Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Adaptive positive and negative runs test

Jiarong Cui, Chunxiao Li, Ke Yang & Wangli Xu

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2018
VOL. 88, NO. 7, 1314–1335
https://doi.org/10.1080/00949655.2018.1430802

Adaptive positive and negative runs test

Jiarong Cui, Chunxiao Li, Ke Yang and Wangli Xu

School of Statistics, Renmin University of China, Beijing, People's Republic of China

ABSTRACT

As a nonparametric randomness test, the positive and negative runs test is widely used in practice due to the simplicity of its procedures. The test can lose efficiency if the alternative distribution is symmetrical at 0.5. In addition, the test can only be applied to test the randomness of a sequence from the uniform distribution. In this paper, we introduce an adaptive positive and negative runs test method to maximize the power function by choosing the optimal cut point. Also, the test is extended to check the randomness of a sequence generated from any other given distributions. Furthermore, we derive the exact distribution and obtain the asymptotical critical values of the proposed test statistics. Compared with the existed test, the efficiency of the proposed adaptive positive and negative runs test is competitive through simulation study.

ARTICLE HISTORY

Received 19 October 2017
Accepted 17 January 2018

KEYWORDS

Adaptive positive and negative runs test; test power; nonparametric randomness test

1. Introduction

In practice, it is common to consider whether a sample is generated randomly from the standard uniform distribution. As one of the methods to test the randomness of a sample, the positive and negative runs test, also called the binary runs test, belongs to a kind of incoherency test. The incoherency test consisting of the binary runs test and the up and down runs test proposed by Taussky and Todd [1], as well as the multi-class runs test proposed in the RAND Corporation's million random number table [2], is used to test whether the elements in a pseudo-random sample sequence come in random order. The positive and negative runs test is used to test whether a sequence of numbers $\{r_i : i = 1, 2, \ldots, n\}$, each of which comes from the interval $(0, 1)$, is generated randomly from the standard uniform distribution. The classic method sets a fixed cut point, 0.5, and divides the sequence into two classes. The elements less than 0.5 are denoted as 0, while the elements larger than 0.5 are denoted as 1. Then, the hypothesis test problem is transformed into whether a 0/1 binary sequence comes in a random order.

In fact, the positive and negative runs test, which has the same idea of constructing the test statistics as the Wald–Wolfowitz runs test and the Mood runs test, can be understood as a special case developed from them. The Wald–Wolfowitz runs test proposed by Wald and Wolfowitz [3] is used to test whether two random sample sequences come from the same

CONTACT Wangli Xu wxu.stat@gmail.com School of Statistics, Renmin University of China, Beijing 100872, People's Republic of China

population. Based on the idea of the two-sample Wald–Wolfowitz test, Mood [4] proposed a one-sample runs test examining whether the elements of a 0/1 binary sequence come in random order. The sample sequence in the Mood runs test is a binary sequence, and the null hypothesis is that the order of the sample sequence is random. The runs test is widely used not only because of its ability to process both quantitative and qualitative data but also because of its wide range of applications in various fields. For instance, in economics, the runs test can be used to study the stock market [5]; in the productive process, it can be used in quality control [6]; in academics, it can be used to solve randomness problems in contingency table analysis [7] and regression analysis [8]; and in survival analysis, it can also be applied in the hypothesis testing of the Cox proportional hazards model when the data are randomly censored [9]. In spite of the extensiveness of its applications, though, the main applications of the runs test focus on the following three aspects: the time-series analysis [10,11], the symmetry test of the single-variable distribution [12,13], and the multivariate runs test, which is emerging more recently. As for the multivariate approach, early studies have mainly focused on exploring the construction approach of the multivariate runs test [14]. However, nowadays, there are mainly two specific research directions in the multivariate field: the randomness test of the multi-dimensional sequence [15,16] and the symmetry test of the multi-dimensional distribution [15,17,18]. The development history and application fields of the runs test are thoroughly introduced in Runs test: theory [19].

In this paper, we mainly focus on improving the positive and negative runs test by selecting the cut point adaptively. According to the definition of the positive and negative runs test, the cut point is fixed at 0.5, and the corresponding statistics are thus obtained. Therefore, once the probability that the elements of the sequence lie within $(0, 0.5)$ and the probability that they lie within $(0.5, 1)$ under the alternative hypothesis are the same, the null hypothesis can lose efficiency. Hence, we propose an adaptive positive and negative runs test. On the basis of the alternative hypothesis, we seek an optimal cut point accordingly, which means we allow a variable cut point depending on different alternative hypotheses, rather than setting a fixed cut point, 0.5, all the time. The power of the adaptive positive and negative runs test is improved significantly. Additionally, the adaptive method is appropriate for more general null hypotheses, or rather, it is able to test whether a sample sequence is randomly generated from an arbitrarily given distribution.

This paper is organized as follows. In Section 2, we introduce the test statistics and derive their properties. In addition, two random variables are defined in order to simplify the process of obtaining the asymptotical critical values of the rejection region. Also, the selection method for the optimal cut point is introduced. In Section 3, we perform the numerical simulation under different alternative hypotheses to illustrate the efficiency of the adaptive test. Finally, we conclude this paper with a brief discussion in Section 4.

2. The adaptive positive and negative runs test

First, we will briefly introduce the classical positive and negative runs test. For a random sequence $\{r_i, i = 1, 2, \ldots, n\}$, the null hypothesis is that the random sample sequence follows the standard uniform distribution. Let $u_i = r_i - a$ with $a = 0.5$. We use positive and negative signs to indicate whether $u_i \geq 0$ or not and obtain a corresponding data set $\{\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_n\}$. Based on the sequence $\{\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_n\}$ with two kinds of elements, we can obtain the total number of runs, T. It is proved that the variable T follows the normal distribution with the mean $0.5n + 1$ and the variance $0.25(n - 1)$. Under the alternative hypothesis, if r_i is symmetrical about 0.5, then we have $P(u_i <= 0.5) = P(u_i > 0.5)$, which means that T follows the same distribution as under the null hypothesis. Thus, in this situation, the test cannot separate the alternative hypothesis from the null hypothesis, even if the null hypothesis is not true. One idea is letting $u_i = r_i - a$ with all the $a \in (0, 1)$ to obtain the corresponding total number of runs T. The distribution of T under the null and alternative hypotheses should be different for some $a \in (0, 1)$. Below, we study the properties of the test statistics T.

2.1. The properties of statistics under the null hypothesis

Based on the set $\{\tilde{r}_i,\ i = 1, 2, \ldots, n\}$, the continuous elements with the same sign between two elements with different signs are defined as runs. The number of elements in a run is denoted by l. The number of the runs with length l is denoted by T_l, and the total number of runs is denoted by $T = \sum T_l$. Let $n_0 = \sum_{i=1}^{n} I(u_i \geq 0)$ and $n_1 = \sum_{i=1}^{n} I(u_i < 0) = n - n_0$. Under the condition that $n_0 = n$ or $n_1 = n$, $T = 1$ is the smallest value of T. As for the largest value of T, the number of the runs with length 1 should be maximized and the corresponding value is $2\min(n_0, n_1)$. Because there is another run with length $\max(n_0, n_1) - \min(n_0, n_1)$, we have $T = 2\min(n_0, n_1) + 1$. Thus, the range of total number of runs T is $1 \leq T \leq 2\min(n_0, n_1) + 1$ for the given n_0, n_1.

