

Statistica Sinica Preprint No: SS-2019-0218

Title	A Sequential Probability Ratio Test for Sparse Gaussian Mixtures
Manuscript ID	SS-2019-0218
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0218
Complete List of Authors	Wenhua Jiang and Cun-Hui Zhang
Corresponding Author	Wenhua Jiang
E-mail	wenhuajiang@yahoo.com
Notice: Accepted version subject to English editing.	

A Sequential Probability Ratio Test for Sparse Gaussian Mixtures

Wenhua Jiang and Cun-Hui Zhang

Fudan University and Rutgers University

Abstract: We develop a one-sided sequential probability ratio test (SPRT) for detecting the presence of rare and weak signals. We prove that the test is consistent throughout the detectable region of the sparse Gaussian mixtures. This makes an interesting connection between the test of power one and the higher criticism. Unlike existing methods which uses simulated critical values, for the SPRT, the actual size of type I error is always strictly less than the nominal level. We approximate the ratio between the actual size and the nominal level by nonlinear renewal theory. The accuracy of the approximation is shown by simulation. The strong power of the SPRT for moderately sparse signals is demonstrated and explained.

Key words and phrases: Sequential probability ratio test, sparse Gaussian mixtures, test of power one, higher criticism, nonlinear renewal theory.

1. Introduction

This paper concerns a sequential probability ratio test (SPRT) for sparse Gaussian mixtures detection. The problem has been considered in Donoho and Jin (2004), Jager and Wellner

(2007), Walther (2013), Jiang and Zhang (2016) and Moscovich, Nadler and Spiegelman (2016), among others. A prominent feature of the SPRT is that it controls the type I error under any preassigned level. Besides proving the consistency of the SPRT, we prove that the ratio between the actual size and the nominal level converges to a constant via renewal theory.

Let Z_i be independent observations associated with parameter θ_i . Consider testing $H_{0,n}: \theta_i = 0$ for all $i \leq n$ against $H_{1,n}: \theta_i \neq 0$ for some $i \leq n$. Thus, the purpose is to detect the existence of some signals, not necessarily to accurately find out where the signals are or to estimate their values. Tukey (1976) introduced the notion of the higher criticism for this problem. Donoho and Jin (2004) proves that in a Gaussian sequence model with rare and weak signals, the signal vector $(\theta_1, \dots, \theta_n)$ can be essentially divided into three regimes: an undetectable regime where the signal and zero vectors are indistinguishable, a detectable regime where the signal is distinguishable from zero but a nonzero estimator cannot perform much better than zero, and an estimable regime where the signal is not only detectable but a certain nonzero estimator would significantly outperform the zero estimator.

To formalize, let (Z_i, θ_i) be i.i.d. vectors with

$$Z_i|\theta_i \sim N(\theta_i, 1), \quad \theta_i \sim G_n, \quad i = 1, \dots, n. \quad (1.1)$$

Let G_0 be the degenerate distribution at zero and $G_{1,n}$ be the distribution under which $\theta_i = \delta_i \mu_n$ where δ_i are i.i.d. Bernoulli random variables with probability ε_n . The problem of sparse Gaussian mixtures detection can be formulated as testing $H_{0,n}: G_n = G_0$ against

$H_{1,n}: G_n = G_{1,n}$. Let $\mu_n = \sqrt{2r \log n}$ and $\varepsilon_n = n^{-\beta}$ be the calibration. For convenience, denote the hypotheses by $H_{0,n}: \mathbb{P}_0$ and $H_{1,n}: \mathbb{P}_{n,r,\beta}$ with probability measures \mathbb{P}_0 and $\mathbb{P}_{n,r,\beta}$ such that

$$G_n = G_0 \text{ under } \mathbb{P}_0 \text{ and } G_n = G_{1,n} \text{ under } \mathbb{P}_{n,r,\beta} \text{ in (1.1), respectively.} \quad (1.2)$$

According to Ingster (1999), \mathbb{P}_0 and $\mathbb{P}_{n,r,\beta}$ are asymptotically orthogonal when $r > \rho^*(\beta)$ and contiguous when $r < \rho^*(\beta)$, where

$$\rho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \leq 3/4, \\ (1 - \sqrt{1 - \beta})^2, & 3/4 < \beta < 1. \end{cases} \quad (1.3)$$

A test for \mathbb{P}_0 has optimal detection boundary if its power under $\mathbb{P}_{n,r,\beta}$ approaches one when $r > \rho^*(\beta)$. Ingster (2002) provided such an adaptive test by a relatively complex procedure. Let $p_i = \Phi(-Z_i)$ be the p -values and $p_{(1)} \leq \dots \leq p_{(n)}$ be the ordered p -values. The higher criticism statistic is

$$\text{HC}_n^* = \max_{1 \leq i \leq \alpha_0 n} \frac{\sqrt{n}(i/n - p_{(i)})}{\sqrt{p_{(i)}(1 - p_{(i)})}}. \quad (1.4)$$

Let $h(n, \alpha)$ be the critical value such that $\mathbb{P}_0\{\text{HC}_n^* > h(n, \alpha)\} = \alpha$. Donoho and Jin (2004) proposed to reject H_0 when $\text{HC}_n^* > h(n, \alpha_n)$ and proved its optimality when the level $\alpha_n \rightarrow 0$ slowly enough.

The sequential probability ratio test (Wald, 1947) is a well-known optimal stopping rule for testing two simple statistical hypotheses. Suppose Y_1, Y_2, \dots are sequentially observed. Let \mathbb{P}_0 and \mathbb{P}_1 be two measures and $\Lambda_k = (d\mathbb{P}_1/d\mathbb{P}_0)(Y_1, \dots, Y_k)$ be the likelihood ratio

between the two measures based on the first k observations. The SPRT takes the stopping rule

$$N = \inf \{k: \log \Lambda_k \notin (-a, b)\}, \quad (1.5)$$

where a and b are two positive constants. When $N < \infty$, the SPRT accepts \mathbb{P}_1 if $\log \Lambda_N \geq b$ and accept \mathbb{P}_0 if $\log \Lambda_N \leq -a$. We note that this leaves the option of not to make a decision when $N = \infty$. Let $\alpha = \mathbb{P}_0\{N < \infty, \log \Lambda_N \geq b\}$ and power = $\mathbb{P}_1\{N < \infty, \log \Lambda_N \geq b\}$. It is well-known that the size of the SPRT is bounded by $\alpha \leq (\text{power})e^{-b}$ and $(1 - \text{power}) \leq (1 - \alpha)e^{-a}$. In particular, for one-sided SPRT with $a = \infty$,

$$\mathbb{P}_0\{\Lambda_k \geq 1/\alpha \text{ for some } k \geq 1\} \leq \alpha. \quad (1.6)$$

We note that the observations Y_1, Y_2, \dots are not required to be independent under \mathbb{P}_0 , although the SPRT has been extensively studied in the case where Y_1, Y_2, \dots are i.i.d. under \mathbb{P}_0 .

Robbins (1970) proposed the test of power one, which is a one-sided SPRT to test a simple hypothesis \mathbb{P}_0 against a composite alternative $\{\mathbb{P}_{1,\eta}, \eta \in B\}$. In suitable settings, the one-sided SPRT with a preassigned size α can achieve power one against all alternatives $\mathbb{P}_{1,\eta}, \eta \in B$, even when $\mathbb{P}_{1,\eta}$ can be arbitrarily close to \mathbb{P}_0 . Robbins' idea is to consider a mixed likelihood ratio. Let $\Lambda_k(y_1, \dots, y_k|\eta)$ be the likelihood ratio based on the first k observations under $\mathbb{P}_{1,\eta}$ with respect to \mathbb{P}_0 . The test of power one typically takes the following form:

$$\text{Reject } \mathbb{P}_0 \text{ when } N_\alpha = \inf \left\{ k: \int \Lambda_k(Y_1, \dots, Y_k|\eta) dG(\eta) \geq 1/\alpha \right\} < \infty, \quad (1.7)$$

where G is a probability measure on B .

The test of power one is closely related to the sparse Gaussian mixture detection as the alternatives can be arbitrarily close to the null in both problems. We are interested in developing a one-sided SPRT as a solution to the problem of detecting sparse Gaussian mixtures. The proposed test makes an interesting connection between the test of power one and the higher criticism. However, it seems that the SPRT has some advantages. Firstly, as can be seen from (1.6), if we set the critical value as $1/\alpha$, the SPRT controls the type I error under α . This advantage of the SPRT is parallel to the main feature of the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) which controls the false discovery rate in multiple testing. Secondly, simulation is required to find the critical values for the procedures in the literature mentioned at the beginning of this paper. The SPRT avoids simulating critical values. Thirdly, by plotting the boundaries of the SPRT and the higher criticism in Subsection 4.2, it turns out that the SPRT might have strong performance for moderately sparse signals ($1/2 < \beta \leq 3/4$). This is confirmed in power comparison in Subsection 4.3.

The rest of this paper is organized as follows. In Section 2 we introduce a one-sided SPRT based on ordered p -values and prove its adaptive optimality. We study the asymptotic ratio between size of the test and the nominal level in Section 3. This is done by approximating the boundary values and then applying nonlinear renewal theory. In Section 4 we provide some simulation results of size approximation and power comparison. The strong power of

the SPRT for moderately sparse signals is demonstrated and explained. Proofs of theorems, lemmas and propositions are given in Section 5.

2. A One-sided SPRT Based on p -values

2.1 The test statistic

Consider probability measures \mathbb{P}_0 and \mathbb{P}_1 and a random vector $\mathbf{X} = (X_1, \dots, X_n)$ with $0 < X_i < 1, i = 1, \dots, n$. Let $X_{(1)} \leq \dots \leq X_{(n)}$ be the ordered statistics of the components of \mathbf{X} . For $m = 0$ and 1 , denote the joint density of the first k ordered statistics $(X_{(1)}, \dots, X_{(k)})$ under \mathbb{P}_m by $f_{m,k,n}(x_1, \dots, x_k)$. To test \mathbb{P}_0 against \mathbb{P}_1 based on $(X_{(1)}, \dots, X_{(k)})$, the one-sided SPRT rejects \mathbb{P}_0 when

$$\Lambda_{k,n} = \frac{f_{1,k,n}(X_{(1)}, \dots, X_{(k)})}{f_{0,k,n}(X_{(1)}, \dots, X_{(k)})} \geq 1/\alpha \text{ for some } k = 1, \dots, n. \quad (2.1)$$

In the sparse Gaussian mixture detection, $X_k = p_k$ are independent p -values with the uniform distribution in $(0, 1)$ under \mathbb{P}_0 .

We shall develop alternative measures \mathbb{P}_1 so that the SPRT has optimal detection boundary in sparse Gaussian mixture detection. When a small fraction of Z_i in (1.1) are contaminated by $N(\mu_n, 1)$ with $\mu_n > 0$, the p -values have a tendency to be small. Let $\mathbb{P}_{1,\eta}$ be probability measures under which the components of \mathbf{X} are i.i.d. uniform variables in $(0, 1/\eta)$ with $\eta > 1$. Let

$$\mathbb{P}_0 = \mathbb{P}_{1,1} \text{ and } \mathbb{P}_1 = \int \mathbb{P}_{1,\eta} dG(\eta), \quad (2.2)$$

where G is a probability measure on $(1, \infty)$. Under $\mathbb{P}_{1,\eta}$, the joint density of the ordered statistics $X_{(1)} \leq \dots \leq X_{(k)}$ is

$$f_{k,n}(x_1, \dots, x_k; \eta) = \frac{n! \eta^k (1 - \eta x_k)^{n-k}}{(n-k)!}, \quad x_k < 1/\eta.$$

The mixture likelihood ratio based on $p_{(1)}, \dots, p_{(k)}$ is then

$$\Lambda_{k,n}(p_{(k)}) = \int \frac{f_{k,n}(p_{(1)}, \dots, p_{(k)}; \eta)}{f_{k,n}(p_{(1)}, \dots, p_{(k)}; 1)} dG(\eta) = \int_1^{1/p_{(k)}} \frac{\eta^k (1 - \eta p_{(k)})^{n-k}}{(1 - p_{(k)})^{n-k}} dG(\eta). \quad (2.3)$$

Denote the stopping rule by

$$N \equiv N_{n,\alpha} = \inf \{k \leq n : \Lambda_{k,n}(p_{(k)}) \geq 1/\alpha\}. \quad (2.4)$$

The one-sided SPRT rejects \mathbb{P}_0 when $N \leq n$.

