcal{B}(s) - (1-s)/(1-t) \{\mathcal{B}(1) - \mathcal{B}(t)\}]^2 ds$.

Similarly, with the same w_k as above, we propose a randomized version of WC:

$$S_{n,q}(\mathbf{Y}_{\bullet}) = \sum_{k=1}^{n-1} \frac{1}{w_k} \left\{ \sum_{i=1}^k (Y_{i\bullet} - \bar{Y}_{\bullet\bullet}) \right\}^2.$$
 (22)

By the functional central limit theorem, when $n \to \infty$,

$$\frac{S_{n,q}(\mathbf{Y}_{\boldsymbol{\cdot}})}{q} \rightarrow_d \int_0^1 \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^2}{D_{1,0,t} + D_{2,t,1}} dt.$$

3. Power and Change Point Estimation

Considering the WC statistic $S_n(\overline{Y}, \tau, 2)$ in (14), we now consider the power of change point test based on

$$\frac{S_n(\mathbf{Y}_{.};\tau,2)}{(n-1)^{-1}\left\{\sum_{i=1}^n (\overline{Y}_{i.} - \hat{\mu}_{1,n})\right\}^2},$$
(23)

under the alternative hypothesis in Section 2.4. We assume some weak dependence conditions in Section 2.6. We note that (23) has the same asymptotic null distribution as (4) in Theorem 1. The asymptotic distribution is shown in Theorem 2. To establish the consistency of the test, we make a further assumption that the change point index k^* is bounded away from the endpoints.

Theorem 3. Assume $E(Y_{i,j}) = \mu_i = \mu_-$ if $i \le k^*$, μ_+ otherwise. Under the alternative hypothesis, the change magnitude $\Delta = \mu_+ - \mu_- \ne 0$. Under a weak dependence satisfying (19) and (20), $0 < \tau_1 \le k^*/n \le \tau_2 < 1$, τ_1 and τ_2 are two constants, $nq\Delta^2 \rightarrow \infty$, and $n^{3/2}q^{-1/2}|\Delta| \rightarrow \infty$, $qS_n(Y_*;\tau,2) \rightarrow_p \infty$.

The proof of Theorem 3 is in Appendix E. As expected, the power of the test based on (23) increases with n, q, and the size of the change in the mean.

The estimated change point is

$$\hat{k}(\tau) = \arg \max_{1 \le k < n} w_k^{-1/2}(\tau) \left| \sum_{i=1}^k (Y_{i\bullet} - \overline{Y}_{\bullet\bullet}) \right|.$$
(24)

We refer the reader to Bai [32,33] for some early works on the asymptotic distribution of $\hat{k}(n/2)$ and [34] for a treatment on the convergence rate of $\hat{k}(n/2)$.

11 of 23

4. Simulations

The main purpose of this simulation is to assess the effect of different values for $w_k(\tau)$, n, q, and change magnitude on the power of our test in (23), and that of the graph-based tests [25,26,35], as both can handle high-dimensional data, and the distance of the graph can be changed to test different changes of parameters for a fair comparison. For example, if we are not sure whether the mean or variance changes, the Euclidean distance can be used to measure the distance between any two nodes in the graph:

$$d_{i_1,i_2} = \left\{ \sum_{j=1}^{q} (Y_{i_1,j} - Y_{i_2,j})^2 \right\}^{1/2};$$
(25)

see Chen and Zhang [35], and Shi, Wu and Rao [25]. Another pseudo-distance can be used

$$d_{i_1,i_2}^* = |Y_{i_1,\bullet} - Y_{i_2,\bullet}|, \tag{26}$$

if only the change in the mean needs to be detected; see Shi, Wu and Rao [26]. We denote the maximal test of Chen and Zhang based on Euclidean distance by MST and based on the pseudo-distance by MST*. The associated algorithm is in the R package gSeg [36]. Similarly, we denote Shi, Wu, and Rao's test (Shi, Wu and Rao [25,26]) based on Euclidean distance by SHP and based on the pseudo-distance by SHP*, and the associated R package can be accessed from [37].

First, we simulate $\{Y_{i,j}, 1 \le i \le k^*, 1 \le j \le q\}$ independent standard normal random variables and $\{Y_{i,j}, k^* + 1 \le i \le n, 1 \le j \le q\}$ independent normal random variables with mean Δ and variance 1. The critical values for $\alpha = 0.05$ are given in Table 1 with $p = 1 - \alpha$. We use these critical values and generate 200 simulations with sample sizes n = 40, 80, dimensions q = 50, 100, change point locations $k^* = n/4, n/2, 3n/4$, and change magnitude $\Delta = 0.1, 0.2$.

In Table 2, we show the percentage of rejections of the null hypothesis at level 0.05 for each of the change point tests. We can see that the power of the graph-based method MST* or SHP* is higher than that of MST and SHP, which use the pseudo-distance for detecting changes in the mean. Interestingly, the power of the graph-based method for change point detection is still not as high as that of (23). This aspect of the comparison, which we have not seen in other literature so far, is considered a new and meaningful comparison, and at least we can claim that there is room for improvement in the change point detection of the graph-based method.

Now we look at the effect of the weights on the power. This weight $w_k(n/2)$ yields the highest power when the change point is in the middle; however, the $w_k(n)$ weight yields the highest power when the change point is near the beginning of the sequence, and conversely, the $w_k(0)$ weight yields the highest power when the change point is near the end of the sequence. Moreover, the power increases with increasing n, q, and Δ , which agrees with Theorem 3.

Now, we introduce a mixture distribution and slightly change the way the random variables are generated. We simulate $\{Y_{i,j}, k^* + 1 \le i \le n, 1 \le j \le q\}$ from a mixture of two normal distributions with mixture weights (0.5, 0.5) or (0.8, 0.2), means (0, 0.2) or (0, 1), and variance always being (1, 1), which corresponds to $\Delta = 0.1$ or $\Delta = 0.2$. We keep the other settings from the previous comparison. As we expected, the difference between Table 2 and Table 3 is very small.

п	40										80													
9	50							100					50					100						
Δ		0.1			0.2			0.1			0.2			0.1			0.2			0.1			0.2	
$\frac{k^*}{n}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$																					
$w_k(0)$	23	41	36	67	91	84	39	72	72	96	100	100	43	74	67	97	100	100	75	96	91	100	100	100
$w_k(n/2)$	31	43	31	82	92	82	51	73	62	99	100	99	57	77	61	99	100	100	87	97	89	100	100	100
$w_k(n)$	32	43	23	89	92	64	59	73	46	99	100	91	61	74	49	100	100	97	91	97	76	100	100	100
MST	3	5	4	7	6	10	3	5	5	6	19	9	3	4	5	7	21	10	5	8	4	13	35	14
MST*	18	24	16	44	38	44	33	29	38	75	77	75	21	21	21	65	80	69	36	43	35	95	99	98
SHP	4	6	5	7	9	12	4	6	9	10	16	6	3	8	6	9	22	13	5	6	8	13	24	16
SHP^*	10	13	8	33	37	33	17	23	18	67	77	65	9	12	13	49	71	54	22	32	21	90	97	92

Table 2. Estimated power (%) for the $w_k(0)$, $w_k(n/2)$, and $w_k(n)$ in (23), MST, MST^{*}, SHP, and SHP^{*}, based on 200 simulations; *n* are the sample sizes, *q* are the dimensions, k^* are the change point locations, and Δ is the size of the change in the mean of the normal random variables.