Below, in order to study the property of statistic T, we first study the properties of $P\{T = 2k \mid n_1\}$ and $P\{T = 2k + 1 \mid n_1\}$. Given n_1, there are $C_n^{n_1}$ combinations if we choose n_1 negative elements among the total n elements. If $T = 2k$, then there are k negative runs and k positive runs. For given n_1 and k, there are $C_{n_1-1}^{k-1}$ combinations if we choose k runs among the sequence of n_1 negative elements. Similarly, we have $C_{n-n_1-1}^{k-1}$ combinations for k runs among the sequence of $n - n_1$ positive elements. Note that the first run of the n elements can be from the n_1 negative elements or $n - n_1$ positive elements, then the total number of combinations for $T = 2k$ is $2C_{n-n_1-1}^{k-1}C_{n_1-1}^{k-1}$. Hence,

$$P\{T = 2k|n_1\} = \frac{2C_{n-n_1-1}^{k-1}C_{n_1-1}^{k-1}}{C_n^{n_1}}. \tag{1}$$

When $T = 2k+1$, there are $k+1$ negative runs and k positive runs or k negative runs and $k+1$ positive runs. It can be similarly derived that

$$P\{T = 2k + 1 \mid n_1\} = \frac{C_{n-n_1-1}^{k-1}C_{n_1-1}^{k} + C_{n-n_1-1}^{k}C_{n_1-1}^{k-1}}{C_n^{n_1}}. \tag{2}$$

Then we have,

$$\begin{aligned} P\{T = 2k, n_1\} &= P\{T = 2k|n_1\}P\{n_1\} \\ &= (2C_{n_1-1}^{k-1}C_{n-n_1-1}^{k-1}/C_n^{n_1})a^{n_1}(1-a)^{n-n_1}C_n^{n_1} \\ &= 2C_{n_1-1}^{k-1}C_{n-n_1-1}^{k-1}a^{n_1}(1-a)^{n-n_1}. \end{aligned}$$

When $T = 2k$, there are k negative runs and k positive runs. Since the length of each run is no less than 1, we have $k \leq \min(n_0, n_1)$. Hence,

$$P\{T = 2k\} = \sum_{n_1=k}^{n-k} 2C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} a^{n_1}(1-a)^{n-n_1}. \tag{3}$$

It can be similarly derived that

$$P\{T = 2k+1\} \\ = \sum_{n_1=k+1}^{n-k} C_{n_1-1}^{k} C_{n-n_1-1}^{k-1} a^{n_1}(1-a)^{n-n_1} + \sum_{n_1=k}^{n-k-1} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k} a^{n_1}(1-a)^{n-n_1} \tag{4}$$

and $P(T = 1) = a^n + (1-a)^n$. The following theorem shows the properties of the statistics T.

Theorem 2.1: *Under the null hypothesis, the mean and variance of the statistics T are* $E(T) = 2a(1-a)(n-1) + 1 = \mu$ *and* $\mathrm{Var}(T) = 4(5-3n)a^2(1-a)^2 + 2(2n-3)a(1-a) = \sigma^2$, *respectively. If a = 0.5 (which is defined in the classical positive and negative runs test), T obeys the normal distribution asymptotically. If* $a \neq 0.5$, *T does not converge to any smooth continuous distribution.*

According to Theorem 2.1, if $a \neq 0.5$, the statistics T diverge. Therefore, we plot Figure 1 for intuitive illustration under the condition that $n = 500$ and $a = 0.1, 0.3, 0.5$. From Figure 1, we can see that the probability of T being odd and even is quite different when $a \neq 0.5$. The farther a departs from 0.5, the more separative the probability of T is from being odd to being even. Note that the probability of T is symmetrical about 0.5. We only show the case of $a \leq 0.5$.

Since it is impossible to compute the asymptotical critical value according to the asymptotic distribution of T, one idea is to make use of the exact distribution of T in Equations (3) and (4) as follows.

Theorem 2.2: *Under the null hypothesis, the critical values of statistics T are*

$$\tilde{t}_1 = \min_x \left\{ x : \sum_{k=1}^{x} P(T = k) > \alpha/2 \right\},$$

$$\tilde{t}_2 = \max_x \left\{ x : \sum_{k=1}^{x} P(T = k) < 1 - \alpha/2 \right\},$$

where the range of x is $1 \leq x \leq 2\min(n_0, n_1) + 1$. *The null hypothesis should be rejected when* $T < \tilde{t}_1$ *or* $T > \tilde{t}_2$.

Since the expressions for solving $\tilde{t}_1$ and $\tilde{t}_2$ are too complicated based on Theorem 2.2, any explicit formula of the critical values cannot be derived. Based on Theorem 2.1, we know that when $a \neq 0.5$, T does not converge to any smooth continuous distribution. In

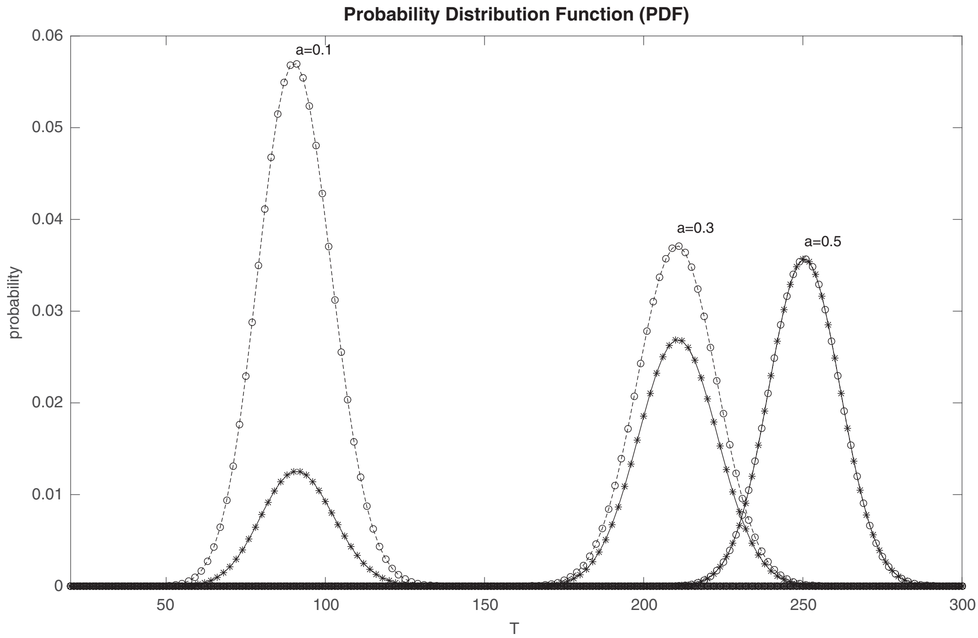

Figure 1. The probability value for T with $n = 500$, and the left to right three figures are corresponding to $a = 0.1, 0.3, 0.5$. The '$*$' and 'o' denote $P(T = 2k)$ $(k = 1, 2, \ldots, \lceil n/2 \rceil)$ and $P(T = 2k + 1)$ $(k = 0, 1, 2, \ldots, \lfloor n/2 \rfloor - 1)$, respectively.