In (2.2), we take the alternative measure \mathbb{P}_1 as a mixture of $\mathbb{P}_{1,\eta}$ given by the uniform distribution $(0, 1/\eta)$ with $\eta > 1$. However, in the higher criticism problem with Gaussian mixtures the true alternative distribution of p_i , given by $\mathbb{P}_{n,r,\beta}$, is not \mathbb{P}_1 or $\mathbb{P}_{1,\eta}$. Thus, in this paper the SPRT for testing \mathbb{P}_0 against \mathbb{P}_1 serves as a general machinery for signal detection against the null measure \mathbb{P}_0 . This is somewhat different from Robbins (1970) where the focus is on testing \mathbb{P}_0 against $\mathbb{P}_{1,\eta}$.

We can write $\Lambda_{k,n}(p_{(k)}) = \int \exp(n\psi(p_{(k)}, k/n; \eta)) dG(\eta)$ where

$$\psi(x, y; \eta) = y \log \eta + (1 - y) \log \frac{1 - \eta x}{1 - x}, \quad \eta x < 1. \quad (2.5)$$

Define $\psi(x, y) = (1/n) \log \int \exp(n\psi(x, y; \eta)) dG(\eta)$. Then the SPRT statistic in (2.3) can be

2.2 The optimal detection boundary of the SPRTs

written as

$$\Lambda_{k,n}(p_{(k)}) = \exp \{n\psi(p_{(k)}, k/n)\}. \quad (2.6)$$

Formula (2.6) expressed the SPRT statistic $\Lambda_{k,n}(p_{(k)})$ as a function of $\psi(p_{(k)}, k/n)$. It can be checked that $\psi(x, y; \eta)$ is decreasing in x and increasing in y . Thus by the definition, $\psi(x, y)$ is also decreasing in x and increasing in y . In some sense $\psi(p_{(k)}, k/n)$ is like a distance between $p_{(k)}$ and k/n . This will be quite useful in our analysis of the SPRT.

Another tool for the analysis of (2.3) is the Kullback-Leibler information between Bernoulli variables: $K(x, y) = x \log(x/y) + (1 - x) \log((1 - x)/(1 - y))$. Because $\psi(x, y; \eta) = K(y, x) - K(y, \eta x)$, we have

$$\exp \{n\psi(x, y)\} = \exp \{nK(y, x)\} \int \exp \{-nK(y, \eta x)\} dG(\eta). \quad (2.7)$$

One property of $K(x, y)$ is that for $y < x \leq 1/2$,

$$\frac{(x - y)^2}{2x(1 - x)} \leq K(x, y) = \int_y^x \frac{x - t}{t(1 - t)} dt \leq \frac{(x - y)^2}{2y(1 - y)}. \quad (2.8)$$

We will use (2.7) in combination with (2.6) to prove in Theorem 1 below the optimality of the SPRT.

2.2 The optimal detection boundary of the SPRT

Theorem 1. Let (Z_i, θ_i) be i.i.d. vectors as in (1.1). Let $p_i = \Phi(-Z_i)$ be the p -values and $p_{(1)} \leq \dots \leq p_{(n)}$ be the ordered p -values. Take $dG(\eta) = I_{\{\eta > 1\}} \eta^{-2} d\eta$. Let $\Lambda_{k,n}(p_{(k)})$ be

2.2 The optimal detection boundary of the SPRT9

defined as in (2.3). The SPRT rejects \mathbb{P}_0 in (1.2) when $\Lambda_{k,n}(p_{(k)}) \geq 1/\alpha$ for some $k = 1, \dots, n$.

Then,

$$\mathbb{P}_0\{\text{SPRT rejects } \mathbb{P}_0\} \leq \alpha, \quad (2.9)$$

and the test possesses the optimal detection boundary for sparse Gaussian mixtures detection. That is, for every $\mathbb{P}_{n,r,\beta}$ in (1.2) with $r > \rho^*(\beta)$,

$$\mathbb{P}_{n,r,\beta}\{\text{SPRT rejects } \mathbb{P}_0\} \rightarrow 1. \quad (2.10)$$

Theorem 1 will be proved in Section 5. Here we first give a brief explanation of the optimal detection boundary in higher criticism. The notation will be useful in the proof of Theorem 1. Let $t_n = \sqrt{2c \log n}$ be a critical value with a constant $c \geq r$, $\bar{S}_n(t_n) = n^{-1} \sum_{i=1}^n I\{Z_i > t_n\}$ be the corresponding proportion of significant observations, $q_{0,n} = \mathbb{E}_0 \bar{S}_n(t_n) = \mathbb{P}_0\{Z_i > t_n\}$ be the significance level, and $q_{1,n} = \mathbb{E}_{n,r,\beta} \bar{S}_n(t_n) = \mathbb{P}_{n,r,\beta}\{Z_i > t_n\}$ be the power. As $Z_i \sim N(0, 1)$ under the null, $q_{0,n} = \Phi(-t_n) = n^{-c+o(1)}$. As $\mathbb{P}_{n,r,\beta}$ puts mass $n^{-\beta}$ at $\mu_n = \sqrt{2r \log n}$, $q_{1,n} = (1 - n^{-\beta})\Phi(-t_n) + n^{-\beta}\Phi(-t_n + \mu_n)$. Let k^* be the critical index given by $k^*/n = \bar{S}_n(t_n)$. Since $q_{0,n}$ is the significance level at t_n , we have $p_{(k^*)} \leq \Phi(-t_n) = q_{0,n}$. It follows that

$$\frac{\sqrt{n}(k^*/n - p_{(k^*)})}{\sqrt{p_{(k^*)}(1 - p_{(k^*)})}} \geq \frac{\sqrt{n}((k^*/n) \wedge q_{1,n} - q_{0,n})}{\sqrt{q_{0,n}}}. \quad (2.11)$$

Since $\mathbb{E}_{n,r,\beta}(\bar{S}_n(t_n) - q_{1,n})^2 = q_{1,n}(1 - q_{1,n})/n$, the difference $|k^*/n - q_{1,n}| = O_{\mathbb{P}_{n,r,\beta}}(\sqrt{q_{1,n}/n})$.

2.2 The optimal detection boundary of the SPRT10

Therefore,

$$\frac{\sqrt{n}(k^*/n - p_{(k^*)})}{\sqrt{p_{(k^*)}(1 - p_{(k^*)})}} \geq \frac{\sqrt{n}(q_{1,n} - q_{0,n})}{\sqrt{q_{0,n}}} - O_{\mathbb{P}_{n,r,\beta}}\left(\sqrt{\frac{q_{1,n}}{q_{0,n}}}\right). \quad (2.12)$$

By Mill's ratio we have $q_{0,n} = n^{-c+o(1)}$, $q_{1,n} = (1 - n^{-\beta})q_{0,n} + n^{-\beta-(\sqrt{c}-\sqrt{r})^2+o(1)}$ and

$$\frac{\sqrt{n}(q_{1,n} - q_{0,n})}{\sqrt{q_{0,n}}} = n^{1/2-c/2+o(1)}(n^{-\beta-r+2\sqrt{cr}+o(1)} - n^{-\beta}). \quad (2.13)$$

Optimizing the growth rate of $n^{1/2-c/2-\beta-r+2\sqrt{cr}}$ gives that

$$\Delta = \max_{0 < c \leq 1} \{1/2 - c/2 - \beta - r + 2\sqrt{cr}\} = \begin{cases} 1/2 - \beta + r, & r \leq 1/4, \\ 1 - \beta - (1 - \sqrt{r})^2, & r > 1/4. \end{cases}$$

The maximizer is $c = \min\{1, 4r\}$. For this choice of c it is clear from (1.3) that $0 < \Delta < 1$ when $\rho^*(\beta) < r < 1$. By algebra, when c is optimized, $\sqrt{n}(q_{1,n} - q_{0,n})/\sqrt{q_{0,n}} = n^{\Delta+o(1)}$ and $\sqrt{q_{1,n}/q_{0,n}} = n^{(\Delta-(1-c)/2)/2+o(1)}$. Since $\Delta > 0$ and $c \leq 1$, the right-hand side of (2.12) is of the order $n^{\Delta+o(1)}$, which is greater than the threshold level $h(n, \alpha_n)$ for the HC statistic (1.4). This gives the optimal detection boundary.

Remark 1. Theorem 1 holds as long as dG has a mixture component $\pi_0 I_{\{\eta > 1\}} \eta^{-2} d\eta$. It also holds for many other choices of G . Another sufficient condition for the SPRT to possess the detection boundary is $G'(\frac{(k/n) \wedge q_{1,n}}{q_{0,n}}) \geq n^\kappa$ for some κ (in view of (5.6) in the proof of Theorem 1). For example, $dG(\eta) = I_{\{\eta > 1\}} \gamma \eta^{-(\gamma+1)} d\eta$ with $\gamma > 0$.

Remark 2. From (2.6) and (2.7), the SPRT is related to the Berk-Jones test statistic

$$\text{BJ}_n = n \max_{1 \leq i \leq n/2} K^+(i/n, p_{(i)}), \quad (2.14)$$

2.2 The optimal detection boundary of the SPRT₁₁

where $K^+(x, y) = K(x, y)I\{y < x\}$. Jager and Wellner (2007) proved that the Berk-Jones test possesses the optimal detection boundary. In (5.4), we actually provide the somewhat stronger $\mathbb{P}_{n,r,\beta}\{\text{BJ}_n \geq n^{\Delta+o(1)}\} \rightarrow 1$.

To discuss the connection of the SPRT to test of power one, we state a well-known fact as the following lemma.

Lemma 1. Let $p_{(1)} \leq \dots \leq p_{(n)}$ be ordered p -values. Consider a fixed $\eta > 1$. Under $\mathbb{P}_{1,\eta}$, we can write

$$p^{(k)} = \frac{Y_1 + \dots + Y_k}{(Y_1 + \dots + Y_{n+1})\eta},$$

where Y_i are i.i.d. exponential variables with mean $1/\eta$.

Lemma 1 expresses $p^{(k)}$ as a random walk because $(Y_1 + \dots + Y_{n+1})\eta/n \rightarrow 1$. This will also be useful in test size approximation in Section 3. By (2.5) and Lemma 1, the distribution of the limit of the likelihood ratio $d\mathbb{P}_{1,\eta}/d\mathbb{P}_0$ is

$$\lim_{n \rightarrow \infty} \frac{\eta^k (1 - \eta p^{(k)})^{n-k}}{(1 - p^{(k)})^{n-k}} \sim \eta^k \exp \left\{ -(\eta - 1) \sum_{i=1}^k Y_i \right\}, \quad (2.15)$$

where Y_i be i.i.d. exponential variables with mean $1/\eta$.