Table 3. Estimated power (%) for the $w_k(0)$, $w_k(n/2)$, $w_k(n)$, MST, MST^{*} in (23), SHP, and SHP^{*}, based on 200 simulations; *n* are the sample sizes, *q* are the dimensions, k^* are the change point locations, and Δ is the size of the change in the mean of mixed normal distributions.

п	40										80													
q	50							100					50					100						
Δ		0.1			0.2			0.1			0.2			0.1			0.2			0.1			0.2	
$\frac{k^*}{n}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$																					
$w_k(0)$	18	35	40	64	92	81	37	74	62	89	100	98	43	72	59	94	100	99	73	98	90	100	100	100
$w_k(n/2)$	28	36	37	80	94	74	50	76	57	98	100	97	53	74	52	98	100	99	84	98	87	100	100	100
$w_k(n)$	32	38	27	83	95	56	58	75	42	99	100	92	60	72	44	100	100	96	89	99	73	100	100	100
MST	4	5	6	4	5	4	6	6	6	5	10	6	4	4	6	4	24	20	5	6	2	8	25	20
MST*	20	11	24	41	44	39	27	28	27	70	72	64	22	21	24	60	75	65	38	39	33	95	100	95
SHP	4	4	8	10	15	13	6	5	10	17	28	21	8	6	5	10	29	18	9	11	6	28	54	41
SHP*	9	7	11	26	41	26	16	20	18	55	70	57	12	12	15	44	59	49	22	28	23	92	96	88

5. Data Analysis

Here, we analyze the video data provided by Dr. Mathieu Lihorea, which are available from [26]. In Lihoreau, Chittka and Raine [38], the authors used artificial pollen to attract bees and an automatic monitoring camera to capture the bee's flight path. However, this automatic monitoring feature does not fully start recording when the bee enters and stops recording when the bee leaves, in fact in this video, the recording starts before the bee enters and does not stop when the bee leaves. Since we only care about the part of the video with bees, detecting the arrival and departure of bees helps us to automatically cut the original video. Although the video contains the interference of ants, the bees are much larger compared to the ants, so it can be assumed that the presence and departure of the bees cause a change in the mean value of the pixel values of the image.

This video has a length of 49 seconds, a frame width of 352, a frame height of 288, and a frame rate of 29.97 frames per second. Shi, Wu and Rao [26] extracted the video into n = 49 images according to the rate of one frame per second. From these 49 images, we can obtain that the image positions corresponding to the bee entering and leaving are 4 and 40, respectively. Moreover, we can extract this video into more images according to the rate of 2 or 5 frames per second. So, the number of images obtained, n, increases to 98 or 245, and at the same time, the positions of the image locations where these bees appear and leave as change points, k^* , we assume that k^*/n is constant with respect to n and close to 0 or 1, respectively. In Figure 3 the first row is four images located at 4 (change point), 5, 40

(change point), and 41 from extracted 49 images; the second row is four images located at 7 (change point), 8, 79 (change point), and 80 from extracted 98 images; and the third row is four images located at 19 (change point), 20, 198 (change point), and 199 from extracted 245 images. Since the images contain R, G, and B components, we use a weighted average of the R, G, and B components and same-scale transformations on the weighted average as suggested by Shi, Wu and Rao [26].



Figure 3. Typical images in three different image sets extracted from the same video data with different frame rates. The first row contains four images located at 4 (change point), 5, 40 (change point), and 41 from the first set of extracted 49 images (1 frame per second); the second row contains four images located at 7 (change point), 8, 79 (change point), and 80 from the second set of extracted 98 images (2 frames per second); and the third row contains four images located at 19 (change point), 20, 198 (change point), and 199 from the third set of extracted 245 images (5 frames per second).

Our quadratic weight test statistics are able to detect these two change points. We compared them to the graph-based change point estimates by applying the method of SHP* and MST* once to the whole sequence. As shown in Table 4, all tests are significant at a level 0.05 except the quadratic weight $w_k(0)$ for the size of 49 returns a *p*-value 0.067; $w_k(0)$ and $w_k(n)$ give the estimates of the second and first change points, respectively; $w_k(n/2)$ gives the same estimates of change points as $w_k(n)$, and both cannot give the estimates of the second change point, such as SHP* and MST*. Thus, we recommend these two weights $w_k(0)$ and $w_k(n)$ for detecting the departure and arrival of the bee.

Table 4. Estimated change points for the $w_k(0)$, $w_k(n/2)$, $w_k(n)$, MST^{*}, and SHP^{*}, based on extracted 49, 98, and 245 images; *n* are the sample sizes and k^* are the change point locations.

п	4	.9	ç	98	245				
k^*	4	40	7	79	19	198			
$w_k(0)$		41		82		206			
$w_k(n/2)$	4		8		19				
$w_k(n)$	4		8		19				
MST*	4		7		19				
SHP*	4		7		19				

6. Discussion

This paper mainly focuses on single change point detection. However, it is possible to extend our method and apply the WC statistic to the detection of multiple change points. An approach recommended in the literature is to select data intervals where there is evidence for a single change point. Some researchers suggested penalty procedures based on either the adaptive lasso [39] or smoothly clipped absolute deviation [40–42]; others applied CUSUM statistics [4,43–45]. As long as the aforementioned intervals have been chosen, one could use tests based on WC. If the tests are rejected for some of the intervals, then the change point can be estimated by (24).

It would also be of interest, although challenging, to consider other quadratic weights, such as $w_k(n/4)$ and $w_k(3n/4)$, as these statistics may be more powerful to detect some change points that are close to the third-quarter and quarter positions of the sequence. The eigenvalues of these quadratic terms may not have recursive formulas.