order to study the asymptotical property of the statistics, we introduce two discrete random variables T_1 and T_2, whose probability is corresponding to the standardized probability of T being odd and even values. That is, T_1 is defined as

$$P(T_1 = 2k) = \frac{P(T = 2k)}{\theta_1}, \quad k = 1, 2, \ldots, \left\lfloor \frac{n}{2} \right\rfloor.$$

Here, $\lfloor x \rfloor$ stands for x round-down, and

$$\begin{aligned}
\theta_1 &= \sum_{k=1}^{\lfloor n/2 \rfloor} P\{T = 2k\} = \sum_{k=1}^{\lfloor n/2 \rfloor} \sum_{n_1=k}^{n-k} 2C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} a^{n_1} (1-a)^{n-n_1} \\
&= 2a(1-a) \left\{ \sum_{n_1=1}^{\lfloor n/2 \rfloor} \sum_{k=1}^{n_1} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} a^{n_1-1} (1-a)^{n-n_1-1} \right. \\
&\quad \left. + \sum_{n_1=\lfloor n/2 \rfloor+1}^{n-1} \sum_{k=1}^{n-n_1} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} a^{n_1-1} (1-a)^{n-n_1-1}] \right\} \\
&= 2a(1-a) \left\{ \sum_{n_1=1}^{\lfloor n/2 \rfloor} C_{n-2}^{n_1-1} a^{n_1-1} (1-a)^{n-n_1-1} + \sum_{n_1=\lfloor n/2 \rfloor+1}^{n-1} C_{n-2}^{n_1-1} a^{n_1-1} (1-a)^{n-n_1-1} \right\} \\
&= 2a(1-a) \qquad (5)
\end{aligned}$$

is the standardized value so that T_1 is a discrete variable. Another discrete variable T_2 is defined as

$$P(T_2 = 2k+1) = \frac{P(T = 2k+1)}{\theta_2}, \quad k = 0, 1, \ldots, \left\lceil \frac{n}{2} \right\rceil - 1.$$

Here, $\lceil x \rceil$ stand for x round-up, and $\theta_2 = \sum_{k=0}^{\lceil n/2 \rceil - 1} P\{T = 2k+1\}$ is a standardized value. Note that $\theta_1 = 2a(1-a)$ and $\theta_1 + \theta_2 = 1$, we have $\theta_2 = 1 - \theta_1 = a^2 + (1-a)^2$.

The means $E(T_1)$ and $E(T_2)$ can be derived as follows:

$$\mu_1 := E(T_1) = \sum_{k=1}^{\lfloor n/2 \rfloor} 2kP(T_1 = 2k) = 2(n-3)a(1-a) + 2, \tag{6}$$

$$\begin{aligned} \mu_2 := E(T_2) &= \sum_{k=0}^{\lceil n/2 \rceil - 1} (2k+1)P(T_2 = 2k+1) \\ &= (-4(n-3)a^4 + 8(n-3)a^3 - 6(n-3)a^2 \\ &\quad + 2(n-3)a + 1)/(a^2 + (1-a)^2). \end{aligned} \tag{7}$$

In the following, we give the derivation of the equations in Equations (6) and (7). Denote $f_k(a) := P\{T_1 = 2k\} = \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} a^{n_1-1}(1-a)^{n-n_1-1}$. By Taylor's expansions at $a = 0.5$, we have

$$P\{T_1 = 2k\} = \sum_{m=0}^{n-2} \cdot f_k^{(m)}(0.5) \frac{(a-0.5)^m}{m!}.$$

Hence,

$$E(T_1) = \sum_{k=1}^{\lfloor n/2 \rfloor} 2kP\{T_1 = 2k\} = \sum_{m=0}^{n-2} \sum_{k=1}^{\lfloor n/2 \rfloor} 2k \cdot f_k^{(m)}(0.5) \frac{(a-0.5)^m}{m!}.$$

By complicated derivation, we have $\sum_{k=1}^{\lfloor n/2 \rfloor} 2k \cdot f_k^{(m)}(0.5)((a-0.5)^m/m!) = 0$ for $m = 1$ and $m > 2$. According to the fact that

$$\begin{aligned} f_k^{(0)}(0.5) &= C_{n-1}^{2k-1} 0.5^{n-2}, \\ f_k^{(2)}(0.5) &= \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} 0.5^{n-2} \cdot [4(n^2 - n + 2) - 16nn_1 + 16n_1^2], \end{aligned}$$

and Lemma 5.1 in Section 5, we have

$$\begin{aligned} E(T_1) &= \sum_{k=1}^{\lfloor n/2 \rfloor} 2k \cdot f_k^{(0)}(0.5) + \sum_{k=1}^{\lfloor n/2 \rfloor} 2k \cdot \frac{f_k^{(2)}(0.5)}{2!}(a-0.5)^2 \\ &= 0.5(n+1) - 2(n-3)(a-0.5)^2 \\ &= 2(n-3)a(1-a) + 2. \end{aligned}$$

As for $E(T_2)$ in Equation (7), according to the expression of $E(T)$ in Theorem 2.1 and $E(T) = E(T_1)\cdot\theta_1 + E(T_2)\cdot\theta_2$, we have $E(T_2) = (-4(n-3)a^4 + 8(n-3)a^3 - 6(n-3)a^2 + 2(n-3)a + 1)/(a^2 + (1-a)^2)$.

For the variances $\mathrm{Var}(T_1)$ and $\mathrm{Var}(T_2)$, note that

$$\begin{aligned}
E(T_1^2) &= \sum_{k=1}^{\lfloor n/2 \rfloor} (2k)^2 P(T_1 = 2k) \\
&= ((8a + (24n-80)a^2 + (8n^2 - 120n + 304)a^3 \\
&\quad + (-24n^2 + 240n - 552)a^4 + (24n^2 - 216n + 480)a^5 \\
&\quad + (-8n^2 + 72n - 160)a^6))/(2a(1-a)), \\
E(T_2^2) &= \sum_{k=0}^{\lceil n/2 \rceil - 1} (2k+1)^2 P(T_2 = 2k+1) \\
&= (1 + (8n-18)a + (4n^2 - 52n + 114)a^2 + (-16n^2 + 160n - 352)a^3 \\
&\quad + (28n^2 - 260n + 576)a^4 + (-24n^2 + 216n - 480)a^5 \\
&\quad + (8n^2 - 72n + 160)a^6)/(a^2 + (1-a)^2),
\end{aligned}$$

we have

$$\begin{aligned}
\sigma_1^2 &:= \mathrm{Var}(T_1) = E(T_1^2) - (E(T_1))^2, \\
\sigma_2^2 &:= \mathrm{Var}(T_2) = E(T_2^2) - (E(T_2))^2.
\end{aligned}$$

The theorem below illustrates the properties of the random variables T_1, T_2 under the null hypothesis. Based on the results, we obtain the limiting critical values of the test statistics.