Proposition 1. Let Y_1, Y_2, \dots be i.i.d. exponential variables with mean $1/\eta$. Consider testing $H_0: \eta = 1$ against $H_1: \eta > 1$. Take $dG(\eta) = I_{\{\eta > 1\}}\eta^{-2}d\eta$. Then, the test given by (1.7) with $\Lambda_k(Y_1, \dots, Y_k|\eta)$ as in (2.15):

$$\Lambda_k(Y_1, \dots, Y_k|\eta) = \eta^k \exp \left\{ -(\eta - 1) \sum_{i=1}^k Y_i \right\},$$

2.3 Stopping time and power for general alternative¹²

is a level α test of power one.

This can be seen as follows. Define a sequence c_k such that $\int_1^\infty \eta^k \exp\{-(\eta-1)c_k\} dG(\eta) = 1/\alpha$. Then, $c_k = k - (1 + o(1))\sqrt{k \log k}$. Hence by the strong law of large numbers,

$$\mathbb{P}_\eta \left\{ \sum_{i=1}^k Y_i \leq c_k \text{ for some } k \geq 1 \right\} = 1, \quad \forall \eta > 1.$$

2.3 Stopping time and power for general alternative

The above discussion also applies to the p -value based SPRT statistic (2.3) with more general G which can be tailored to find power from $p_{(k)}$ for a wide range of k . To facilitate this discussion, we write

$$\Lambda_{k,n}(x) \equiv \Lambda_{k,n}(x; G) = \int_1^{1/x} \frac{\eta^k (1 - \eta x)^{n-k}}{(1 - x)^{n-k}} dG(\eta).$$

We first note that the mixture distribution $G = \sum_{j=1}^m \pi_j G_j$ can be used to pool power from different G_j . Formally, the $\Lambda_{k,n}(p_{(k)}; G)$ -based mixture test of size α has no smaller power than the $\Lambda_{k,n}(p_{(k)}; G_j)$ -based test of size $\pi_j \alpha$ because

$$\left\{ \max_{k \leq n} \Lambda_{k,n}(p_{(k)}; G) \geq 1/\alpha \right\} \supseteq \cup_{1 \leq j \leq m} \left\{ \max_{k \leq n} \Lambda_{k,n}(p_{(k)}; G_j) \geq 1/(\pi_j \alpha) \right\}.$$

Moreover, the size- α mixture test will also stop earlier than the size- (α/π_j) component tests,

$$N = \inf_k \{k \leq n : \Lambda_{k,n}(p_{(k)}; G) \geq 1/\alpha\} \leq \min_{1 \leq j \leq m} \inf_k \{k \leq n : \Lambda_{k,n}(p_{(k)}; G_j) \geq 1/(\pi_j \alpha)\}.$$

We will show below that the SPRT adaptively matches the detection power of sign tests of the form $I\{p_{(k)} \leq q_{0,n}\}$ for a wide range of k and $q_{0,n}$. Let G be the distribution function

satisfying

$$G(1+t) = \int_1^{1+t} dG(\eta) = \frac{1}{\log(\max\{e, \log(1/t)\})}, \quad t > 0. \quad (2.16)$$

As $(x\eta)^k(1-\eta x)^{n-k}$ is increasing in ηx when $\eta x \leq k/n$, for $k/n = q_{0,n} + \Delta'_n$ and $\Delta'_n > \Delta''_n > 0$,

$$\begin{aligned} \Lambda_{k,n}(q_{0,n}; G) &\geq \int_{1+\Delta''_n/q_{0,n}}^{k/(nq_{0,n})} \frac{\eta^k(1-\eta q_{0,n})^{n-k}}{(1-q_{0,n})^{n-k}} dG(\eta) \\ &\geq \left[\int_{1+\Delta''_n/q_{0,n}}^{1+\Delta'_n/q_{0,n}} dG(\eta) \right] \frac{(q_{0,n} + \Delta''_n)^k(1-q_{0,n} - \Delta''_n)^{n-k}}{q_{0,n}^k(1-q_{0,n})^{n-k}}. \end{aligned} \quad (2.17)$$

As Δ''_n can be made very close to Δ'_n and $q_{0,n} + \Delta'_n = k/n$, the SPRT is at least nearly as powerful as the likelihood ratio test based on the observation of $\#\{i \leq n: p_i \leq q_{0,n}\} = k$, up to an iterated logarithmic term. This is made precise in the following theorem.

Theorem 2. Let $p_{(1)} \leq \dots \leq p_{(n)}$ be ordered p -values. Let $\Lambda_{k,n}(p_{(k)})$ be defined as in (2.3) with G as in (2.16). Let $0 < q_{0,n} < 1/2$, $c_n = 1/(e \vee \log \log(nq_{0,n}))$ and

$$k = \left\lceil nq_{0,n} + \sqrt{2k(1-k/n)} \left\{ \frac{\log(1/\alpha) + L_n}{1 - c_n^2} \right\}^{1/2} \right\rceil \leq (n/2) \wedge \{(1 + e^{-e})nq_{0,n}\}, \quad (2.18)$$

where $L_n = \log(c_n + \log(nq_{0,n})) + 3 \log \log(c_n + \log(nq_{0,n}))$. Then,

$$\Lambda_{k,n}(q_{0,n}; G) \geq 1/\alpha. \quad (2.19)$$

Thus, the SPRT of size α rejects \mathbb{P}_0 in (1.2) when $p_{(k)} \leq q_{0,n}$. Moreover, for any alternative probability measure \mathbb{Q}_n under which p_i are i.i.d. random variables satisfying $\mathbb{Q}_n\{p_i \leq q_{0,n}\} = q_{1,n}$ with

$$\frac{1}{2} \geq q_{1,n} \geq q_{0,n} + \sqrt{2q_{1,n}(1-q_{1,n})/n} \left\{ \sqrt{\log(1/\alpha')} + \left(\frac{\log(1/\alpha) + L_n}{1 - c_n^2} \right)^{1/2} \right\} + \frac{1}{n}, \quad (2.20)$$

2.3 Stopping time and power for general alternative14

the same SPRT has at least power $1 - \alpha'$ at \mathbb{Q}_n ,

$$\mathbb{Q}_n\{\text{SPRT rejects } \mathbb{P}_0 \text{ at } N \leq k\} \geq 1 - \alpha'. \quad (2.21)$$

Consequently, when (2.20) holds with $\alpha' \leq c_0 k/(n-k)$, the expected stopping rule is bounded by $\mathbb{E}_{\mathbb{Q}_n} N \leq k(1 + c_0)$ under the alternative probability measure.

The sign test would control the type I error, $\mathbb{P}_0\{p_{(k)} \leq q_{0,n}\} \leq \alpha$, when

$$k = \left\lceil nq_{0,n} + \sqrt{2k(1 - k/n)}\sqrt{\log(1/\alpha)} \right\rceil. \quad (2.22)$$

Let \mathbb{Q}_n be the alternative probability measure as in Theorem 2. The sign test with the above k also has at least power $1 - \alpha'$ at \mathbb{Q}_n when

$$q_{1,n} \geq q_{0,n} + \sqrt{2q_{0,n}(1 - q_{0,n})/n} \left[\sqrt{\log(1/\alpha)} + \sqrt{\log(1/\alpha')} \right] + \frac{1}{n}. \quad (2.23)$$

The proofs of (2.22) and (2.23) are same as (2.21).

Theorem 2 asserts that the detection power of the SPRT, guaranteed by (2.20), matches that of the sign test $I\{p_{(k)} \leq q_{0,n}\}$ as given in (2.23) up to a logarithmic term

$$L_n \leq (1 + o(1)) \log \log(qn_0) \leq (1 + o(1)) \log \log k.$$

Thus, as the SPRT is adaptive, it pools power from the sign tests in a broad spectrum of $(q_{0,n}, k)$. Theorem 2 also proves that the crossing boundary of the SPRT at k , given by

$$\frac{k/n - p_{(k)}}{\sqrt{k(1 - k/n)}} \leq \frac{k/n - q_{0,n}}{\sqrt{k(1 - k/n)}} \leq (1 + o(1))\sqrt{\log(1/\alpha) + \log \log k}$$

via (2.18) and (2.19), is below the crossing boundary for the HC statistic in (1.4) when $\log(1/\alpha) + \log \log k$ is of smaller order than $\log \log n$.

3. Approximation for the Size of the Test

The inequality (1.6) fails to be an equality only because Λ_k does not have to hit the boundary exactly when the first crossing happens. What is the actual size of the test? The question becomes how much $\Lambda_{k,n}$ jumps over the boundary when stops. This question can be answered by nonlinear renewal theory (e.g., Lai and Siegmund, 1977; Woodroffe, 1976; Zhang, 1988).

3.1 Approximation to boundary values

In this subsection, we derive an analytic expression for the boundary value $b(k)$ such that $\Lambda_{k,n}(b(k)) = 1/\alpha$. It is certainly interesting to know the boundary for the p -value as it directly provides the threshold for the unadjusted significance level and facilitates implementation of the test. A more formal reason is that the problem in (2.4) is a nonlinear function of p -value crossing a constant boundary. Once we have an analytic expression of $b(k)$, we shall have a more standard crossing problem, that is, a drifted random walk crossing a nonlinear boundary, so that nonlinear renewal theory can be applied.

To provide more explicit formulas for implementation and ease of understanding the proof, we take $dG(\eta) = I_{\{\eta>1\}}\eta^{-2}d\eta$ throughout the rest of the section although similar but more tedious analysis can be carried out for more general G .

Let $b(k) \equiv b_{n,\alpha}(k)$ be the solution of

$$\Lambda_{k,n}(b(k)) = 1/\alpha. \tag{3.1}$$

Since $\Lambda_{k,n}(x)$ is monotone decreasing in x , $b(k)$ gives a discrete boundary under which $\Lambda_{k,n}$ is larger than $1/\alpha$. By change of variables in (2.3), we have

$$\Lambda_{k,n}(x) = \frac{\int_0^{1-x} t^{n-k}(1-t)^{k-2} dt}{x^{k-1}(1-x)^{n-k}}. \quad (3.2)$$

Let $B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$ be the incomplete beta function and $B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$. Let $I_x(a, b) = B_x(a, b)/B(a, b)$ be the regularized incomplete beta function. Then

$$\Lambda_{k,n}(x) = \frac{B(n-k+1, k-1)}{x^{k-1}(1-x)^{n-k}} I_{1-x}(n-k+1, k-1). \quad (3.3)$$

Remark 3. Expression (3.3) provides a much more convenient way to compute $\Lambda_{k,n}(p_{(k)})$ than the definition in (2.3), as it avoids integration. Indeed, $I_x(a, b)$ is the cumulative distribution function of the beta distribution $\text{Beta}(a, b)$, while $B(n-k+1, k-1)/x^{k-1}(1-x)^{n-k}$ can be computed by the density function of the beta distribution $\text{Beta}(n-k+1, k-1)$. Expression (3.3) also provides a way to compute the first two derivatives of $\log \Lambda_{k,n}(x)$, which will be useful in the analysis of the excess over the boundary in Theorem 4.