Author Contributions: X.S., X.-S.W. and N.R. designed research; X.S., X.-S.W. and N.R. performed research; X.S. analyzed data; X.S., X.-S.W. and N.R. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: Shi's work was supported by NSERC Discovery Grant RGPIN 2022-03264, the Interior Universities Research Coalition and the BC Ministry of Health, and the University of British Columbia Okanagan (UBC-O) Vice Principal Research in collaboration with UBC-O Irving K. Barber Faculty of Science.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank two anonymous reviewers for helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Theorem 1. The exact distribution of $S_n(Y; \tau, 2)$ is determined by the eigenvalues of Q. Define the $n \times (n-1)$ matrix $B = (B_1, \dots, B_{n-1})$ with $B_k = p_k^{-1/2}(0, \dots, 0, 1, -1, 0, \dots, 0)^\top$ such that all entries of B_k are zeros except the k entry $p_k^{-1/2}$ and the k + 1 entry $-p_k^{-1/2}$. It is readily seen that $A^\top B = nI_{n-1}$ and thus $B^\top QB = I_{n-1}$. Note that

$$B^{\top}B = P^{-1/2}TP^{-1/2} = \begin{pmatrix} p_1^{-1/2} & & \\ & \ddots & \\ & & p_{n-1}^{-1/2} \end{pmatrix} \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \begin{pmatrix} p_1^{-1/2} & & \\ & \ddots & \\ & & p_{n-1}^{-1/2} \end{pmatrix},$$

where *P* is a diagonal matrix with $P_{kk} = p_k$ and *T* is a tridiagonal matrix with $T_{kk} = 2$ and $T_{k,k+1} = T_{k+1,k} = -1$.

We shall find the relationship between Q and $B^{\top}B$. We diagonalize $B^{\top}B = R\Gamma R^{\top}$ with $RR^{\top} = I_n$ and Γ diagonal matrix with $\Gamma_{kk} = \gamma_k$. Set $C = BR\Gamma^{-1/2}$. We have $C^{\top}C = I_{n-1}$ and $C^{\top}QC = \Gamma^{-1}$. Finally, we introduce $u = (n^{-1/2}, \dots, n^{-1/2})^{\top}$ such that Qu = 0, $C^{\top}u = 0$ and $u^{\top}u = 1$. Define U = (C, u). It then follows that $U^{\top}U = I_n$ and

$$U^{\top}QU = \begin{pmatrix} C^{\top}QC & C^{\top}Qu \\ u^{\top}QC & u^{\top}Qu \end{pmatrix} = \begin{pmatrix} \Gamma^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

This implies that the nonzero eigenvalues of *Q* are reciprocals of those of $B^{\top}B$.

Let $(v_1, \dots, v_{n-1})^{\top}$ be the eigenvector corresponding to an eigenvalue λ of $B^{\top}B$. We have the recurrence identity

$$-p_{k-1}^{-1/2}v_{k-1} + 2p_k^{-1/2}v_k - p_{k+1}^{-1/2}v_{k+1} = \lambda p_k^{1/2}v_k,$$
(A1)

where $v_0 = v_n = 0$, $p_0 = 1$, and $k = 1, \dots, n-1$. The above recurrence relation appears in Gardner [6] (1.7) as an eigenvalue equation for a forward difference operator. As mentioned in Gardner [6], it is difficult to find an explicit formula for the eigenvalues unless the prior distribution is uniform; i.e., p_k is independent of k. To overcome this difficulty, we make use of the above recurrence relation and then apply the classical theory of orthogonal polynomials and special functions. To be more specific, we shall link the eigenvector in (A1) to the dual Hahn polynomial by making some transformations.

Let $v_k = (p_k p_{k-1})^{1/2} \cdots (p_1 p_0)^{1/2} f_k$. The above recurrence relation becomes

$$-\frac{f_{k-1}}{p_k p_{k-1}} + \frac{2f_k}{p_k} - f_{k+1} = \lambda f_k$$

We further denote $g_k = (-1)^k f_k$ and $\pi_k = g_{k+1}$. It is readily seen that

$$\frac{g_{k-1}}{p_k p_{k-1}} + \frac{2g_k}{p_k} + g_{k+1} = \lambda g_k,$$

and (by shifting the index *k*)

$$rac{\pi_{k-1}}{p_k p_{k+1}} + rac{2\pi_k}{p_{k+1}} + \pi_{k+1} = \lambda \pi_k,$$

where $\pi_{-1} = \pi_{n-1} = 0$ and $\pi_0 = 1$. By induction, π_k is a monic *k*th order polynomial of λ . Now, we consider three quadratic weights in (2).

Case I. If $p_k = 1/\{k(n-k)\}$ for $k = 1, \dots, n-1$, then π_k is related to the dual Hahn polynomial

$$\pi_k = \lim_{N \to n-2} (2)_k (-N)_k R_k (\lambda - 2; 1, 1, N) = \lim_{N \to n-2} (2)_k (-N)_k \sum_{j=0}^k \frac{(-k)_j (-x)_j (x+3)_j}{(1)_j (2)_j (-N)_j},$$

where $(a)_j := \prod_{i=0}^{j-1} (a+i)$ is the Pochhammer symbol which is commonly used in the field of orthogonal polynomials and special functions, R_k is the dual Hahn polynomial of degree k and $\lambda - 2 = x(x+3)$. In particular,

$$\pi_{n-1} = (-1)^{n-1} (-x)_{n-1} (x+3)_{n-1} = \prod_{j=0}^{n-2} (x-j)(x+3+j) = \prod_{j=0}^{n-2} [\lambda - (j+1)(j+2)].$$

This implies that the eigenvalues of $B^{\top}B$ are (j+1)(j+2) for $j = 0, \dots, n-2$. Consequently, the eigenvalues of Q are 0 and $1/\{k(k+1)\}$ for $k = 1, \dots, n-1$.

Case II. If $p_k = 1/\{k(2n-k)\}$ for $k = 1, \dots, n-1$, then π_k is related to the dual Hahn polynomial as follows

$$\pi_k = (2)_k (2-2n)_k R_k (\lambda - 2; 1, 1, 2n - 2) = (2)_k (2-2n)_k \sum_{j=0}^k \frac{(-k)_j (-x)_j (x+3)_j}{(1)_j (2)_j (2-2n)_j},$$

where R_k is the dual Hahn polynomial of degree *k* and $\lambda - 2 = x(x + 3)$. In particular,

$$\pi_{n-1} = (2)_{n-1}(2-2n)_{n-1} \sum_{j=0}^{n-1} \frac{(1-n)_j(-x)_j(x+3)_j}{(1)_j(2)_j(2-2n)_j}.$$

By Watson's sum, $\pi_{n-1} = 0$ when x = 2k - 1 or x = -2 - 2k with $k = 1, \dots, n-1$. This implies that the eigenvalues of $B^{\top}B$ are (2k)(2k+1) for $k = 1, \dots, n-1$. Consequently, the eigenvalues of Q are 0 and $1/\{2k(2k+1)\}$ with $k = 1, \dots, n-1$.

Case III. If $p_k = 1/\{(n+k)(n-k)\}$ for $k = 1, \dots, n-1$, then the eigenvalues of Q are also 0 and 1/[2k(2k+1)] with $k = 1, \dots, n-1$ because the sequence $\{p_1, \dots, p_{n-1}\}$ is just the reverse of that in Case II.