Theorem 2.3: *Under the null hypothesis, variables T_1 and T_2 obey the normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Asymptotically, we have $P(t_1 < T < t_2) = \alpha$ for $t_1 \approx -\sigma z_{\alpha/2} + \mu$ and $t_2 \approx \sigma z_{\alpha/2} + \mu$, where $z_{\alpha/2}$ is $\alpha/2$ quantile of the standard normal distribution, $\mu = E(T)$ and $\sigma = (\mathrm{Var}(T))^{1/2}$.*

For an intuitive expression, under the condition of $n = 500$ and $a = 0.1, 0.3, 0.5$, we can plot the histograms of the random variables T_1, T_2 and corresponding normal density function $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ in Figure 2. It is shown that the normal density function fits the histograms very well.

2.2. The properties of statistics under the alternative hypothesis

We assume the alternative hypothesis is that H_1, the random sample sequence $r_1, r_2, \ldots, r_n$, follows any distribution on $(0, 1)$ except the standard uniform distribution. Let $p = P(u_i < 0)$, with $u_i = r_i - a$. Then, the corresponding properties can be similarly derived. Under

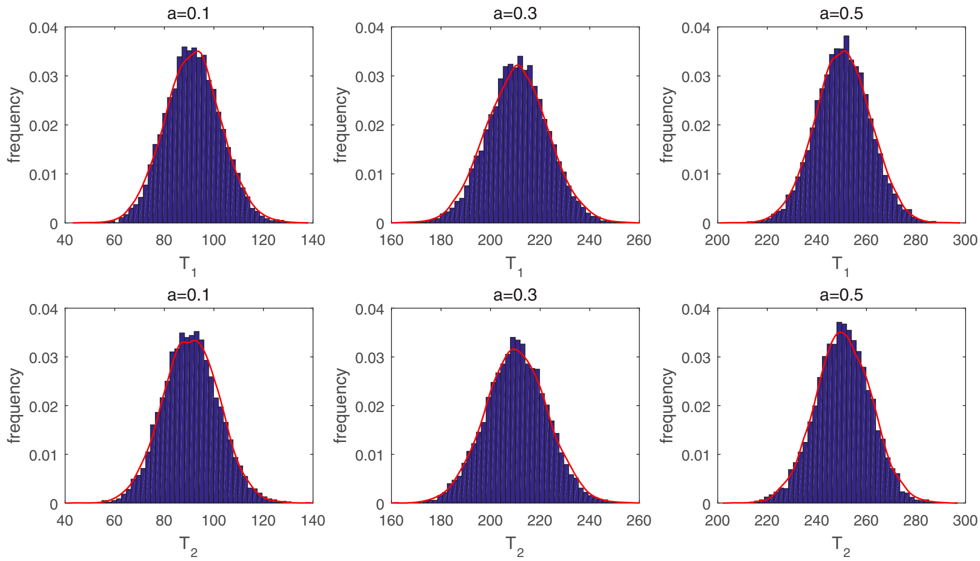

Figure 2. When $a = 0.1, 0.3, 0.5$, the probability density functions of T_1, T_2 ($n = 500$).

the alternative hypothesis, we have

$$\begin{aligned}
\tilde{\mu}_1 &:= E(T_1) \\
&= \frac{4(n-3)p^4 - 8(n-3)p^3 + 4(n-4)p^2 + 4p}{2p(1-p)}, \\
\tilde{\mu}_2 &:= E(T_2) \\
&= \frac{-4(n-3)p^4 + 8(n-3)p^3 - 6(n-3)p^2 + 2(n-3)p + 1}{p^2 + (1-p)^2}, \\
\tilde{\sigma}_1^2 &:= \mathrm{Var}(T_1) \\
&= \frac{8p + (24n-80)p^2 + (8n^2 - 120n + 304)p^3}{2p(1-p)} \\
&\quad + \frac{(-24n^2 + 240n - 552)p^4 + (24n^2 - 216n + 480)p^5 + (-8n^2 + 72n - 160)p^6}{2p(1-p)} \\
&\quad - \left(\frac{4(n-3)p^4 - 8(n-3)p^3 + 4(n-4)p^2 + 4p}{2p(1-p)}\right)^2, \\
\tilde{\sigma}_2^2 &:= \mathrm{Var}(T_2) \\
&= \frac{1 + (8n-18)p + (4n^2 - 52n + 114)p^2 + (-16n^2 + 160n - 352)p^3}{p^2 + (1-p)^2} \\
&\quad + \frac{(28n^2 - 260n + 576)p^4 + (-24n^2 + 216n - 480)p^5 + (8n^2 - 72n + 160)p^6}{p^2 + (1-p)^2} \\
&\quad - \left(\frac{-4(n-3)p^4 + 8(n-3)p^3 - 6(n-3)p^2 + 2(n-3)p + 1}{p^2 + (1-p)^2}\right)^2.
\end{aligned}$$

In the following theorem, we derive the asymptotic properties of the test statistics under the alternative hypothesis.

Theorem 2.4: *Under the alternative hypothesis that the random sample sequence* $r_1, r_2, \ldots, r_n$ *does not follow the standard uniform distribution, variables* T_1 *and* T_2 *obey asymptotically the normal distributions* $N(\tilde{\mu}_1, \tilde{\sigma}_1^2)$ *and* $N(\tilde{\mu}_2, \tilde{\sigma}_2^2)$*, respectively. Based on the results, the asymptotical power of the statistics* T *is*

$$Q(a) = 1 + \Phi\left(\frac{-\sigma z_{\alpha/2} + \mu - \tilde{\mu}_1}{\tilde{\sigma}_1}\right) \cdot \tilde{\theta}_1 + \Phi\left(\frac{-\sigma z_{\alpha/2} + \mu - \tilde{\mu}_2}{\tilde{\sigma}_2}\right) \cdot \tilde{\theta}_2 - \Phi\left(\frac{\sigma z_{\alpha/2} + \mu - \tilde{\mu}_1}{\tilde{\sigma}_1}\right) \cdot \tilde{\theta}_1 - \Phi\left(\frac{\sigma z_{\alpha/2} + \mu - \tilde{\mu}_2}{\tilde{\sigma}_2}\right) \cdot \tilde{\theta}_2. \tag{8}$$

Here, $\tilde{\theta}_1 = 2p(1-p)$*,* $\tilde{\theta}_2 = p^2 + (1-p)^2$*, and* $\Phi(\cdot)$ *stands for the standard normal distribution function.*

According to Theorem 2.4, one option is to select an optimal a that maximizes the function $Q(a)$ in Equation (8). Hence, we denote the optimal a as $\tilde{a}$, which is obtained by

$$\tilde{a} = \max_a Q(a). \tag{9}$$

Since the parameters $\tilde{\mu}_1$, $\tilde{\mu}_2$, $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ in Equation (9) contain the unknown parameter p, when a is given, we use the estimator $\hat{p} = n^{-1} \sum_{j=1}^{n} I(u_j < 0)$ of p to get the corresponding estimators of $\tilde{\mu}_1$, $\tilde{\mu}_2$, $\tilde{\sigma}_1$, and $\tilde{\sigma}_2$. Here, $I(\cdot)$ is the indicative function.