The regularized incomplete beta function $I_x(a, b)$ is related to the binomial distribution by

$$1 - I_x(a, b) = I_{1-x}(b, a) = \sum_{i=0}^{a-1} \binom{a+b-1}{i} x^i (1-x)^{a+b-1-i}. \quad (3.4)$$

This relation and the CLT give that

$$\Lambda_{k,n}(x) \approx \frac{B(n-k+1, k-1)}{x^{k-1}(1-x)^{n-k}} \Phi\left(\frac{k-2-(n-1)x}{\sqrt{(n-1)x(1-x)}}\right). \quad (3.5)$$

Table 1: Some values of λ_k in (3.7), $\alpha = 0.05$.

k	2	3	4	5	10	50	100
λ_k	0.0527	0.254	0.559	0.935	3.397	31.734	72.413

If we could prove that $\Phi\left(\frac{k-2-(n-1)b(k)}{\sqrt{(n-1)b(k)(1-b(k))}}\right) \rightarrow 1$, by (3.5),

$$\Lambda_{k,n}(b(k)) \approx \frac{B(n-k+1, k-1)}{b(k)^{k-1}(1-b(k))^{n-k}} = \frac{1/(k-1)}{\binom{n-1}{k-1}b(k)^{k-1}(1-b(k))^{n-k}}. \quad (3.6)$$

The validity of (3.6) will be proved in Theorem 3. To approximate $b(k)$, we need to solve $\binom{n-1}{k-1}b(k)^{k-1}(1-b(k))^{n-k} = \alpha/(k-1)$. By the Poisson approximation to the binomial distribution, $b(k) \approx \lambda_k/(n-1)$ where λ_k satisfies $\lambda_k^{k-1}e^{-\lambda_k}/(k-1)! = \alpha/(k-1)$. This is equivalent to

$$\log \lambda_k - \frac{\lambda_k}{k-1} = \log(\alpha \Gamma(k-1))^{1/(k-1)}, \quad (3.7)$$

where $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ is the gamma function. We may use iterations to compute λ_k in (3.7). In Table 1 we show some values of λ_k with $\alpha = 0.05$.

Applying Stirling's approximation to (3.7), we have $\log(\lambda_k/(k-1)) - \lambda_k/(k-1) + 1 = (\log \alpha)/(k-1) + o(1)$. Because $p_{(k)} \approx k/(n\eta)$ under $\mathbb{P}_{1,\eta}$, λ_k/k is close to $1/\eta$. Expanding $\log(\lambda_k/(k-1))$ at $1/\eta$ gives an analytic expression of λ_k :

$$\lambda_k = (1 + o(1)) \left\{ \frac{(2\eta - 1)(k - 1)}{\eta^2} - A_\eta(k; \alpha) \right\}, \quad (3.8)$$

where

$$A_\eta(k; \alpha) = (1 + o(1)) \frac{k-1}{\eta^2} \left\{ (\eta-1)^2 + 2\eta^2 \left(\frac{\log(1/\alpha)}{k-1} + \frac{\eta-1}{\eta} + \log \frac{1}{\eta} \right) \right\}^{1/2}. \quad (3.9)$$

The following theorem gives the approximation to boundary values $b(k)$.

Theorem 3. Let $\Lambda_{k,n}(x)$ be as in (2.3) with $dG(\eta) = I_{\{\eta>1\}}\eta^{-2}d\eta$. Let $b(k)$ be as in (3.1).

Let $\nu > 1$ be an arbitrary constant. Suppose $\alpha \rightarrow 0$ and $\log(1/\alpha)/n \rightarrow 0$. Then, for

$$k = (1 + o(1)) \frac{\log(1/\alpha)}{1/\nu - 1 - \log(1/\nu)},$$

$$\frac{\lambda_k/n}{b(k)} \rightarrow 1, \quad (3.10)$$

where λ_k is as in (3.8) with $\eta = \nu$.

3.2 Nonlinear renewal theory

Let N be the stopping rule as in (2.4). To approximate the size of the test, we have

$$\begin{aligned} \alpha^{-1} \mathbb{P}_0 \{ \text{reject } \mathbb{P}_0 \} &= \sum_{k=1}^n \mathbb{E}_1 \{ \alpha \Lambda_{k,n}(p(k)) \}^{-1} I \{ N = k \} \\ &= \int_1^\infty \mathbb{E}_{1,\eta} \exp \left\{ - (\log \Lambda_{N,n}(p(N)) - \log \Lambda_{N,n}(b(N))) \right\} I \{ N \leq n \} dG(\eta). \end{aligned} \quad (3.11)$$

Inside the integral in (3.11) involves the overshoot $\log \Lambda_{N,n}(p(N)) - \log \Lambda_{N,n}(b(N))$. We will focus on studying this overshoot under $\mathbb{P}_{1,\eta}$ for $\eta > 1$.

Since Stirling's approximation is used to analyze (3.7), (3.8) only holds when $k \rightarrow \infty$.

Actually, (3.8) is an under-approximation of λ_k and thus $b(k)$ for small k . So the probability of crossings at those k is under-approximated. Nevertheless, it does not matter. Here is the reason. Since $\mathbb{E}_0 \Lambda_{k,n}(p(k)) = 1$ by change of measure, we have

$$\lim_{\alpha \rightarrow 0} \mathbb{P}_{1,\eta} \{ N \leq k \} = \lim_{\alpha \rightarrow 0} \mathbb{E}_0 \Lambda_{k,n}(p(k)) I \{ N \leq k \} = 0, \quad \forall k > 0.$$

Thus by (3.11), we have

$$\lim_{\alpha \rightarrow 0} \alpha^{-1} \mathbb{P}_0 \{N \leq k\} = 0, \quad \forall k > 0. \quad (3.12)$$

Therefore, it suffices to consider $\Lambda_{k,n}(p(k))$ for large k in the analysis of (3.11).

The stopping rule in (2.4) can be written as

$$N = \inf \{k \leq n: \Lambda_{k,n}(p(k)) \geq 1/\alpha\} = \inf \{k \leq n: p(k) \leq b(k)\}, \quad (3.13)$$

where $\Lambda_{k,n}(b(k)) = 1/\alpha$. Let $S_n = \sum_{i=1}^n Y_i$ where Y_i are i.i.d. exponential variables with mean $1/\eta$. The representation in Lemma 1 and the boundary approximation in Theorem 3 give that as $\alpha \rightarrow 0$ and $\log(1/\alpha)/n \rightarrow 0$,

$$N \stackrel{\mathbb{P}_{1,\eta}}{\sim} \inf \left\{ k \leq n: \frac{S_k}{\eta S_{n+1}} \leq \frac{\lambda_k}{n} (1 + o(1)) \right\}, \quad (3.14)$$

where $\stackrel{\mathbb{P}_{1,\eta}}{\sim}$ means having the same distribution under $\mathbb{P}_{1,\eta}$. Define a random walk

$$\tilde{S}_k = \sum_{i=1}^k (2/\eta - 1/\eta^2 - Y_i). \quad (3.15)$$

By the strong law of large numbers, (3.14) and (3.8), as $\alpha \rightarrow 0$ and $\log(1/\alpha)/n \rightarrow 0$,

$$N \stackrel{\mathbb{P}_{1,\eta}}{\sim} \inf \{k \leq n: S_k \leq (1 + o(1))\lambda_k\} = \inf \{k \leq n: \tilde{S}_k \geq (1 + o(1))A_\eta(k; \alpha)\}, \quad (3.16)$$

where $A_\eta(k; \alpha)$ is as in (3.9). The stopping rule in (3.16) involves a random walk with drift crossing a nonlinear boundary, which is studied in nonlinear renewal theory. Following Zhang (1988), define the critical index

$$k_\eta = \sup \{k: (1 + o(1))A_\eta(k; \alpha) \geq (1/\eta - 1/\eta^2)k\}. \quad (3.17)$$

Then by (3.9),

$$k_\eta = (1 + o(1)) \frac{\log(1/\alpha)}{1/\eta - 1 - \log(1/\eta)}. \quad (3.18)$$

Let $b_\eta^* = b(k_\eta)$ be the boundary value at k_η . By Theorem 3, (3.8) and (3.18),

$$b_\eta^* = (1 + o(1)) \frac{\lambda_{k_\eta}}{n} = (1 + o(1)) \frac{k_\eta}{\eta n} = (1 + o(1)) \frac{\log(1/\alpha)/n}{1 - \eta - \eta \log(1/\eta)}. \quad (3.19)$$

Lemma 2. Let $\Lambda_{k,n}(p(k))$ be as in (2.3) with $dG(\eta) = I_{\{\eta>1\}}\eta^{-2}d\eta$. Let $b(k)$ be as in (3.1) and $b_\eta^* = b(k_\eta)$ where k_η is as in (3.18). Suppose $\alpha \rightarrow 0$. When $|k - k_\eta| = O(k_\eta^{1/2})$ and $|x - b(k)| = o(1)$,

$$\begin{aligned} & \log \Lambda_{k,n}(x) - \log \Lambda_{k,n}(b(k)) \\ &= (1 + o(1)) \log' \Lambda_{k,n}(b_\eta^*)(x - b(k)) + (1 + o(1)) \log'' \Lambda_{k,n}(b_\eta^*) b'(k_\eta) (k - k_\eta) (x - b(k)). \end{aligned} \quad (3.20)$$

Lemma 2 is a crucial element in our analysis. Although the boundary value $b(k)$ is different for each $\Lambda_{k,n}$, we wish to expand the overshoot at a fixed point b_η^* . Lemma 2 means that if we expand the overshoot at b_η^* instead of at $b(k)$, it needs to consider an additional term which involves the second derivative $\log'' \Lambda_{k,n}(b_\eta^*)$. The following theorem gives the relation between the overshoots for the renewal processes in terms of $\log \Lambda_{k,n}(p(k))$ and $np(k)$.

Theorem 4. Let $\Lambda_{k,n}(p(k))$ be as in (2.3) with $dG(\eta) = I_{\{\eta>1\}}\eta^{-2}d\eta$. Let $b(k)$ be as in (3.1) and $b_\eta^* = b(k_\eta)$ where k_η is as in (3.18). Let N be the stopping rule as in (2.4). Suppose $\alpha \rightarrow 0$ and $\log(1/\alpha)/n \rightarrow 0$. Then, under $\mathbb{P}_{1,\eta}$,

$$\log \Lambda_{N,n}(p(N)) - \log \Lambda_{N,n}(b(N)) = (1 + o_{\mathbb{P}_{1,\eta}}(1)) \log' \Lambda_{N,n}(b_\eta^*) (p(N) - b(N)), \quad (3.21)$$

where $p_{(1)} \leq \dots \leq p_{(n)}$ are ordered uniform $(0, 1/\eta)$ variables under $\mathbb{P}_{1,\eta}$.

Remark 4. Theorem 4 will be proved in Section 5 using Lemma 2. We shall prove that $\log'' \Lambda_{N,n}(b_\eta^*) b'(k_\eta)(N - k_\eta) = o(1) \log' \Lambda_{N,n}(b_\eta^*)$. The first two derivatives of $\log \Lambda_{k,n}(x)$ can be computed by (3.3).