Since it is a quadratic form in normal random variables, the results follow. \Box

Appendix **B**

Proof of Theorem 2. Case I. For the weight $2t - t^2$, we define $X_t = \frac{\mathcal{B}(t) - t\mathcal{B}(1)}{\sqrt{2t - t^2}}$. By the Karhunen–Loève expansion,

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t),$$

where random variables Z_k are stochastically independent normal and $e_k(\cdot)$ are an orthonormal basis. Then, the integral of the square of X_t becomes $\sum_{k=1}^{\infty} Z_k^2$, and we need the variance of Z_k . We consider the covariance of X_t , called the Mercer Kernel:

$$K_X(t,s) = E(X_t X_s) = \frac{\min(t,s) - ts}{\sqrt{2t - t^2}\sqrt{2s - s^2}}.$$

By Mercer's theorem, there exists a set $\{\lambda_k, e_k(t)\}$ such that

$$K_X(t,s) = \sum_{k=1}^{\infty} \lambda_k e_k(t) e_k(s)$$

where λ_k are eigenvalues and $e_k(t)$ are eigenfunctions satisfying the Fredholm integral equation $\int_0^1 K_X(t,s)e_k(s)ds = \lambda_k e_k(t)$.

Thus, we have the eigenvalue problem

$$\int_0^1 \frac{\min(t,s) - ts}{\sqrt{2t - t^2}\sqrt{2s - s^2}} e(s)ds = \lambda e(t).$$

Denote $e(t) = \sqrt{2t - t^2} f(t)$. After the multiplication of $\sqrt{2t - t^2}$ on both sides, the above eigenvalue problem becomes

$$\int_0^1 [\min(t,s) - ts] f(s) ds = \lambda (2t - t^2) f(t).$$

Let $f(t) = \sum_{k=0}^{\infty} f_k t^k$. It follows that

$$\sum_{k=0}^{\infty} f_k \int_0^1 [\min(t,s) - ts] s^k ds = \lambda \sum_{k=0}^{\infty} f_k (2t - t^2) t^k.$$

Note that

$$\int_0^1 [\min(t,s) - ts] s^k ds = \frac{t - t^{k+2}}{(k+1)(k+2)}.$$

We then have

$$\sum_{k=0}^{\infty} \frac{f_k t}{(k+1)(k+2)} - \sum_{k=0}^{\infty} \frac{f_k t^{k+2}}{(k+1)(k+2)} = \lambda \sum_{k=0}^{\infty} 2f_k t^{k+1} - \lambda \sum_{k=0}^{\infty} f_k t^{k+2}.$$

Obviously, $\lambda \neq 0$; otherwise $f_k = 0$ for all $k \ge 0$. Now, we compare the coefficients of t^k with $k \ge 0$ on both sides of the above identity. It follows that

$$2\lambda f_0 = \sum_{k=0}^{\infty} \frac{f_k}{(k+1)(k+2)},$$

and

$$-\frac{f_k}{(k+1)(k+2)} = \lambda(2f_{k+1} - f_k).$$

We shall prove that $\lambda_n = \{1/[(2n)(2n+1)], n = 1,...\}$ are eigenvalues of $K_X(t,s)$. To see this, we obtain from the above recurrence relation

$$\frac{f_{k+1}}{f_k} = \frac{(k+1)(k+2) - (2n)(2n+1)}{2(k+1)(k+2)} = \frac{(k+1-2n)(k+2n+2)}{2(k+1)(k+2)}$$

Making use of the Pochhammer symbol $(a)_j := \prod_{i=0}^{j-1} (a+i)$, we obtain

$$\frac{f_k}{f_0} = \frac{(1-2n)_k(2n+2)_k}{2^k(1)_k(2)_k}$$

Consequently,

$$2\lambda f_0 = \sum_{k=0}^{\infty} \frac{f_k}{(k+1)(k+2)} = f_0 \sum_{k=0}^{\infty} \frac{(1-2n)_k (2n+2)_k}{2^k (1)_{k+1} (2)_{k+1}} = \frac{2f_0}{(-2n)(2n+1)} \sum_{j=1}^{\infty} \frac{(-2n)_j (2n+1)_j}{2^j (1)_j (2)_j}$$

Since

$$\sum_{j=0}^{\infty} \frac{(-2n)_j (2n+1)_j}{2^j (1)_j (2)_j} = 0,$$

we then have

$$2\lambda f_0 = \frac{-2f_0}{(-2n)(2n+1)} = \frac{2f_0}{(2n)(2n+1)},$$

which agrees with $\lambda = 1/[(2n)(2n+1)]$. By normalizing $f_0 = 1$, we can express the eigenfunction as

$$f(t) = \sum_{k=0}^{\infty} \frac{(1-2n)_k (2n+2)_k}{2^k (1)_k (2)_k} t^k = \frac{1}{2n} P_{2n-1}^{(1,1)} (1-t),$$

where

$$P_n^{(\alpha,\beta)}(z) = \frac{(\alpha+1)_n}{n!} \sum_{j=0}^n \frac{(-n)_k(n+\alpha+\beta+1)_k}{(1)_k(\alpha+1)_k} (\frac{1-z}{2})^j$$

is the Jacobi polynomial.

So, we have

$$\int_0^1 \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^2}{2t - t^2} dt = \sum_{k=1}^\infty \frac{1}{2k(2k+1)} Z_k^2.$$

where Z_k are independent normal random variables, each having mean zero and variance 1. That proves (5).

Case II. For the weight $1 - t^2$, we intend to solve the eigenvalue problem

$$\int_0^1 \frac{\min(t,s) - ts}{\sqrt{1 - t^2}\sqrt{1 - s^2}} e(s) ds = \lambda e(t).$$

Denote $e(t) = \sqrt{1 - t^2} f(t)$. After the multiplication of $\sqrt{1 - t^2}$ on both sides, the above eigenvalue problem becomes

$$\int_0^1 [\min(t,s) - ts] f(s) ds = \lambda (1 - t^2) f(t)$$

Let $f(t) = \sum_{k=0}^{\infty} f_k t^k$. It follows that

$$\sum_{k=0}^{\infty} f_k \int_0^1 [\min(t,s) - ts] s^k ds = \lambda \sum_{k=0}^{\infty} f_k (1 - t^2) t^k.$$

Note that

$$\int_0^1 [\min(t,s) - ts] s^k ds = \frac{t - t^{k+2}}{(k+1)(k+2)}.$$

We then have

$$\sum_{k=0}^{\infty} \frac{f_k t}{(k+1)(k+2)} - \sum_{k=0}^{\infty} \frac{f_k t^{k+2}}{(k+1)(k+2)} = \lambda \sum_{k=0}^{\infty} f_k t^k - \lambda \sum_{k=0}^{\infty} f_k t^{k+2} + \lambda \sum_{k=0}^{\infty$$