To determine the optimal a, first we consider obtaining the estimate of $\tilde{a}$ using the following Newton–Raphson algorithm:

$$a_{i+1} = a_i - (Q''(a_i))^{-1} Q'(a_i). \tag{10}$$

Here, $Q'(a)$ is the first derivative of $Q(a)$, and $Q''(a)$ is the second derivative of $Q(a)$. Theoretically, the parameter p is a function of a and p is contained in $Q(a)$. Hence, in order to obtain the derivatives $Q'(a)$ and $Q''(a)$, an explicit form of p as a function of a is needed. But in practice, according to the definition of p, $p = P(r_i < a)$, it is not easy to derive the explicit form of p, since the explicit form of the random sample sequence's distribution under the alternative hypothesis is unknown. Therefore, the Newton–Raphson algorithm does not work here. Instead, we estimate the parameters using $p = n^{-1} \sum_{j=1}^{n} I(u_j < 0)$ and use the grid method to obtain the optimal a from the definition of $Q(a)$ in Equation (9).

2.3. The null hypothesis is any other specific distribution

In this section, we extend the statistics to examine whether the random sample sequence $r_1, r_2, \ldots, r_n$ follows an arbitrary distribution. In other words, the null hypothesis is that H_0: the random sample sequence $r_1, r_2, \ldots, r_n$ follows the distribution with density function $g(x)$ on (x_1, x_2), versus the alternative hypothesis H_1, which states that the random sample sequence $r_1, r_2, \ldots, r_n$ does not follow the density function $g(x)$ on (x_1, x_2). Here, x_1, x_2 can be $-\infty$ or $+\infty$. Without loss of generality, we also let $u_i = r_i - a, a \in (x_1, x_2)$ and S be the total number of runs. It can be derived similarly as in the case of the uniform distribution that the asymptotical rejection region is $W_S = \{S : S < s_1, \text{ or } S > s_2\}$, where

$$s_1 = -\sigma_s z_{\alpha/2} + \mu_s, \quad s_2 = \sigma_s z_{\alpha/2} + \mu_s.$$

Here,

$$\mu_s := E(S) = 2G_0(a)(1 - G_0(a))(n-1) + 1,$$
$$\sigma_s^2 := \mathrm{Var}(S) = 4(5-3n)G_0(a)^2(1-G_0(a))^2 + 2(2n-3)G_0(a)(1-G_0(a)),$$

where $n_0 = \sum_{i=1}^n I(u_i \geq 0)$, $n_1 = \sum_{i=1}^n I(u_i < 0) = n - n_0$, and $G_0(a) = P(u_i \leq 0)$ under the null hypothesis. As for the derivation of μ_s and σ_s^2, Theorem 2.1 shows the expectation and variance of T when the null hypothesis is uniform distribution on $[0,1]$, in which a is the probability for $u_i < 0$. When the null hypothesis is any other specific distribution, the probability for $u_i < 0$ is $G_0(a)$. Under H_0, we replace a as $G_0(a)$ and obtain the corresponding μ_s and σ_s^2 by following the proof under null hypothesis.

Similar to the case of the uniform distribution, we define two random variables S_1 and S_2 with probability $P(S_1 = 2k) = P(S = 2k)/\tilde{\theta}_{s1}$ for $k = 1, 2, \ldots, \lfloor n/2 \rfloor$ and $P(S_2 = 2k+1) = P(S = 2k+1)/\tilde{\theta}_{s2}$ for $k = 0, 1, \ldots, \lceil n/2 \rceil - 1$. Denote $\tilde{\mu}_{s1} = E(S_1)$, $\tilde{\mu}_{s2} = E(S_1)$, $\tilde{\sigma}_{s1}^2 = \mathrm{Var}(S_1)$ and $\tilde{\sigma}_{s2}^2 = \mathrm{Var}(S_2)$ under the alternative hypothesis. It can also be derived that the power under the alternative hypothesis is

$$\begin{aligned} Q(a) = 1 &+ \Phi\left(\frac{-\sigma_s z_{\alpha/2} + \mu_s - \tilde{\mu}_{s1}}{\tilde{\sigma}_{s1}}\right) \cdot \tilde{\theta}_{s1} + \Phi\left(\frac{-\sigma_s z_{\alpha/2} + \mu_s - \tilde{\mu}_{s2}}{\tilde{\sigma}_{s2}}\right) \cdot \tilde{\theta}_{s2} \\ &- \Phi\left(\frac{\sigma_s z_{\alpha/2} + \mu_s - \tilde{\mu}_{s1}}{\tilde{\sigma}_{s1}}\right) \cdot \tilde{\theta}_{s1} - \Phi\left(\frac{\sigma_s z_{\alpha/2} + \mu_s - \tilde{\mu}_{s2}}{\tilde{\sigma}_{s2}}\right) \cdot \tilde{\theta}_{s2}. \end{aligned} \tag{11}$$

Here, $\tilde{\theta}_{s1} = 2p(1-p)$, $\tilde{\theta}_{s2} = p^2 + (1-p)^2$.

As for obtaining the maximum $Q(a)$ in Equation (11), if the explicit form of the random sample sequence's distribution denoted as $G_1(\cdot)$ under the alternative hypothesis is known, we can use the theoretical optimal cut point $p = G_1(a)$. Otherwise, we use the estimator of p, $\hat{p} = n^{-1}\sum_{j=1}^n I(u_j < 0)$, instead. In this paper, the grid method is applied to obtain the optimal a, since the alternative hypothesis is unknown in most cases.

3. Numerical analysis

In this section, we mainly use simulation analysis to study the efficiency of the proposed method. Different examples are investigated here to illustrate the efficiency of the proposed test.

Example 3.1: The null hypothesis is H_0 and reads as follows. The random sample sequence $r_1, r_2, \ldots, r_n$ follows the standard uniform distribution. The alternative hypothesis is H_1. It states that the random sample sequence $r_1, r_2, \ldots, r_n$ follows the Beta(2,2) distribution on $(0,1)$. Also, the size of the proposed test is computed.