In (5.9) in the proof of Theorem 4, we will show that $\log' \Lambda_{N,n}(b_\eta^*) = -(1 + o_{\mathbb{P}_{1,\eta}}(1))(\eta - 1)n$ under $\mathbb{P}_{1,\eta}$. Bringing this into (3.21) and then by Lemma 1 and (3.8), we have

$$\log \Lambda_{N,n}(p_{(N)}) - \log \Lambda_{N,n}(b(N)) \stackrel{\mathbb{P}_{1,\eta}}{\sim} (1 + o_{\mathbb{P}_{1,\eta}}(1))(\eta - 1)(\tilde{S}_N - (1 + o(1))A_\eta(N; \alpha)). \quad (3.22)$$

Thus by the Portmanteau theorem, for the expectation inside the integral in (3.11),

$$\begin{aligned} & \lim_{\alpha \rightarrow 0} \mathbb{E}_{1,\eta} \exp \left\{ - (\log \Lambda_{N,n}(p_{(N)}) - \log \Lambda_{N,n}(b(N))) \right\} \\ &= \lim_{\alpha \rightarrow 0} \mathbb{E}_{1,\eta} \exp \left\{ - (\eta - 1)(\tilde{S}_N - (1 + o(1))A_\eta(N; \alpha)) \right\}. \end{aligned} \quad (3.23)$$

In (3.22) and (3.23), $\tilde{S}_N - (1 + o(1))A_\eta(N; \alpha)$, the overshoot of a drifted random walk, is in the form of a more standard nonlinear renewal problem. It is the overshoot of a drifted random walk crossing a nonlinear boundary. Define $\tau(c, u) = \inf\{n: \tilde{S}_n - un > c\}$ and $R(c, u) = \tilde{S}_{\tau(c,u)} - u\tau(c, u) - c$. Applying Theorem 1 of Zhang (1988), we obtain the Blackwell-type renewal result

$$\lim_{\alpha \rightarrow 0} \mathbb{P}_{1,\eta} \{ \tilde{S}_N - (1 + o(1))A_\eta(N; \alpha) > y \} = \frac{1}{\mathbb{E}_{1,\eta} R(0, d^*)} \int_y^\infty \mathbb{P}_{1,\eta} \{ R(0, d^*) > t \} dt, \quad (3.24)$$

where $d^* = \lim_{\alpha \rightarrow 0} (\partial/\partial k)(1 + o(1))A_\eta(k_\eta; \alpha) = 2/\eta - 1/\eta^2 - (\log \eta)/(\eta - 1)$. Once the limiting distribution of the overshoot $\tilde{S}_N - (1 + o(1))A_\eta(N; \alpha)$ is obtained, the remaining

calculation of (3.23) and thus (3.11) is relatively standard in sequential analysis literature (cf. Corollary 8.45 of Siegmund, 1985).

Theorem 5. Let (Z_i, θ_i) be i.i.d. vectors as in (1.1). Let $p_i = \Phi(-Z_i)$ be the p -values and $p_{(1)} \leq \dots \leq p_{(n)}$ be the ordered p -values. Let $\Lambda_{k,n}(p_{(k)})$ be as in (2.3) with $dG(\eta) = I_{\{\eta>1\}}\eta^{-2}d\eta$. The SPRT rejects \mathbb{P}_0 when $\Lambda_{k,n}(p_{(k)}) \geq 1/\alpha$ for some $k = 1, \dots, n$. Suppose $\alpha \rightarrow 0$ and $\log(1/\alpha)/n \rightarrow 0$. Then

$$\begin{aligned} & \lim_{\alpha \rightarrow 0} \alpha^{-1} \mathbb{P}_0 \{ \text{SPRT rejects } \mathbb{P}_0 \} \\ &= \int_1^\infty \frac{\eta}{\eta \log \eta - \eta + 1} \exp \left\{ - \sum_{k=1}^\infty \frac{1}{k} \left(1 - F_k \left(\frac{k\eta \log \eta}{\eta - 1} \right) + F_k \left(\frac{k \log \eta}{\eta - 1} \right) \right) \right\} dG(\eta), \end{aligned} \quad (3.25)$$

where F_k is the cumulative distribution function of gamma distribution $\Gamma(k, 1)$.

Remark 5. Theorems 3-5 also hold for $dG(\eta) = I_{\{\eta>1\}}\gamma\eta^{-(\gamma+1)}d\eta$ with $\gamma > 0$ by essentially the same proof.

4. Simulation Results

4.1 Size approximation

Here we study the size of the test and demonstrate the validity of the approximation formula in a simulation study. We use $dG(\eta) = I_{\{\eta>1\}}\eta^{-2}d\eta$ in simulation. In (3.25), $\eta = 1$ is a singular point of $\eta/(\eta \log \eta - \eta + 1)$. So near $\eta = 1$, it requires large n for an accurate approximation due to slow divergence of $\sum_{n=1}^\infty 1/n$. As a remedy, we may divide the whole

Table 2: Size approximation from (3.25).

ε	0.1	0.05	0.01	0.005	0.001
$\lim_{\alpha \rightarrow 0} \alpha^{-1} \mathbb{P}_0 \{\text{reject } \mathbb{P}_0\}$	0.765	0.757	0.7545	0.7544	0.7544

Table 3: Type I errors of the SPRT based on 10^5 replications. In the parentheses are the ratios between the type I errors and the corresponding nominal levels.

	n			
α	500	1000	5000	10^4
0.05	0.03662 (0.732)	0.03786 (0.757)	0.03727 (0.745)	0.03816 (0.763)
0.01	0.00751 (0.751)	0.00737 (0.737)	0.00771 (0.771)	0.00757 (0.757)
0.005	0.00367 (0.734)	0.00378 (0.756)	0.00371 (0.742)	0.00387 (0.774)
0.002	0.00141 (0.705)	0.00153 (0.765)	0.00141 (0.705)	0.00162 (0.81)

region of η into two subintervals: $(1, 1 + \varepsilon)$ and $(1 + \varepsilon, \infty)$. Over $(1 + \varepsilon, \infty)$, the integral can be computed relatively quickly, which is denoted by I_1 . In view of (3.11), the integrand in (3.25) is always smaller than 1. So the integral over $(1, 1 + \varepsilon)$ is smaller than ε . Hence $I_1 \leq \lim_{\alpha \rightarrow 0} \alpha^{-1} \mathbb{P}_0 \{\text{reject } \mathbb{P}_0\} \leq I_1 + \varepsilon$ and thus $\lim_{\alpha \rightarrow 0} \alpha^{-1} \mathbb{P}_0 \{\text{reject } \mathbb{P}_0\} \approx I_1 + \varepsilon$. The values of $I_1 + \varepsilon$ with different ε are reported in Table 2. Clearly, the limit is about 0.754.

As an indication of the accuracy of our size approximation, we report type I errors with various sample sizes and significance levels in Table 3. The ratios between the type I errors and the corresponding nominal levels are in parentheses. The results demonstrate the

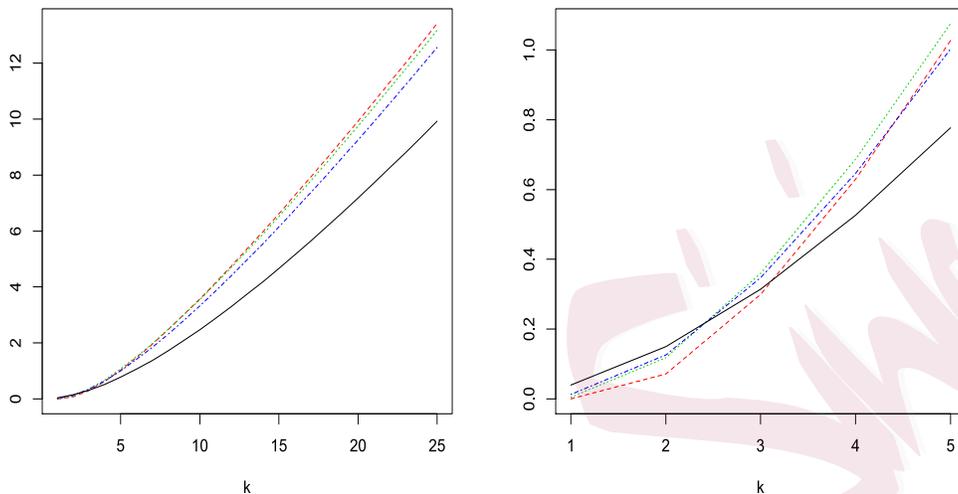


Figure 1: Left panel: scaled boundary values with $n = 5000$ and $\alpha = 0.05$. Red dashed: $nb(k)$ of the SPRT with $\gamma = 1$; green dotted: $nb(k)$ with $\gamma = 0.5$; blue dotdashed: $nb(k)$ with $\gamma = 0.1$; black solid: $nb_{\text{HC}}(k)$ of the HC. Right panel: zoom in on $k = 1, \dots, 5$.

validity of our theory of size approximation. Although (3.25) concerns the asymptotic case, it works well in practice.

4.2 Choice of G

As mentioned in Remarks 1 and 5, any prior in the family $dG(\eta) = I_{\{\eta>1\}}\gamma\eta^{-(\gamma+1)}d\eta$, $\gamma > 0$ leads to a procedure having optimal detection boundary. How to choose G ? We may compare the corresponding boundary curves. Let

$$b_{\text{HC}}(k) = \min \left\{ p_{(k)} : \frac{\sqrt{n}(k/n - p_{(k)})}{\sqrt{p_{(k)}(1 - p_{(k)})}} \leq h(n, \alpha) \right\}$$

be the boundary values of the HC, where $h(n, \alpha)$ satisfies $\mathbb{P}_0\{\text{HC}_n^* > h(n, \alpha)\} = \alpha$. We set $n = 5000$ and $\alpha = 0.05$. In the left panel of Figure 1, we plot the scaled boundary values for $k = 1, \dots, 25$. The black solid curve represents $nb_{\text{HC}}(k)$ of the HC. The red dashed, green dotted and blue dotdashed curves represent $nb(k)$ of the SPRT with $\gamma = 1, 0.5$ and 0.1 , respectively. Since the size of the SPRT is always smaller than the nominal level, the levels have been calibrated to their corresponding $\alpha / \lim_{\alpha \rightarrow 0} \alpha^{-1} \mathbb{P}_0\{\text{reject } \mathbb{P}_0\}$ by using (3.25) in Theorem 5. The calibration makes the actual levels of the SPRT and HC same. In the right panel, we zoom in on $k = 1, \dots, 5$. Larger boundary values are in favor of the detection of alternative. The green dotted curve, i.e., $\gamma = 0.5$ seems to be the preferred choice for the SPRT. Indeed, it can be seen from the right panel of Figure 1 that with $\gamma = 0.5$, $b(k) < b_{\text{HC}}(k)$ for $k = 1$ and 2 , and $b(k) > b_{\text{HC}}(k)$ for $k \geq 3$. For moderately sparse signals, there are relatively abundant small p -values. So when $b_{\text{HC}}(k) < p_{(k)} < b(k)$ for some $k \geq 3$, the SPRT wins. For very sparse signals, there are not so many small p -values. So it is more likely that $p_{(k)} > b(k)$ for all $k \geq 3$. When $b(k) < p_{(k)} < b_{\text{HC}}(k)$ for $k = 1$ or 2 , the HC wins. This explains why the moderately sparse case is in favor of the SPRT, which is confirmed in power comparison in Table 4.