Obviously, $\lambda \neq 0$; otherwise $f_k = 0$ for all $k \ge 0$. Now, we compare the coefficients of t^k with $k \ge 0$ on both sides of the above identity. It follows that $f_0 = 0$, and

$$\lambda f_1 = \sum_{k=0}^{\infty} \frac{f_k}{(k+1)(k+2)}$$

and

$$\frac{f_k}{(k+1)(k+2)} = \lambda(f_{k+2} - f_k)$$

On account of $f_0 = 0$, the above recurrence relation implies that $f_{2j} = 0$ for all $j \ge 0$. Now, we set $g_j = f_{2j+1}$ with $j \ge 0$. The above recurrence relation (with k = 2j + 1) becomes

$$\frac{g_j}{(2j+2)(2j+3)} = \lambda(g_{j+1} - g_j).$$

For convenience, we let $1/\lambda = 2\mu(2\mu + 1)$. It is readily seen that

$$\frac{g_{j+1}}{g_j} = \frac{(2j+2)(2j+3) - (2\mu)(2\mu+1)}{(2j+2)(2j+3)} = \frac{(j+1-\mu)(j+3/2+\mu)}{(j+1)(j+3/2)}$$

Making use of the Pochhammer symbol $(a)_j := \prod_{i=0}^{j-1} (a+i)$, we obtain

$$\frac{f_{2j+1}}{f_1} = \frac{g_j}{g_0} = \frac{(1-\mu)_j(3/2+\mu)_j}{(1)_j(3/2)_j}.$$

Consequently,

$$\sum_{k=0}^{\infty} \frac{f_k}{(k+1)(k+2)} = \sum_{j=0}^{\infty} \frac{f_{2j+1}}{(2j+2)(2j+3)} = \frac{f_1}{6} \sum_{j=0}^{\infty} \frac{(1-\mu)_j (3/2+\mu)_j}{(2)_j (5/2)_j}$$

The left-hand side is λf_1 . When $\mu = n + 1$ with $n = 0, 1, \dots$, by Pfaff-Saalschütz identity, we calculate the right-hand side as

$$\frac{f_1}{6}\sum_{j=0}^{\infty}\frac{(1-\mu)_j(3/2+\mu)_j(1)_j}{(2)_j(5/2)_j(1)_j} = \frac{f_1}{6}\cdot\frac{(1)_n(-1/2-n)_n}{(2)_n(-3/2-n)_n} = \frac{f_1}{(2n+2)(2n+3)}.$$

Hence, $\lambda_n = \frac{1}{(2n+2)(2n+3)}$, $n = 0, 1, \cdots$ are the eigenvalues. By normalizing $f_1 = 1$, we can express the eigenfunction as

$$f(t) = \sum_{j=0}^{\infty} f_{2j+1} t^{2j+1} = \sum_{j=0}^{n} \frac{(-n)_j (5/2+n)_j}{(1)_j (3/2)_j} t^{2j+1} = \frac{tn!}{(3/2)_n} P_n^{(1/2,1)} (1-2t^2).$$

b, we have
$$\int_{0}^{1} \left\{ \mathcal{B}(t) - t\mathcal{B}(1) \right\}_{dt}^2 dt = \sum_{j=0}^{\infty} \frac{1}{(1-2t^2)} T_j^2 dt = \frac{1}{2} T_j^2$$

So

$$\int_0^1 \frac{\{\mathcal{B}(t) - t\mathcal{B}(1)\}^2}{1 - t^2} dt = \sum_{k=1}^\infty \frac{1}{2k(2k+1)} Z_{k,k}^2$$

where Z_k are independent normal random variables, each having mean zero and variance 1. This gives (6). \Box

Appendix C

Proof of Lemma 1. It is well known that $(1 - t)\mathcal{B}(\frac{t}{1-t}) =_d \mathcal{B}(t) - t\mathcal{B}(1)$ We first show the covariance: $Cov\{(1-s)^2 \mathcal{B}^2(\frac{s}{1-s}), (1-t)^2 \mathcal{B}^2(\frac{t}{1-t})\}$ for 0 < s < t < 1. Note that

 $\mathcal{B}(\frac{s}{1-s}) \sim N(0, \frac{s}{1-s}).$ Since $\mathcal{B}^2(\frac{t}{1-t}) - \mathcal{B}^2(\frac{s}{1-s}) = \{\mathcal{B}(\frac{t}{1-t}) - \mathcal{B}(\frac{s}{1-s})\}^2 + 2\{\mathcal{B}(\frac{t}{1-t}) - \mathcal{B}(\frac{s}{1-s})\}\mathcal{B}(\frac{s}{1-s}), \text{ we}$ have $E\{\mathcal{B}^2(\frac{t}{1-t}) - \mathcal{B}^2(\frac{s}{1-s})\} = E\{\mathcal{B}(\frac{t}{1-t}) - \mathcal{B}(\frac{s}{1-s})\}^2 \text{ and } E[\mathcal{B}^2(\frac{s}{1-s})\{\mathcal{B}^2(\frac{t}{1-t}) - \mathcal{B}^2(\frac{s}{1-s})\}]$ $= E\{\mathcal{B}^2(\frac{s}{1-s})\}E\{\mathcal{B}(\frac{t}{1-t}) - \mathcal{B}(\frac{s}{1-s})\}^2$ by the independence of increments. These lead to

$$\mathbf{E}[\mathcal{B}^{2}(\frac{s}{1-s})\{\mathcal{B}^{2}(\frac{t}{1-t}) - \mathcal{B}^{2}(\frac{s}{1-s})\}] = \mathbf{E}\{\mathcal{B}^{2}(\frac{s}{1-s})\}\mathbf{E}\{\mathcal{B}^{2}(\frac{t}{1-t}) - \mathcal{B}^{2}(\frac{s}{1-s})\}$$

Therefore, we have

$$Cov\{(1-s)^{2}\mathcal{B}^{2}(\frac{s}{1-s}), (1-t)^{2}\mathcal{B}^{2}(\frac{t}{1-t})\}$$

= $(1-s)^{2}(1-t)^{2}E\{\mathcal{B}^{2}(\frac{s}{1-s})\mathcal{B}^{2}(\frac{t}{1-t})\} - s(1-s)t(1-t),$

where

$$\begin{split} \mathbf{E}\{\mathcal{B}^2(\frac{s}{1-s})\mathcal{B}^2(\frac{t}{1-t})\} &= \mathbf{E}[\mathcal{B}^2(\frac{s}{1-s})\{\mathcal{B}^2(\frac{s}{1-s}) + \mathcal{B}^2(\frac{t}{1-t}) - \mathcal{B}^2(\frac{s}{1-s})\}]\\ &= \mathbf{E}\{\mathcal{B}^4(\frac{s}{1-s})\} + \mathbf{E}\{\mathcal{B}^2(\frac{s}{1-s})\}\mathbf{E}\{\mathcal{B}^2(\frac{t}{1-t}) - \mathcal{B}^2(\frac{s}{1-s})\}\\ &= 3(\frac{s}{1-s})^2 + \frac{s}{1-s}(\frac{t}{1-t} - \frac{s}{1-s}). \end{split}$$