First, we plot the function $Q(a)$ in Equation (8) to illustrate the consistency of $Q(a)$ and its estimators in Figure 3 with sample size $n = 50$ and 100. The figures for (a), (b), and (c)

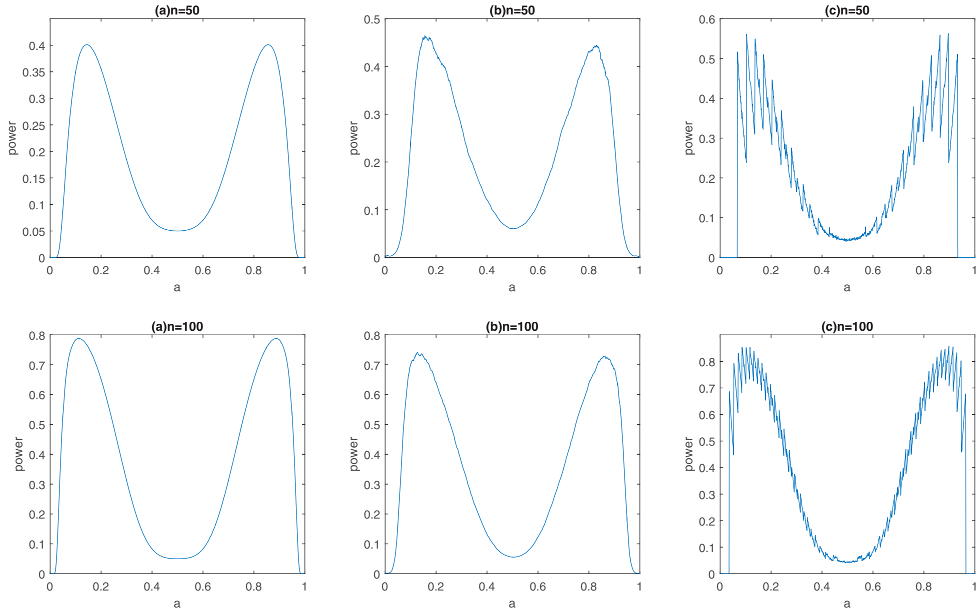

Figure 3. When $n = 50,100$, (a) the function $Q(a)$ with $p = P(r < a)$; (b) the function $Q(a)$ with the estimator $\hat{p} = n^{-1} \sum_{i=1}^{n} I(r_i < a)$; and (c) the practical power computed by rejection regions.

correspond to the function $Q(a)$ with $p = P(r < a)$, $\hat{p} = n^{-1} \sum_{i=1}^{n} I(r_i < a)$, and that the $\hat{Q}(a)$ estimated by $\hat{Q}(a) = \hat{P}(T < \hat{t}_1$, or $T > \hat{t}_2)$. Here, $\hat{t}_1$, $\hat{t}_2$, and $\hat{P}(\cdot)$ are the estimators of t_1, t_2, and $P(\cdot)$, respectively. As shown in Figure 3, the estimators of $Q(a)$ can approximate the true function very well. Therefore, it is reasonable to estimate the optimal $\tilde{a}$ based on the estimator of $Q(a)$ with the estimated p. The larger the sample size is, the more efficient the estimators $\hat{Q}(a)$ will be. In addition, since $P(r_i < a) + P(r_i < 1 - a) = 1$, the test power is symmetrical with respect to $a = 0.5$, which means that the optimal a may not be unique. The powers are the same for any optimal a if it is not unique.

Table 1 presents type I error for the uniform distribution over (0, 1), from which we can see that the size of the proposed test satisfies the nominal level constraint. Table 2 presents the simulation results with different sample sizes for $n = 20,30,50,100,200,500$, and 1000. The corresponding test powers of classical positive and negative runs tests with $a = 0.5$ and the K–S test are shown for comparison. From Table 2, we can see that the power is more efficient for the proposed method. However, the classical positive and negative runs

Table 1. Comparison of type I error ($H_0 : U(0, 1)$, $H_1 : U(0, 1)$).

n	Adaptive test	Classical test	K–S test	Optimal a
20	0.0518	0.0607	0.0524	0.1940
30	0.0503	0.0570	0.0507	0.1870
50	0.0527	0.0467	0.0501	0.1080
100	0.0453	0.0470	0.0508	0.1060
200	0.0545	0.0464	0.0504	0.0350
500	0.0513	0.0472	0.0522	0.0220
1000	0.0443	0.0520	0.0531	0.0260

Table 2. Comparison of power ($H_0 : U(0,1), H_1 : Be(2,2)$).

n	Adaptive test	Classical test	K–S test	Optimal a
20	0.1493	0.0657	0.0642	0.2520
30	0.2455	0.0587	0.1079	0.2330
50	0.3120	0.0450	0.2232	0.1690
100	0.8299	0.0467	0.6249	0.1340
200	0.9859	0.0464	0.9772	0.1060
500	1.0000	0.0460	1.0000	0.0820
1000	1.0000	0.0520	1.0000	0.0840

test cannot detect the alternative hypothesis due to the fact that the density function of $Be(2,2)$ is symmetrical about 0.5. The power of the classical tests is only about 0.05 under the alternative hypothesis, which is much worse than that of the adaptive tests. With the same sample, the power of the proposed test is more efficient than K–S test as well.

Example 3.2: The null hypothesis is H_0 and reads as follows. The random sample sequence $r_1, r_2, \ldots, r_n$ follows the standard uniform distribution. The alternative hypothesis is H_1. It states that the random sample sequence $r_1, r_2, \ldots, r_n$ follows the distribution with density function $f(x) = (\pi/2)\cos(\pi/2)x,\ x \in (0,1)$.

We use this example to illustrate the power performance when the density function under the alternative hypothesis is a cosine function, which means that the density is a frequency function. The results are presented in Table 3, from which we can see that the trend of the simulation results is very similar to that of Example 3.1, and the adaptive test is more efficient than classical test and K–S test under all circumstances.

Example 3.3: The null hypothesis is H_0 and reads as follows. The random sample sequence $r_1, r_2, \ldots, r_n$ follows the standard normal distribution $N(0,1)$. The alternative hypothesis is H_1. It states that the random sample sequence $r_1, r_2, \ldots, r_n$ follows the normal distribution $N(\mu, 1)(\mu \neq 0)$.

In this example, the density function under the null hypothesis is the normal distribution $N(0,1)$ with domain $(-\infty, +\infty)$ instead of $(0,1)$. We fix $n = 30$ and gradually increase μ to pull it away from the null hypothesis. Then, we compute the test power. Note that the classical test is proposed for the null hypothesis when the random sample sequence is from the uniform distribution on $(0,1)$. Thus, we compare the adaptive test proposed in this paper to the classical test with the fixed cut point $a = 0.5$. As the simulation result

Table 3. Comparison of power ($H_0 : U(0,1),\ H_1 : f(x) = (\pi/2)\cos(\pi/2)x$).

n	Adaptive test	Classical test	K–S test	Optimal a
20	0.6429	0.1059	0.5097	0.9740
30	0.8196	0.1281	0.6739	0.9710
50	0.9298	0.1306	0.8832	0.9550
100	0.9954	0.2033	0.9945	0.9540
200	1.0000	0.3502	1.0000	0.9380
500	1.0000	0.6712	1.0000	0.9220
1000	1.0000	0.9258	1.0000	0.9170

Table 4. Comparison of power ($n = 30$) ($H_0 : N(0, 1), H_1 : N(\mu, 1)$).