Imaging that $p_{(k)}$ attains the SPRT boundary value $b(k)$. The HC-type values of the SPRT are given by

$$\text{HC-type}(k) = \frac{\sqrt{n}(k/n - b_{(k)})}{\sqrt{b_{(k)}(1 - b_{(k)})}}. \quad (4.1)$$

In Figure 2, we plot the HC-type values for $k = 1, \dots, 15$ with $n = 5000$ and $\alpha = 0.05$. We

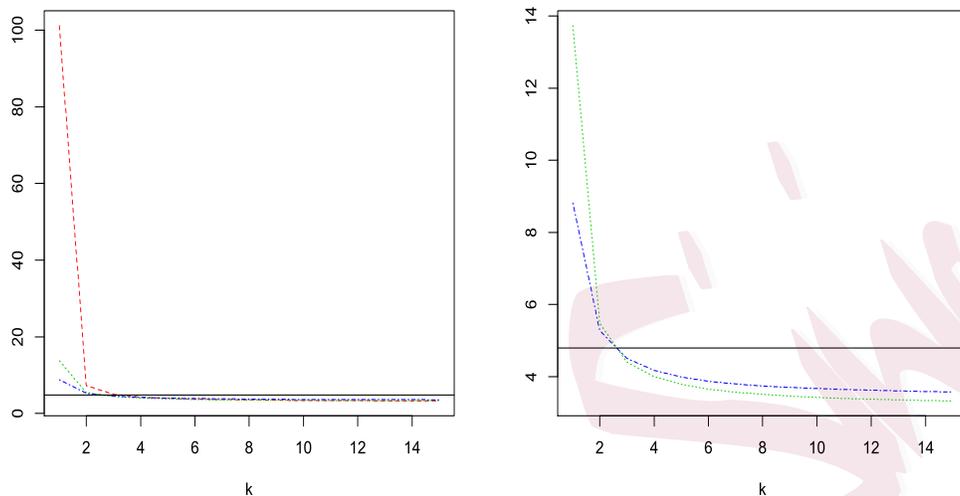


Figure 2: Left panel: SPRT's HC-type values in (4.1) with $n = 5000$ and $\alpha = 0.05$. Red dashed: $\gamma = 1$; green dotted: $\gamma = 0.5$; blue dotdashed: $\gamma = 0.1$. The black solid line represents $h(n, \alpha)$. Right panel: only $\gamma = 0.5$ and 0.1 are included.

still include $\gamma = 1, 0.5$ and 0.1 . The black solid line is $h(n, \alpha) = 4.79$. For $\gamma = 1$, HC-type(1) is extremely large. This implies that if the crossing happened at $k = 1$ with $\gamma = 1$, the corresponding HC statistic would be much larger than the critical value $h(n, \alpha)$, which is nearly impossible. In the right panel, only $\gamma = 0.5$ and 0.1 are included. Considering that HC-type(1) with $\gamma = 0.1$ is considerably larger than $h(n, \alpha)$, and HC-type(k) with $\gamma = 0.5$ are smaller than HC-type(k) with $\gamma = 0.1$ for $k \geq 3$, $\gamma = 0.5$ should be preferred. This confirms the analysis from Figure 1. In Figure 2, the effect of difference between $b(k)$ and $b_{\text{HC}}(k)$ is reflected in the domain of HC statistic.

In the very sparse case, the strongest evidence against \mathbb{P}_0 is found at the few smallest

p -values (Donoho and Jin, 2004). In view of the analysis above, certain mixture of $dG(\eta) = I_{\{\eta>1\}}\gamma\eta^{-(\gamma+1)}d\eta$ might improve upon the power. Similar as Walther (2013), suppose we start with a noninformative uniform prior for the parameter η on $(1/2, 1)$, so we have half probabilities of $(1/2, 3/4]$ and $(3/4, 1)$ respectively. This leads to consider, for example, a combination of $dG(\eta) = I_{\{\eta>1\}}0.5\eta^{-1.5}d\eta$ and $dG(\eta) = I_{\{\eta>1\}}0.1\eta^{-1.1}d\eta$ with equal weights. We call this procedure the mix-SPRT.

4.3 Power comparison

We conduct simulation experiments to evaluate the power of the SPRT in testing \mathbb{P}_0 against $\mathbb{P}_{n,r,\beta}$. Under $\mathbb{P}_{n,r,\beta}$, $\theta_i = \delta_i\mu_n$ where δ_i are i.i.d. Bernoulli random variables with probability $\varepsilon_n = n^{-\beta}$. In Table 4, we display our simulation results for seven procedures: the HC, the HC+ (a variation of HC), the Berk-Jones (BJ) test (Jager and Wellner, 2007), the exact BJ test (Moscovich, Nadler and Spiegelman, 2016), the GLRT (Jiang and Zhang, 2016), the SPRT with $dG(\eta) = I_{\{\eta>1\}}0.5\eta^{-1.5}d\eta$, and the mix-SPRT with $dG(\eta) = I_{\{\eta>1\}}0.5\eta^{-1.5}d\eta/2 + I_{\{\eta>1\}}0.1\eta^{-1.1}d\eta/2$. We set the significant level 0.05. Still, the levels are calibrated. Since the ratios are 0.615 and 0.278 for $\gamma = 0.5$ and 0.1 respectively, we set $\alpha = 0.05/0.615$ for the SPRT and $\alpha = 0.05/0.447$ for the mix-SPRT. The simulated power in Table 4 is based on 1000 replications. Except the two versions of the SPRT, the critical values are computed by simulations. We consider $n = 5000$ and $\varepsilon_n = 0.005, 0.002$ and 0.001. Let $\mu^* = \sqrt{2\rho^*(\beta)\log n}$ be the thresholding effect value of μ_n , where $\rho^*(\beta)$ is the detection boundary in (1.3). The

Table 4: Power comparison for testing $H_{0,n}: \mathbb{P}_0$ against $H_{1,n}: \mathbb{P}_{n,r,\beta}$ at level $\alpha = 0.05$. Under $\mathbb{P}_{n,r,\beta}$, $\theta_i = \delta_i \mu_n$ where δ_i are i.i.d. Bernoulli random variables with probability $\varepsilon_n = n^{-\beta}$.

	μ_n	2	2.25	2.5	2.75	3	3.25	3.5	3.75
$n = 5000$ $\varepsilon_n = 0.005$ ($\beta = 0.622$)	SPRT	0.524	0.748	0.913	0.986	0.997	1.000	1.000	1.000
	mix-SPRT	0.516	0.745	0.911	0.984	0.997	1.000	1.000	1.000
	HC	0.376	0.602	0.815	0.956	0.994	0.999	1.000	1.000
	HC+	0.490	0.682	0.876	0.969	0.993	0.999	1.000	1.000
	BJ	0.479	0.713	0.880	0.971	0.993	0.999	1.000	1.000
	Exact BJ	0.375	0.602	0.816	0.940	0.990	0.999	1.000	1.000
	GLRT	0.433	0.687	0.871	0.974	0.998	1.000	1.000	1.000
$n = 5000$ $\varepsilon_n = 0.002$ ($\beta = 0.730$)	SPRT	0.186	0.295	0.418	0.619	0.790	0.896	0.978	0.994
	mix-SPRT	0.184	0.300	0.416	0.622	0.794	0.906	0.977	0.994
	HC	0.180	0.306	0.407	0.595	0.773	0.886	0.974	0.994
	HC+	0.131	0.229	0.321	0.453	0.610	0.723	0.862	0.938
	BJ	0.152	0.258	0.358	0.543	0.719	0.870	0.962	0.988
	Exact BJ	0.115	0.183	0.283	0.433	0.610	0.800	0.924	0.981
	GLRT	0.150	0.228	0.369	0.532	0.716	0.858	0.965	0.992
$n = 5000$ $\varepsilon_n = 0.001$ ($\beta = 0.811$)	SPRT	0.108	0.156	0.222	0.339	0.428	0.598	0.722	0.855
	mix-SPRT	0.108	0.155	0.225	0.334	0.444	0.615	0.729	0.861
	HC	0.112	0.152	0.259	0.351	0.468	0.641	0.757	0.870
	HC+	0.086	0.126	0.148	0.193	0.257	0.320	0.405	0.513
	BJ	0.083	0.126	0.193	0.282	0.369	0.532	0.653	0.810
	Exact BJ	0.061	0.090	0.148	0.201	0.275	0.425	0.539	0.729
	GLRT	0.087	0.101	0.177	0.257	0.356	0.521	0.629	0.794

corresponding calibrations are $(\beta, \mu^*) = (0.622, 1.442)$, $(0.730, 1.978)$ and $(0.811, 2.333)$. We let the amplitude parameter μ_n range from 2 to 3.75 with an increment of 0.25.

These simulation results can be summarized as follows. The first group is the moderately sparse case ($1/2 < \beta \leq 3/4$). It is known that for such relatively dense signals, the HC+, BJ and GLRT have strong performance (Jiang and Zhang, 2016). Both versions of the SPRT are more powerful, especially for $\mu_n = 2, 2.25$ and 2.5 . In the second group ($\beta = 0.730$),

Table 5: $\mathbb{E}_0(N|\text{reject } H_0)$ of the SPRT based on 10^5 replications.

α	n			
	500	1000	5000	10^4
0.05	83.4	124.2	348.9	543.6
0.01	97.3	154.1	486	690.5
0.005	97.9	154.9	500	690.9
0.002	112.4	172.2	586.7	755.5

the SPRTs are slightly more powerful than the HC. The third group is the very sparse case ($3/4 < \beta < 1$). The mix-SPRT improves upon the SPRT. Although they rank the second and third, they are more powerful than other methods by large margin. See also Walther (2013) and Moscovich, Nadler and Spiegelman (2016) for strong performance of the HC for very sparse signals.

4.4 Evaluation of $\mathbb{E}_0(N|\text{reject } H_0)$

It might be interesting to report some numerical results of $\mathbb{E}_0(N|\text{reject } H_0)$. This is useful because it provides information at which p -values the SPRT rejects H_0 under the null hypothesis. We report $\mathbb{E}_0(N|\text{reject } H_0)$ with various sample sizes and significance levels in Table 5. The results are based on 10^5 replications. The simulation indicates that the crossings happen at quite early stage on average. As expected, the expectation of the stopping time increases with the decrement of the significance level, or with the increment of the sample size. It might also be interesting to have a glance at the distribution of N . As an example, we choose $n = 10^4$ and $\alpha = 0.01$. In Figure 3, we provide histograms of the stopping time based on 10^5 replications. The stopping times that are no later than 1000, 100 and 10

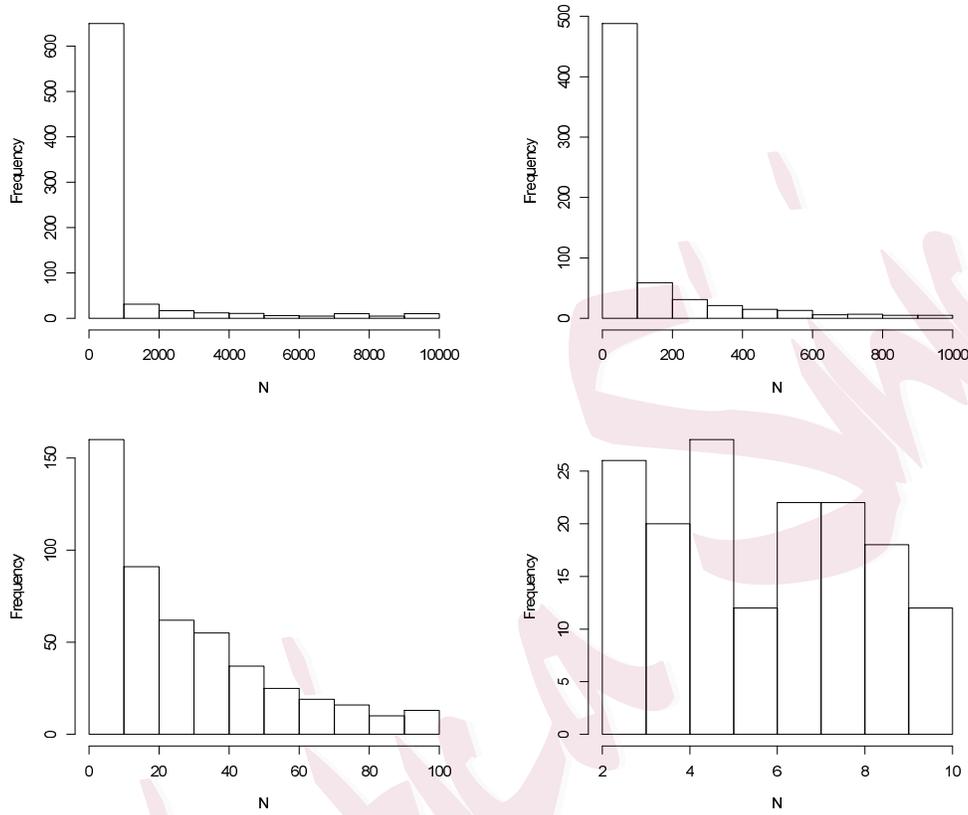


Figure 3: Top left: histogram of N based on 10^5 replications, $n = 10^4$ and $\alpha = 0.01$; top right: zoom in $N \leq 1000$; bottom left: zoom in $N \leq 100$; bottom right: zoom in $N \leq 10$.

are zoomed in turn. As reported in Table 3, there are 757 crossings in all. The numbers of crossings that are no later than 1000, 100 and 10 are 650, 488 and 160, respectively.