So, $\operatorname{Cov}\{(1-s)^2 \mathcal{B}^2(\frac{s}{1-s}), (1-t)^2 \mathcal{B}^2(\frac{t}{1-t})\} = 2s^2(1-t)^2.$ In the next step, we will show that $n^{-1}Cov_{perm} \{ C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{\ell}), C_{\mathcal{P}}(\mathbb{N}_m, \overline{\mathbb{N}}_m) \} \rightarrow 4s^2(1 - 1)$ $(t)^2$ for $\ell = \lfloor sn \rfloor$, $m = \lfloor tn \rfloor$, and $\ell < m$. We first decompose $\operatorname{Cov}_{\operatorname{perm}} \{ C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{\ell}), C_{\mathcal{P}}(\mathbb{N}_m, \overline{\mathbb{N}}_m) \}$. Note that $C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{\ell}) = C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{\ell})$

 $\overline{\mathbb{N}}_{\ell} \cap \mathbb{N}_m$) + $C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_m)$ and $C_{\mathcal{P}}(\mathbb{N}_m, \overline{\mathbb{N}}_m) = C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_m) + C_{\mathcal{P}}(\overline{\mathbb{N}}_{\ell} \cap \mathbb{N}_m, \overline{\mathbb{N}}_m)$. Then,

 $\operatorname{Cov}_{\operatorname{perm}} \{ C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{\ell}), C_{\mathcal{P}}(\mathbb{N}_{m}, \overline{\mathbb{N}}_{m}) \}$ $= \operatorname{Cov}_{\operatorname{perm}} \{ C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{\ell} \cap \mathbb{N}_{m}), C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{m}) \} + \operatorname{Cov}_{\operatorname{perm}} \{ C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{\ell} \cap \mathbb{N}_{m}), C_{\mathcal{P}}(\overline{\mathbb{N}}_{\ell} \cap \mathbb{N}_{m}, \overline{\mathbb{N}}_{m}) \}$ + Cov_{perm} { $C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_m), C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_m)$ } + Cov_{perm} { $C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_m), C_{\mathcal{P}}(\overline{\mathbb{N}}_{\ell} \cap \mathbb{N}_m, \overline{\mathbb{N}}_m)$ }.

To calculate the covariance, we need the following moments for any disjoint subsets A_1 , A_2 , and A_3 of \mathbb{N}_n . Their sizes are denoted as n_1 , n_2 and n_3 with $n_1 + n_2 + n_3 \le n$.

$$E_{\text{perm}}C_{\mathcal{P}}(\mathbb{A}_{1},\mathbb{A}_{2}) = \sum_{i=1}^{n-1} P[\{(v_{i} \in \mathbb{A}_{1}) \cap (v_{i+1} \in \mathbb{A}_{2})\} \cup \{(v_{i+1} \in \mathbb{A}_{1}) \cap (v_{i} \in \mathbb{A}_{2})\}]$$

= $\frac{2n_{1}n_{2}}{n}.$ (A2)

The following calculations of second moments need to consider three cases i = j, |i - j| = 1 and |i - j| > 1 for $1 \le i, j < n$, with $\#\{i = j | 1 \le i, j < n\} = n - 1$, $\#\{|i - j| = 1 | 1 \le i, j < n\} = 2(n - 2)$, and $\#\{|i - j| > 1 | 1 \le i, j < n\} = (n - 2)(n - 3)$. Therefore, we have

$$\begin{split} & \operatorname{E}_{\operatorname{perm}} \{ C_{\mathcal{P}}(\mathbb{A}_{1}, \mathbb{A}_{2}) C_{\mathcal{P}}(\mathbb{A}_{2}, \mathbb{A}_{3}) \} \\ &= \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} P[\{ (v_{i} \in \mathbb{A}_{1}) \cap (v_{i+1}, v_{j+1} \in \mathbb{A}_{2}) \cap (v_{j+1} \in \mathbb{A}_{3}) \} \\ & \cup \{ (v_{i} \in \mathbb{A}_{1}) \cap (v_{i+1}, v_{j+1} \in \mathbb{A}_{2}) \cap (v_{j} \in \mathbb{A}_{3}) \} \\ & \cup \{ (v_{i+1} \in \mathbb{A}_{1}) \cap (v_{i}, v_{j} \in \mathbb{A}_{2}) \cap (v_{j+1} \in \mathbb{A}_{3}) \} \\ & \cup \{ (v_{i+1} \in \mathbb{A}_{1}) \cap (v_{i}, v_{j+1} \in \mathbb{A}_{2}) \cap (v_{j} \in \mathbb{A}_{3}) \}] \\ &= \sum_{i=j,1 \leq i, j < n} 0 + \sum_{|i-j| = 1, 1 \leq i, j < n} \frac{n_{1}n_{2}n_{3}}{n(n-1)(n-2)} + \sum_{|i-j| > 1, 1 \leq i, j < n} \frac{4n_{1}n_{2}(n_{2}-1)n_{3}}{n(n-1)(n-2)(n-3)} \\ &= \frac{2n_{1}n_{3}n_{2}(2n_{2}-1)}{n(n-1)}. \end{split}$$
(A3)

In a similar way, we have

$$E_{\text{perm}}\{C_{\mathcal{P}}^{2}(\mathbb{A}_{1},\mathbb{A}_{2})\} = \sum_{i=j,1\leq i,j< n} \frac{2n_{1}n_{2}}{n(n-1)} + \sum_{|i-j|=1,1\leq i,j< n} \frac{n_{1}n_{2}(n_{1}+n_{2}-2)}{n(n-1)(n-2)} \\ + \sum_{|i-j|>1,1\leq i,j< n} \frac{4n_{1}(n_{1}-1)n_{2}(n_{2}-1)}{n(n-1)(n-2)(n-3)} \\ = \frac{2n_{1}n_{2}}{n} + \frac{2n_{1}n_{2}(n_{1}+n_{2}-2)}{n(n-1)} + \frac{4n_{1}(n_{1}-1)n_{2}(n_{2}-1)}{n(n-1)}.$$
(A4)

After tedious calculations using (A3) and (A4), we have

$$\operatorname{Cov}_{\operatorname{perm}}\{C_{\mathcal{P}}(\mathbb{N}_{\ell},\overline{\mathbb{N}}_{\ell}),C_{\mathcal{P}}(\mathbb{N}_{m},\overline{\mathbb{N}}_{m})\}=2\ell(n-m)\{2\ell(n-m)-n\}/(n^{3}-n^{2}),$$

which leads to n^{-1} Cov_{perm} { $C_{\mathcal{P}}(\mathbb{N}_{\ell}, \overline{\mathbb{N}}_{\ell}), C_{\mathcal{P}}(\mathbb{N}_m, \overline{\mathbb{N}}_m)$ } $\rightarrow 4s^2(1-t)^2$. The proof of Lemma 1 is finished. \Box