μ	Adaptive test	Classical test	K–S test	Optimal a
0.1	0.0697	0.0434	0.0670	1.5090
0.2	0.1156	0.0506	0.1587	1.4120
0.3	0.3276	0.0532	0.2876	1.5890
0.4	0.4094	0.0658	0.4701	1.4080
0.5	0.4128	0.0716	0.6537	1.4750
0.6	0.5306	0.0634	0.8107	1.4940
0.7	0.6435	0.0514	0.9132	1.4900
0.8	0.7780	0.0485	0.9661	1.5990
0.9	0.8092	0.0477	0.9894	1.5180
1.0	0.8853	0.0465	0.9983	1.5290

shows in Table 4, we find that as μ moves from 0.0 to 1.0, the power of the classical test increases first and then decreases, reaching its maximum value at $\mu = 0.5$. However, the maximum test power is only 0.0716, which is clearly not ideal. In addition, it is obvious that the power of the classical test is almost symmetrical with respect to $\mu = 0.5$, due to the fact that when $a = 0.5$, $p = P(r_i < a)$ and $1-p$ are the same with respect to $\mu = 0.5$. Thus, only when $\mu = 0.5$ and $p = 1-p = 0.5$, the random sample sequence is divided most effectively, and the total number of runs T and the test power reach their maximum. However, the adaptive test solves this problem by using the flexible a. As μ grows larger, which means that the alternative hypothesis departs from the null hypothesis gradually, the test power becomes larger and is always better than that of the corresponding classical test with the same μ. Under this circumstance, the proposed test is less efficient than K–S test.

Example 3.4: The null hypothesis is H_0 and reads as follows. The random sample $r_1, r_2, \ldots, r_n$ is from the chi-square distribution $\chi^2(1)$. The alternative hypothesis is H_1. It states that the random sample $r_1, r_2, \ldots, r_n$ is from the truncated normal distribution with density function $f(x) = \phi(x - \mu)/(1 - \Phi(-\mu))(x > 0)$.

The domains of the distributions under the null and alternative hypotheses are both $(0, +\infty)$, which are different from previous examples. In this example, for the fixed n, we increase μ to pull it away from the null hypothesis, and then compute the test power. According to the simulation results presented in Table 5, we find that as μ moves from 0.0 to 1.0, both the power of the classical test with $a = 0.5$ and that of the adaptive test increase.

Table 5. Comparison of Power ($H_0 : \chi^2(1), H_1 : N(\mu, 1)(x \geq 0)$).

	$n = 30$					$n = 100$			
μ	Adaptive test	Classical test	K–S test	Optimal a	μ	Adaptive test	Classical test	K–S test	Optimal a
0.0	0.5647	0.0883	0.5208	0.1430	0.0	0.9923	0.0854	0.9970	0.0700
0.1	0.5989	0.0912	0.5961	0.1530	0.1	0.9951	0.1120	0.9990	0.0810
0.2	0.6388	0.0998	0.6732	0.1690	0.2	0.9986	0.1674	0.9998	0.0760
0.3	0.6891	0.1403	0.7481	0.1700	0.3	0.9988	0.2523	0.9998	0.0990
0.4	0.7158	0.1885	0.8248	0.1900	0.4	0.9994	0.3624	1.0000	0.1020
0.5	0.7592	0.2633	0.8721	0.1980	0.5	0.9996	0.4843	1.0000	0.1110
0.6	0.7707	0.2831	0.9195	0.2060	0.6	0.9998	0.6021	1.0000	0.1190
0.7	0.8752	0.3504	0.9464	0.2120	0.7	0.9999	0.7361	1.0000	0.1230
0.8	0.8859	0.4421	0.9709	0.2540	0.8	1.0000	0.8386	1.0000	0.1630
0.9	0.9073	0.4909	0.9859	0.2690	0.9	1.0000	0.9215	1.0000	0.1680
1.0	0.9432	0.5912	0.9916	0.2550	1.0	1.0000	0.9617	1.0000	0.1700

The adaptive test can always reject the null hypothesis more effectively than the classical test under all circumstances. It can be also seen that when the sample size n is small, the power of the proposed test is less efficient than K–S test. But the powers of the proposed test are almost as efficient as K–S test with large sample size.

According to the simulation study results described, we believe that the performance of the adaptive test is more powerful than that of the classical test under all of the null and alternative hypothesis scenarios we tested. The adaptive test can quickly detect the differences when the alternative hypothesis departs from the null hypothesis. In addition, regarding testing for the uniform distribution over (0, 1), with the same sample, the power of the proposed test is significantly more efficient than K-S test.

4. Conclusion

Motivated by the idea of optimizing the cut point a, this paper improves the classical positive and negative runs test. The adaptive positive and negative runs test relaxes the restriction of the cut point being fixed at 0.5 in the classical method. The idea is to divide the sample sequence $\{r_i\}$ using an uncertain cut point a and obtain the corresponding statistics T first. Additionally, by defining two new, random variables, this paper derives and demonstrates the exact and asymptotic critical values of the rejection region of the statistics T. Eventually, the theoretical optimal cut point a is selected by maximizing the asymptotical power. However, note that the alternative hypothesis is unknown in most cases, so we estimate the cut point $a(a \in (0, 1))$ using the grid method to let a approach the theoretically optimal value, which leads the actual test power to approach the theoretical maximum value. Numerical study has thus verified the efficiency of the adaptive positive and negative runs test.

5. Proof of the theorems

In order to prove the theorems, we first introduce a lemma.

Lemma 5.1: *This lemma shows the property of* $\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^m$ *for* $m = 0, 1, \ldots, n-2$. *It can be derived that*

$$\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} = C_{n-1}^{2k-1}, \tag{12}$$

$$\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1 = \frac{n}{2} C_{n-1}^{2k-1}, \tag{13}$$

$$\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^2 = \left(-\frac{k}{2(2k+1)} n + \frac{k+1}{2(2k+1)} n^2\right) C_{n-1}^{2k-1}, \tag{14}$$

$$\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^3 = \left(-\frac{3k}{4(2k+1)} n^2 + \frac{k+2}{4(2k+1)} n^3\right) C_{n-1}^{2k-1}, \tag{15}$$

$$\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^4$$
$$= \frac{(k+2)(k+3)n^4 - 6k(k+2)n^3 + (3k-1)kn^2 + 2k^2 n}{4(2k+1)(2k+3)} C_{n-1}^{2k-1}, \tag{16}$$

$$\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^m$$
$$= \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n n_1^{m-1} - k^2 \sum_{n_1=k}^{n-k} C_{n1}^{k} C_{n-n_1}^{k} (nn_1 + n + 1) n_1^{m-4}$$
$$+ k^2(k+1)^2 \sum_{n_1=k}^{n-k} C_{n1+1}^{k+1} C_{n-n_1+1}^{k+1} n_1^{m-4}. \tag{17}$$

Based on the above equations, the sum notation in $\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^m$ can be simplified for any m.

Proof for Lemma* 5.1*: For proving Equation (12), let X_1, Y_1 be independent Pascal random variables with $X_1, Y_1 \sim \text{Pascal}(k, 0.5)$, we have $Z_1 = X_1 + Y_1 \sim \text{Pascal}(2k, 0.5)$. That is

$$P(Z_1 = n) = C_{n-1}^{2k-1} 0.5^{2k} 0.5^{n-2k}. \tag{18}$$

On the other hand, based on the convolution formula, it can be derived that

$$\begin{aligned} P(Z_1 = n) &= P\{X_1 + Y_1 = n\} \\ &= \sum_{n_1=k}^{n-k} P\{X_1 = n_1\} P\{Y_1 = n - n_1\} \\ &= \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} 0.5^{2k} 0.5^{n-2k}. \end{aligned} \tag{19}$$

Hence, we have $\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} = C_{n-1}^{2k-1}$ based on Equations (18) and (19).