5. Proofs

Proof of Theorem 1. We shall use the same notation as in Section 2: $t_n = \sqrt{2c \log n}$ with $c = \min\{1, 4r\}$, $\bar{S}_n(t_n) = n^{-1} \#\{i \leq n: Z_i > t_n\}$, $q_{0,n} = \mathbb{E}_0 \bar{S}_n(t_n) = \Phi(-t_n) = n^{-c+o(1)}$,

$q_{1,n} = \mathbb{E}_{n,r,\beta} \bar{S}_n(t_n)$ and $k^*/n = \bar{S}_n(t_n)$. We want to prove

$$\mathbb{P}_{n,r,\beta} \left\{ \Lambda_{k^*,n}(p_{(k^*)}) = \exp \{ n\psi(p_{(k^*)}, k^*/n) \} \geq 1/\alpha \right\} \rightarrow 1. \quad (5.1)$$

Because $\psi(x, y)$ is decreasing in x and increasing in y , and $p_{(k^*)} \leq q_{0,n}$, it suffices to prove

$$\mathbb{P}_{n,r,\beta} \left\{ n\psi(q_{0,n}, (k^*/n) \wedge q_{1,n}) \geq \log(1/\alpha) \right\} \rightarrow 1. \quad (5.2)$$

Now we apply the Kullback-Leibler expression in (2.7). Under $\mathbb{P}_{n,r,\beta}$, $k^* = n\bar{S}_n(t_n)$ is a binomial($n, q_{1,n}$) variable, so that $|k^*/n - q_{1,n}| = O_{\mathbb{P}_{n,r,\beta}}(\sqrt{q_{1,n}/n})$. It follows from (2.8) that

$$\begin{aligned} nK((k^*/n) \wedge q_{1,n}, q_{0,n}) &\geq \frac{n((k^*/n) \wedge q_{1,n} - q_{0,n})^2}{2q_{1,n}} \\ &= \left(\frac{q_{0,n}}{q_{1,n}} \right) \frac{n(q_{1,n} - q_{0,n})^2}{2q_{0,n}} + O_{\mathbb{P}_{n,r,\beta}}(1). \end{aligned} \quad (5.3)$$

In the explanation for HC, we have shown that for $c = \min\{1, 4r\}$, $\sqrt{n}(q_{1,n} - q_{0,n})/\sqrt{q_{0,n}} = n^{\Delta+o(1)}$ and $q_{1,n}/q_{0,n} \leq n^\Delta$ with $\Delta > 0$. These and (5.3) imply that

$$\mathbb{P}_{n,r,\beta} \left\{ nK((k^*/n) \wedge q_{1,n}, q_{0,n}) \geq n^{\Delta+o(1)} \right\} \rightarrow 1. \quad (5.4)$$

The second factor in the right-hand side of (2.7) is $\int \exp \{ -nK((k^*/n) \wedge q_{1,n}, \eta q_{0,n}) \} dG(\eta)$.

Consider the case where $|(k^*/n) \wedge q_{1,n} - \eta q_{0,n}| \leq \sqrt{q_{0,n}/n}$. By (2.8), $K((k^*/n) \wedge q_{1,n}, \eta q_{0,n}) \leq 1/(\eta n^2)$. So that

$$\begin{aligned} &\int \exp \{ -nK((k^*/n) \wedge q_{1,n}, \eta q_{0,n}) \} dG(\eta) \\ &\geq \int_{|(k^*/n) \wedge q_{1,n} - \eta q_{0,n}| \leq \sqrt{q_{0,n}/n}} \exp \{ -1/(\eta n) \} dG(\eta). \end{aligned} \quad (5.5)$$

The integral is over $\frac{(k^*/n) \wedge q_{1,n}}{q_{0,n}} \pm 1/(n\sqrt{q_{0,n}})$ when $|(k^*/n) \wedge q_{1,n} - \eta q_{0,n}| \leq \sqrt{q_{0,n}}/n$. Hence, by $dG(\eta) = I_{\{\eta>1\}}\eta^{-2}d\eta$ and $q_{1,n}/q_{0,n} \leq n^\Delta$,

$$\int \exp\{-nK((k^*/n) \wedge q_{1,n}, \eta q_{0,n})\}dG(\eta) \geq c_0 \left(\frac{q_{0,n}}{q_{1,n}}\right)^2 \frac{1}{n\sqrt{q_{0,n}}} \geq n^{-2\Delta+c/2-1+o(1)}, \quad (5.6)$$

where c_0 is certain positive constant. Putting (2.7), (5.4) and (5.6) together,

$$\mathbb{P}_{n,r,\beta} \left\{ n\psi(q_{0,n}, (k^*/n) \wedge q_{1,n}) \geq n^{\Delta+o(1)} - (2\Delta - c/2 + 1 + o(1)) \log n \right\} \rightarrow 1.$$

This completes the proof of (5.2) and thus (5.1). □

Proof of Proposition 1. Let c_k be a sequence such that $\int_1^\infty \eta^k \exp\{-(\eta-1)c_k\}dG(\eta) = 1/\alpha$. Let $x = \eta - 1$ and $c_k = k(1 - \varepsilon_k)$. Then

$$\begin{aligned} 1/\alpha &= (1 + o(1)) \int_0^\infty \exp\left\{-\frac{1}{2}(k-2)x^2 + (k\varepsilon_k - 2)x\right\}dx \\ &= (1 + o(1)) \frac{c_0}{\sqrt{k-2}} \exp\left\{\frac{(k\varepsilon_k - 2)^2}{2(k-2)}\right\} \end{aligned}$$

with certain positive constant c_0 . This leads to

$$k\varepsilon_k = (1 + o(1)) \left\{ (k-2)(\log(k-2) - 2\log(c_0\alpha)) \right\}^{1/2}.$$

Hence, $c_k = k - (1 + o(1))\sqrt{k \log k}$ for fixed α . Since $c_k/k \rightarrow 1$ and $\sum_{i=1}^k Y_i/k \rightarrow 1/\eta < 1$ as $k \rightarrow \infty$, the type II error

$$\mathbb{P}_\eta \left\{ \sum_{i=1}^k Y_i > c_k \text{ for all } k \geq 1 \right\} = 0, \quad \forall \eta > 1.$$

This gives Proposition 1. □

Proof of Theorem 2. Let $A_k = q_{0,n}/\Delta'_n$. We have $A_k = nq_{0,n}/(k - nq_{0,n}) \leq nq_{0,n} \leq k$ and due to $k \leq nq_{0,n}(1 + e^{-e})$, $A_k \geq e^e$. Let $\Delta''_n = e^{-c_n} \Delta'_n$ with $c_n = 1/\log \log(nq_{0,n}) \leq 1/e$. We have

$$\int_{1+\Delta''_n/q_{0,n}}^{1+\Delta'_n/q_{0,n}} dG(\eta) = \int_{\log A_k}^{c_n + \log A_k} \frac{1}{\eta \log^2 \eta} d\eta \geq \frac{1}{\log^3(c_n + \log(nq_{0,n}))(c_n + \log(nq_{0,n}))}.$$

As $k/n - q_{0,n} = \Delta'_n > 0$ and $\Delta''_n = e^{-c_n} \Delta'_n$,

$$\begin{aligned} & \log \left(\frac{(q_{0,n} + \Delta''_n)^k (1 - q_{0,n} - \Delta''_n)^{n-k}}{q_{0,n}^k (1 - q_{0,n})^{n-k}} \right) \\ &= n \int_{q_{0,n}}^{q_{0,n} + \Delta''_n} \frac{k/n - t}{t(1-t)} dt \\ &\geq \frac{n}{(k/n)(1 - k/n)} \int_{q_{0,n}}^{q_{0,n} + \Delta''_n} (k/n - t) dt \\ &= \frac{n}{(k/n)(1 - k/n)} \left(\frac{(k/n - q_{0,n})^2}{2} - \frac{(k/n - q_{0,n} - \Delta''_n)^2}{2} \right) \\ &= \frac{n(k/n - q_{0,n})^2 (1 - (1 - e^{-c_n})^2)}{2(k/n)(1 - k/n)}. \end{aligned}$$

It follows from (2.17) and the above calculations that

$$\Lambda_{k,n}(q_{0,n}; G) \geq \frac{\exp \{ (k - nq_{0,n})^2 (1 - c_n^2) / (2k(1 - k/n)) \}}{\log^3(c_n + \log(nq_{0,n}))(c_n + \log(nq_{0,n}))}.$$

Consequently, $\Lambda_{k,n}(q_{0,n}; G) \geq 1/\alpha$ when (2.18) holds.

By the Chernoff-Hoeffding bound,

$$\mathbb{Q}_n \{ p_{(k)} > q_{1,n} \} \leq \mathbb{Q}_n \{ \text{Bin}(n, q_{1,n}) < k \} \leq \exp \{ -nK(q_{1,n}, k/n) \}.$$

As $1/n \leq q_{1,n} \leq 1/2$, $K(q_{1,n}, k/n) \geq (q_{1,n} - k/n)^2 / (2q_{1,n}(1 - q_{1,n}))$ by (2.8). Thus, (2.21)

holds under condition (2.20). \square

Proof of Theorem 3. It remains to argue that the approximation in (3.6) and the Poisson approximation are valid. That is, $\Phi\left(\frac{k-2-\lambda_k}{\sqrt{\lambda_k(1-\lambda_k/n)}}\right) \rightarrow 1$ and $\lambda_k/n \rightarrow 0$. Denote $k_\nu = \frac{\log(1/\alpha)}{1/\nu-1-\log(1/\nu)}$. It follows from (3.9) that $A_\nu(k_\nu; \alpha) = (1 + o(1))(1/\nu - 1/\nu^2)k_\nu$. So when $k = (1 + o(1))k_\nu$, $\lambda_k = (1 + o(1))k_\nu/\nu$, and thus $(k - 2 - \lambda_k)/\sqrt{\lambda_k(1 - \lambda_k/n)} \rightarrow \infty$. \square

Proof of Lemma 2. To allow differentiation, we treat $\Lambda_{k,n}(x)$, $b(k)$, λ_k and $A_\eta(k; \alpha)$ as defined for all real $k > 1$ in (3.2), (3.1), (3.7) and (3.9), respectively.