Appendix D

Proof of Lemma 2. Consider the integral

$$\mathcal{I}(\rho; a, b, \nu^2) = \int_{-\infty}^{\infty} \exp\{au - be^u - (u - \rho)^2 / (2\nu^2)\} du,$$

where a > 0 is large and b > 0. The saddle point for the phase function $au - be^u$ is $c = \log(a/b)$. We set s = u - c and define

$$az^{2}/2 = be^{u} - au - (be^{c} - ac) = a(e^{s} - s - 1)$$

such that *z* is analytic near s = 0 and $z \sim s$ as $s \rightarrow 0$. It is easily seen that

$$z = s + \frac{s^2}{6} + \frac{s^3}{36} + \frac{s^4}{270} + \cdots, \ s = z - \frac{z^2}{6} + \frac{z^3}{36} - \frac{z^4}{270} + \cdots$$

Moreover,

$$\frac{du}{dz} = \frac{az}{be^u - a} = \frac{z}{e^s - 1} = \frac{z}{s + z^2/2}.$$

Now, we can rewrite the integral as

$$\mathcal{I}(\rho; a, b, \nu^2) = (\frac{a}{be})^a \int_{-\infty}^{\infty} e^{-az^2/2} \frac{ze^{-(u-\rho)^2/(2\nu^2)}}{s+z^2/2} dz.$$

A simple calculation gives

$$\begin{split} e^{-\frac{(u-\rho)^2}{2v^2}} &= e^{-\frac{(c-\rho)^2}{2v^2} - \frac{(c-\rho)s}{v^2} - \frac{s^2}{2v^2}} = e^{-\frac{(c-\rho)^2}{2v^2}} [1 - \frac{(c-\rho)s}{v^2} - \frac{s^2}{2v^2} + \frac{(c-\rho)^2s^2}{2v^4} + O(s^3)] \\ &= e^{-\frac{(c-\rho)^2}{2v^2}} [1 - \frac{(c-\rho)(z-z^2/6)}{v^2} - \frac{z^2}{2v^2} + \frac{(c-\rho)^2z^2}{2v^4} + O(z^3)] \\ &= e^{-\frac{(c-\rho)^2}{2v^2}} [1 - \frac{(c-\rho)}{v^2}z + \frac{(c-\rho)v^2 - 3v^2 + 3(c-\rho)^2}{6v^4}z^2 + O(z^3)], \end{split}$$

and

$$\frac{z}{s+z^2/2} = \frac{1}{1+z/3+z^2/36+O(z^3)} = 1 - \frac{z}{3} + \frac{z^2}{12} + O(z^3).$$

Consequently,

$$\frac{ze^{-(u-\rho)^2/(2\nu^2)}}{s+z^2/2} = e^{-\frac{(c-\rho)^2}{2\nu^2}} \left[1 - \frac{3(c-\rho)+\nu^2}{3\nu^2}z + \left(\frac{(c-\rho)\nu^2-\nu^2+(c-\rho)^2}{2\nu^4} + \frac{1}{12}\right)z^2 + O(z^3)\right]$$

By Watson's lemma, we obtain

$$\begin{split} \mathcal{I}(\rho;a,b,\nu^2) \\ &= (\frac{a}{be})^a e^{-\frac{(c-\rho)^2}{2\nu^2}} \bigg\{ \int_{-\infty}^{\infty} e^{-az^2/2} [1 - \frac{3(c-\rho) + \nu^2}{3\nu^2} z + (\frac{c\nu^2 - \nu^2 + (c-\rho)^2}{2\nu^4} + \frac{1}{12})z^2] dz + O(a^{-5/2}) \bigg\} \\ &= (\frac{a}{be})^a e^{-\frac{(c-\rho)^2}{2\nu^2}} \bigg[\frac{\sqrt{2\pi}}{a^{1/2}} + (\frac{(c-\rho)\nu^2 - \nu^2 + (c-\rho)^2}{2\nu^4} + \frac{1}{12})\frac{\sqrt{2\pi}}{a^{3/2}} + O(a^{-5/2}) \bigg] \\ &= (\frac{a}{be})^a \sqrt{\frac{2\pi}{a}} e^{-\frac{(c-\rho)^2}{2\nu^2}} \bigg[1 + (\frac{(c-\rho)\nu^2 - \nu^2 + (c-\rho)^2}{2\nu^4} + \frac{1}{12})a^{-1} + O(a^{-2}) \bigg]. \end{split}$$

Appendix E

Proof of Theorem 3. We denote $\overline{Y}_{i_{\bullet}}^* = \overline{Y}_{i_{\bullet}} - E(\overline{Y}_{i_{\bullet}})$, $\hat{\mu}_{1,n}^* = \hat{\mu}_{1,n} - E\hat{\mu}_{1,n}$. Under the alternative hypothesis, $E\hat{\mu}_{1,n} = \{k^*\mu_- + (n-k^*)\mu^+\}/n$.

We first find a lower bound $S_n(\overline{Y}, \tau, 2)$, i.e., $S_n(\overline{Y}, \tau, 2) \ge S_{k^*}(\overline{Y}, \tau, 2)$. Then, we decompose the lower bound into three terms:

where $S_n(\overline{Y}^*, \tau, \gamma) = \sum_{k=1}^n w_k^{-1}(\tau) \left\{ \sum_{i=1}^k (\overline{Y}^*_{i\bullet} - \hat{\mu}^*_{1,n}) \right\}^{\gamma}$.

By the weak dependence, $qS_{k^*}(\overline{Y}^*, \tau, 2) = O_p(1)$ and $qS_{k^*}(\overline{Y}^*, \tau, 1) = O_p(|\Delta|\sqrt{\frac{q}{n}})$. Furthermore, $n^{-2}(n-k^*)^2(\mu_--\mu^+)^2\sum_{k=1}^{k^*}k^2w_k^{-1}(\tau) = O(n\Delta^2)$.

$$qS_n(\overline{Y}, \tau, 2) \to_p \infty$$
 holds because $nq\Delta^2 \to \infty$ and $n^{3/2}q^{-1/2}|\Delta| \to \infty$. \Box

References

- 1. Csörgő, M.; Horváth, L. Limit Theorems in Change-Point Analysis; Wiley: Chichester, UK, 1997.
- 2. Jiang, F.; Zhao, Z.; Shao, X. Modeling the COVID-19 infection trajectory: A piecewise linear quantile trend model. J. R. Statist. Soc. B 2021, accepted.
- 3. Liu, B.; Zhou, C.; Zhang, X.; Liu, Y. A unified data-adaptive framework for high dimensional change point detection. *J. R. Statist. Soc. B* **2020**, *82*, 933–963. [CrossRef]
- Yu, M.; Chen, X. Finite sample change point inference and identification for high-dimensional mean vectors. J. R. Statist. Soc. B 2021, 83, 247–270. [CrossRef]
- 5. Jandhyala, V.; Fotopoulos, S.; MacNeill, I.; Liu, P. Inference for single and multiple change-points in time series. *J. Time Ser. Anal.* **2013**, *34*, 423–446. [CrossRef]
- 6. Gardner, J.A. On detecting changes in the mean of normal variates. Ann. Math. Statist. 1969, 40, 116–126. [CrossRef]
- Perron, P. Dealing with structural breaks. In *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*; Mills, T.C., Patterson, K., Eds.; Publisher: Palgrave Macmillan, London, UK, 2006; pp. 278–352.
- 8. MacNeill, I. Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *Ann. Statist.* **1978**, *6*, 422–433. [CrossRef]
- 9. Daniels, H.E. Saddlepoint approximations in statistics. Ann. Math. Statist. 1954, 25, 631–650. [CrossRef]
- 10. Reid, N. Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* 1988, *3*, 213–238.
- 11. Reid, N. Approximations and asymptotics, In *Statistics Theory Model*; Essays in Honor of D.R. Cox; Chapman and Hall: London, UK, 1991; pp. 287–334.
- 12. Shi, X.; Wang, X.-S.; Reid, N. Saddlepoint approximation of nonlinear moments. Statist. Sinica 2014, 24, 1597–1611. [CrossRef]
- 13. Shi, X.; Reid, N.; Wu, Y. Approximation to the moments of ratios of cumulative sums. Can. J. Statist. 2014, 42, 325–336. [CrossRef]
- 14. Akman, V.E.; Raftery, A.E. Asymptotic inference for a change-point Poisson process. Ann. Statist. 1986, 14, 1583–1590. [CrossRef]
- 15. Loader, C.R. A log-linear model for a Poisson process change point. *Ann. Statist.* **1992**, 20, 1391–1411. [CrossRef]
- 16. Imhof, J.P. Computing the distribution of quadratic forms in normal variables. Biometrika 1961, 48, 419–426. [CrossRef]
- 17. Kuonen, D. Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika* **1999**, *86*, 929–935. [CrossRef]
- 18. Daniels, H.E. Tail probability approximations. Int. Statist. Rev. 1987, 55, 37–48. [CrossRef]
- 19. Lugannani, R.; Rice, S.O. Saddlepoint approximations for the distribution of the sum of independent random variables. *Adv. Appl. Probab.* **1980**, *12*, 475–490. [CrossRef]
- 20. Anderson, T.; Darling, D. Asymptotic theory of certain "goodness of fit"criteria based on stochastic processes. *Ann. Math. Statist.* **1952**, *23*, 193–212. [CrossRef]
- 21. de Micheaux, P. L. R Package CompQuadForm. 2017. Available online: https://cran.r-project.org/web/packages/CompQuad Form/index.html (accessed on 25 December 2020).
- 22. Anderson, T.; Darling, D. A test of "goodness of fit". J. Amer. Statist. Assoc. 1954, 49, 765–769. [CrossRef]
- 23. Wald, A.; Wolfowitz, J. On a test whether two samples are from the same distribution. *Ann. Math. Statist.* **1940**, *11*, 147–162. [CrossRef]
- 24. Biswas, M.; Mukhopadhyay, M.; Ghosh, A.K. A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika* **2014**, *101*, 913–926. [CrossRef]
- 25. Shi, X.; Wu, Y.; Rao, C.R. Consistent and powerful graph-based change-point test for high-dimensional data. *Proc. Natl. Acad. Sci.* USA 2017, 114, 3969–3974. [CrossRef]
- Shi, X.; Wu, Y.; Rao, C.R. Consistent and powerful non-Euclidean graph-based change-point test with applications to segmenting random interfered video data. *Proc. Natl. Acad. Sci. USA* 2018, 115, 5914–5919. [CrossRef]
- 27. Hall, P.; Ormerod, J.T.; Wand, M.P. Theory of Gaussian variational approximation for a Poisson mixed model. *Statist. Sinica* 2011, 21, 369–389.
- 28. Hall, P. ; Pham, T.; Wand, M.P.; Wang, S.S.J. Asymptotic normality and valid inference for Gaussian variational approximation. *Ann. Statist.* **2011**, *39*, 2502–2532. [CrossRef]
- 29. Peligrad, M. An invariance principle for *φ*-mixing sequences. Ann. Probab. **1985**, *13*, 1304–1313. [CrossRef]
- 30. Phillips, P.C.B.; Solo, V. Asymptotics for linear processes. Ann. Statist. 1992, 20, 971–1001. [CrossRef]
- 31. Shao, X.; Zhang, X. Testing for change points in time series. J. Am. Statist. Assoc. 2010, 105, 1228–1240. [CrossRef]
- 32. Bai, J. Least square estimation of a shift in linear processes. J. Time Ser. Anal. 1994, 15, 453–472. [CrossRef]
- 33. Bai, J. Estimation of a change point in multiple regressions. Rev. Econ. Stat. 1997, 79, 551–563. [CrossRef]

- 34. Kokoszka, P.; Leipus, R.D. Change-point in the mean of dependent observations. *Statist. Probab. Lett.* **1998**, *40*, 385–393. [CrossRef]
- 35. Chen, H.; Zhang, N. Graph-based change-point detection. Ann. Statist. 2015, 43, 139–176. [CrossRef]
- 36. Chen, H.; Zhang, N. gSeg: Graph-Based Change-Point Detection (G-Segmentation). R Package Version 0.1. 2014. Available online: https://cran.r-project.org/web/packages/gSeg/index.html (accessed on 27 December 2020).
- 37. Chen, M.; Shi, X.; Li, H. GraphCpClust: Graph-Based Change-Point Detection and Clustering. R Package Version 0.1. 2021. Available online: https://github.com/Meigian-Chen/GraphCpClust (accessed on 27 April 2021).
- 38. Lihoreau, M.; Chittka, L.; Raine, N.E. Monitoring flower visitation networks and interactions between pairs of bumble bees in a large outdoor flight cage. *PLoS ONE* **2016**, *11*, e0150844. [CrossRef]
- 39. Zou, H. The adaptive Lasso and its oracle properties. J. Am. Statist. Assoc. 2006, 101, 1418–1429. [CrossRef]
- 40. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Statist. Assoc.* 2001, *96*, 1348–1360. [CrossRef]
- 41. Jin, B.; Shi, X.; Wu, Y. A novel and fast methodology for simultaneous multiple structural break estimation and variable selection for non-stationary time series models. *Statist. Comput* **2013**, *23*, 221–231. [CrossRef]
- 42. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. Ann. Statist. 2010, 38, 894–942. [CrossRef]
- 43. Cho, H.; Fryzlewicz, P. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Statist. Soc. B* 2015, 77, 475–507. [CrossRef]
- 44. Fryzlewicz, P. Wild binary segmentation for multiple change-point detection. Ann. Statist. 2014, 42, 2243–2281. [CrossRef]
- 45. Wang, T.; Samworth, R.J. High dimensional change point estimation via sparse projection. J. R. Statist. Soc. B 2017, 80, 57–83. [CrossRef]