By (3.8), (3.9) and (3.18), $(\partial/\partial k)A_\eta(k_\eta; \alpha) = (1 + o(1))(2/\eta - 1/\eta^2 - (\log \eta)/(\eta - 1))$ and thus $(\partial/\partial k)\lambda_k|_{k=k_\eta} = (1 + o(1))(\log \eta)/(\eta - 1)$. Note that $k_\eta \asymp \log(1/\alpha)$. Hence $b'(k_\eta)(k - k_\eta) \asymp \sqrt{\log(1/\alpha)/n}$ when $|k - k_\eta| = O(k_\eta^{1/2})$. A similar argument gives that $(\partial^2/\partial k^2)\lambda_k|_{k=k_\eta} \asymp 1/\log(1/\alpha)$. So $b''(k_\eta)(k - k_\eta)^2 \asymp 1/n = o(1)b'(k_\eta)(k - k_\eta)$ when $|k - k_\eta| = O(k_\eta^{1/2})$. Therefore,

$$b(k) = b(k_\eta) + (1 + o(1))b'(k_\eta)(k - k_\eta). \quad (5.7)$$

By Taylor expansion, when $|x - b(k)| = o(1)$,

$$\log \Lambda_{k,n}(x) - \log \Lambda_{k,n}(b(k)) = (1 + o(1)) \log' \Lambda_{k,n}(b(k))(x - b(k)).$$

Bringing (5.7) into $\log' \Lambda_{k,n}(b(k))$ and using the first order Taylor expansion of $\log' \Lambda_{k,n}(x)$ at b_η^* complete the proof of Lemma 2. \square

Proof of Theorem 4. We first prove that when $|k - k_\eta| = O(k_\eta^{1/2})$ and $|x - b(k)| = o(1)$,

$$\log \Lambda_{k,n}(x) - \log \Lambda_{k,n}(b(k)) = (1 + o(1)) \log' \Lambda_{k,n}(b_\eta^*)(x - b(k)). \quad (5.8)$$

In view of Lemma 2, it only needs to show that the second term in (3.20) is negligible with respect to the first term. So we calculate $\log' \Lambda_{k,n}(b_\eta^*)$ and $\log'' \Lambda_{k,n}(b_\eta^*)$. By (3.3), $\log \Lambda_{k,n}(x) = -(k-1) \log x - (n-k) \log(1-x) + \log I_{1-x}(n-k+1, k-1) + C$. Since $I_x(a, b)$ is the cumulative distribution function of $\text{Beta}(a, b)$,

$$\begin{aligned} \frac{\partial}{\partial x} \log I_{1-x}(n-k+1, k-1) &= \frac{(\partial/\partial x)I_{1-x}(n-k+1, k-1)}{I_{1-x}(n-k+1, k-1)} \\ &= -\frac{(1-x)^{n-k}x^{k-2}}{I_{1-x}(n-k+1, k-1)B(n-k+1, k-1)}. \end{aligned}$$

By Stirling's approximation and some algebra, when $k = (1+o(1))k_\eta$,

$$\log \left(\frac{(1-b_\eta^*)^{n-k}(b_\eta^*)^{k-2}}{B(n-k+1, k-1)} \right) = \log n - c_1 \log(1/\alpha) - c_2 \log \log(1/\alpha) + c_3$$

with constants $c_1 > 0$, $c_2 > 0$ and c_3 . Moreover, the median of $\text{Beta}(n-k+1, k-1)$ is $1 - (1+o(1))k_\eta/n$ when $k = (1+o(1))k_\eta$. As $b_\eta^* = (1+o(1))k_\eta/(\eta n)$, $I_{1-b_\eta^*}(n-k+1, k-1) \geq 1/2$. Hence $(\partial/\partial x) \log I_{1-x}(n-k+1, k-1)|_{x=b_\eta^*} = o(n)$. A similar argument gives that $(\partial^2/\partial x^2) \log I_{1-x}(n-k+1, k-1)|_{x=b_\eta^*} = o(n^2/\log(1/\alpha))$. Therefore, when $k = (1+o(1))k_\eta$,

$$\begin{aligned} &\log' \Lambda_{k,n}(b_\eta^*) \\ &= (1+o(1)) \left(-\frac{k-1}{b_\eta^*} + \frac{n-k}{1-b_\eta^*} \right) + o(n) \\ &= (1+o(1)) \left\{ -\frac{\frac{1}{1/\eta-1-\log(1/\eta)} \log(1/\alpha) - 1}{\frac{1}{1-\eta-\eta \log(1/\eta)} (\log(1/\alpha)/n)} + \frac{n - \frac{1}{1/\eta-1-\log(1/\eta)} \log(1/\alpha)}{1 - \frac{1}{1-\eta-\eta \log(1/\eta)} (\log(1/\alpha)/n)} \right\} + o(n) \\ &= -(1+o(1))(\eta-1)n, \end{aligned} \tag{5.9}$$

and

$$\begin{aligned}
 & \log'' \Lambda_{k,n}(b_\eta^*) \\
 = & (1 + o(1)) \left(\frac{k-1}{(b_\eta^*)^2} + \frac{n-k}{(1-b_\eta^*)^2} \right) + o\left(\frac{n^2}{\log(1/\alpha)}\right) \\
 = & (1 + o(1)) \left\{ \frac{\frac{1}{1/\eta-1-\log(1/\eta)} \log(1/\alpha) - 1}{\left(\frac{1}{1-\eta-\eta \log(1/\eta)} \log(1/\alpha)/n\right)^2} + \frac{n - \frac{1}{1/\eta-1-\log(1/\eta)} \log(1/\alpha)}{\left(1 - \frac{1}{1-\eta-\eta \log(1/\eta)} \log(1/\alpha)/n\right)^2} \right\} \\
 & + o\left(\frac{n^2}{\log(1/\alpha)}\right) \\
 = & (1 + o(1)) \eta^2 (1/\eta - 1 - \log(1/\eta)) \frac{n^2}{\log(1/\alpha)}. \tag{5.10}
 \end{aligned}$$

It has been shown in the proof of Lemma 2 that $b'(k_\eta)(k-k_\eta) \asymp \sqrt{\log(1/\alpha)}/n$ when $|k-k_\eta| = O(k_\eta^{1/2})$. This, (5.9) and (5.10) give that $\log'' \Lambda_{k,n}(b_\eta^*) b'(k_\eta)(k-k_\eta) \asymp n/\sqrt{\log(1/\alpha)} = o(1) \log' \Lambda_{k,n}(b_\eta^*)$. Hence the second term in (3.20) is negligible.

By Proposition 1 of Zhang (1988), $k_\eta^{-1/2}(N-k_\eta)$ converges weakly to a normal distribution with mean 0 and constant variance under $\mathbb{P}_{1,\eta}$. Therefore, $|N-k_\eta| = O_{\mathbb{P}_{1,\eta}}(k_\eta^{1/2})$. As $b(k) = (1+o(1))k/(\eta n)$ when $k = (1+o(1))k_\eta$, by Lemma 1,

$$\frac{p(k)}{b(k)} - 1 \stackrel{\mathbb{P}_{1,\eta}}{\sim} (1+o(1)) \frac{(Y_1 + \dots + Y_k)/k}{(Y_1 + \dots + Y_{n+1})/n} - 1 = O_{\mathbb{P}_{1,\eta}}(k^{-1/2}).$$

So $|p(N) - b(N)| = O_{\mathbb{P}_{1,\eta}}(k_\eta^{1/2}/n) = o_{\mathbb{P}_{1,\eta}}(1)$. By (5.8), the proof is complete. \square

Proof of Theorem 5. Using (3.24), (3.23) could be evaluated as

$$\begin{aligned}
 & \lim_{\alpha \rightarrow 0} \mathbb{E}_{1,\eta} \exp \left\{ -(\eta - 1)(\tilde{S}_N - (1 + o(1))A_\eta(N; \alpha)) \right\} \\
 &= \frac{1}{\mathbb{E}_{1,\eta} R(0, d^*)} \int_0^\infty e^{-(\eta-1)y} \mathbb{P}_{1,\eta} \{ R(0, d^*) > y \} dy \\
 &= \frac{1}{(\eta - 1)\mathbb{E}_{1,\eta} R(0, d^*)} \left(1 - \mathbb{E}_{1,\eta} \exp \left\{ -(\eta - 1)R(0, d^*) \right\} \right) \\
 &= \frac{1}{(\eta - 1)\mu^*} \exp \left\{ -\sum_{k=1}^\infty \frac{1}{k} \mathbb{E}_{1,\eta} e^{-(\eta-1)(\tilde{S}_k - kd^*)_+} \right\}, \tag{5.11}
 \end{aligned}$$

where $\mu^* = \mathbb{E}_{1,\eta}(2/\eta - 1/\eta^2 - Y_i) - d^* = (\log \eta - 1 + 1/\eta)/(\eta - 1)$. The last equation above follows from Corollary 8.45 of Siegmund (1985). Since $\sum_{i=1}^k Y_i \stackrel{\mathbb{P}_{1,\eta}}{\sim} \Gamma(k, 1)/\eta$, we have

$$\tilde{S}_k - kd^* \stackrel{\mathbb{P}_{1,\eta}}{\sim} \frac{1}{\eta} \left\{ \frac{k\eta \log \eta}{\eta - 1} - \Gamma(k, 1) \right\}.$$

Thus, direct calculation shows that

$$\mathbb{E}_{1,\eta} e^{-(\eta-1)(\tilde{S}_k - kd^*)_+} = 1 - F_k \left(\frac{k\eta \log \eta}{\eta - 1} \right) + F_k \left(\frac{k \log \eta}{\eta - 1} \right), \tag{5.12}$$

where F_k is the cumulative distribution function of $\Gamma(k, 1)$. The proof of Theorem 5 is completed by (3.11), (3.23), (5.11) and (5.12). \square

Acknowledgement

The authors would like to express their sincere thanks to an Associate Editor and two referees for helpful suggestions and comments.

References

- [1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, **57**, 289-300.
- [2] Donoho, D.L. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics*, **32**, 962-994.
- [3] Ingster, Y.I. (1999). Minimax detection of a signal for l_p^n -balls. *Mathematical Methods of Statistics*, **7**, 401-428.
- [4] Ingster, Y.I. (2002). Adaptive detection of a signal of growing dimension, I, II. *Mathematical Methods of Statistics*, **10**, 395-421.
- [5] Jager, L. and Wellner, J.A. (2007). Goodness-of-fit tests via phi-divergences. *Annals of Statistics*, **35**, 2018-2053.
- [6] Jiang, W. and Zhang, C.-H. (2016). Generalized likelihood ratio test for normal mixtures. *Statistica Sinica*, **26**, 955-978.
- [7] Karlin, S. and Taylor, H. (1975). *A First Course in Stochastic Processes*. Academic Press, New York.
- [8] Lai, T.L. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *Annals of Statistics*, **5**, 946-954.

REFERENCES

- [9] Moscovich, A., Nadler, B. and Spiegelman, C. (2016). On the exact Berk-Jones statistics and their p -value calculation. *Electronic Journal of Statistics*, **10**, 2329-2354.
- [10] Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Annals of Mathematical Statistics*, **41**, 1397-1409.
- [11] Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.
- [12] Tukey, J.W. (1976). T13 N: The higher criticism. Course Notes, Statistics 411, Princeton University.
- [13] Wald, A. (1947). *Sequential Analysis*. Wiley, New York.
- [14] Walther, G. (2013). The average likelihood ratio for large-scale multiple testing and detecting sparse mixtures. In *From Probability to Statistics and Back: High Dimensional Models and Processes*, IMS Collections **9**, 317-326.
- [15] Woodroffe, M. (1976). A renewal theorem for curved boundaries and moments of first passage times. *Annals of Probability*, **4**, 67-80.
- [16] Zhang, C.-H. (1988). A nonlinear renewal theory. *Annals of Probability*, **16**, 793-824.

School of Mathematical Sciences, Fudan University and Shanghai Center for Mathematical Sciences, 220 Handan Road, Shanghai 200433, China

REFERENCES

E-mail: wenhuajiang@yahoo.com

Department of Statistics and Biostatistics, Rutgers University, 110 Frelinghuysen Road,
Piscataway, NJ 08854, U.S.A.

E-mail: czhang@stat.rutgers.edu