For proving Equation (13), it can be derived that

$$\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1 = \frac{n}{2} \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} = \frac{n}{2} C_{n-1}^{2k-1}.$$

As for Equation (14), based on Equation (13), it can be proved that

$$\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^2$$
$$= \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n n_1 - \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1 (n - n_1)$$

$$
\begin{aligned}
&= \frac{n^2}{2} C_{n-1}^{2k-1} - k^2 \sum_{n_1=k}^{n-k} C_{n_1}^{k} C_{n-n_1}^{k} \\
&= \frac{n^2}{2} C_{n-1}^{2k-1} - \frac{kn(n+1)}{2(2k+1)} C_{n-1}^{2k-1} \\
&= \left(-\frac{k}{2(2k+1)} n + \frac{k+1}{2(2k+1)} n^2 \right) C_{n-1}^{2k-1}.
\end{aligned}
$$

As for Equations (15) and (16), we can derive it by following the proof for Equation (14). For Equation (17), we have

$$
\begin{aligned}
&\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^m \\
&\quad = \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n n_1^{m-1} - \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^{m-1} (n - n_1).
\end{aligned}
$$

Since

$$
\begin{aligned}
&\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^{m-1} (n - n_1) \\
&\quad = k^2 \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} n_1^{m-2} \\
&\quad = k^2 \left(\sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} (n n_1 + n + 1) n_1^{m-4} \right. \\
&\qquad \left. - \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} (n_1 + 1)(n - n_1 + 1) n_1^{m-4} \right) \\
&\quad = k^2 \sum_{n_1=k}^{n-k} C_{n1}^{k} C_{n-n_1}^{k} (n n_1 + n + 1) n_1^{m-4} - k^2 (k+1)^2 \sum_{n_1=k}^{n-k} C_{n1+1}^{k+1} C_{n-n_1+1}^{k+1} n_1^{m-4},
\end{aligned}
$$

Equation (17) is proved. ∎

Proof for Theorem* 2.1*: Based on Equations (3) and (4), we have

$$
E(T \mid n_1) = \sum_{m=1}^{n} m P(T = m \mid n_1) = \frac{2n_1(n - n_1)}{n} + 1, \tag{20}
$$

$$E(T^2 \mid n_1) = \sum_{m=1}^{n} m^2 P(T = m \mid n_1)$$
$$= \frac{4n_1(n-n_1)(n_1-1)(n-n_1-1) + 2n_1(n-n_1)(4n-5)}{n(n-1)} + 1. \tag{21}$$

Below, we give the proof for $E(T \mid n_1)$ in Equation (20). Evidently, $E(T \mid n_1) = 1$ when $n_1 = 0$. Under the condition that $0 < n_1 < n_0$, it can be derived that

$$E(T \mid n_1) = \sum_{k=1}^{n_1} 2k \cdot P\{T = 2k \mid n_1\} + \sum_{k=1}^{n_1} (2k+1) \cdot P\{T = 2k+1 \mid n_1\}$$
$$=: E_1 + E_2. \tag{22}$$

According to the definition of $P\{T = 2k \mid n_1\}$ in Equation (1), note that $(k-1)C_{n_1-1}^{k-1} = (n_1-1)C_{n_1-2}^{k-2}$, $\sum_{k=2}^{n_1} C_{n_0-1}^{k-1} C_{n_1-2}^{k-2} = C_{n_1+n_0-3}^{n_0-2}$ and $\sum_{k=1}^{n_1} C_{n_0-1}^{k-1} C_{n_1-1}^{k-1} = C_{n_0+n_1-2}^{n_0-1}$, we have

$$E_1 = \sum_{k=1}^{n_1} 4(k-1) \cdot \frac{C_{n_0-1}^{k-1} C_{n_1-1}^{k-1}}{C_n^{n_1}} + \sum_{k=1}^{n_1} 4 \cdot \frac{C_{n_0-1}^{k-1} C_{n_1-1}^{k-1}}{C_n^{n_1}}$$
$$= \frac{4(n_1-1) \cdot C_{n-3}^{n_0-2} + 4C_{n-2}^{n_0-1}}{C_n^{n_1}}. \tag{23}$$

As for the expression E_2, based on the equations $\sum_{k=1}^{n_1-1} k \cdot C_{n_0-1}^{k-1} C_{n_1-1}^{k} = (n_1 - 1)C_{n_0+n_1-3}^{n_0-1}$, $\sum_{k=1}^{n_1} k \cdot C_{n_0-1}^{k} C_{n_1-1}^{k-1} = (n_0-1)C_{n_0+n_1-3}^{n_1-1}$, $\sum_{k=1}^{n_1-1} C_{n_0-1}^{k-1} C_{n_1-1}^{k} = C_{n_0+n_1-2}^{n_0}$ and $\sum_{k=1}^{n_1} C_{n_0-1}^{k} C_{n_1-1}^{k-1} = C_{n_0+n_1-2}^{n_1}$, we have the following result by recalling the definition of $P\{T = 2k+1 \mid n_1\}$ in Equation (2):

$$E_2 = \sum_{k=1}^{n_1-1} 2k \cdot \frac{C_{n_0-1}^{k-1} C_{n_1-1}^{k} + C_{n_0-1}^{k} C_{n_1-1}^{k-1}}{C_n^{n_1}} + \sum_{k=1}^{n_1-1} \frac{C_{n_0-1}^{k-1} C_{n_1-1}^{k} + C_{n_0-1}^{k} C_{n_1-1}^{k-1}}{C_n^{n_1}}$$
$$+ (2n_1+1) \frac{C_{n_0-1}^{n_1} C_{n_1-1}^{n_1-1}}{C_n^{n_1}}$$
$$= \frac{2(n_1-1) \cdot C_{n-3}^{n_0-1} + 2(n_0-1)C_{n-3}^{n_1-1} + C_{n-2}^{n_1-2} + C_{n-2}^{n_0-2}}{C_n^{n_1}}. \tag{24}$$

Hence, according to expressions (22), (23) and (24), we have

$$E(T \mid n_1) = \frac{2n_1(n-n_1)}{n} + 1. \tag{25}$$

As for $n_1 = n_0$, we have

$$E(T \mid n_1) = \sum_{k=1}^{n_1} 2k \cdot P\{T = 2k \mid n_1\} + \sum_{k=2}^{n_1} (2k-1) \cdot P\{T = 2k-1 \mid n_1\}.$$

The same result can be derived as that in Equation (25) by a similar way. Under the condition that $0 \le n_1 \le n_0$, note that the symmetrical property of n_0 and n_1 in Equation

(22), hence Equation (25) holds for any n_0 and n_1. The $E(T^2 \mid n_1)$ in Equation (21) can be similarly derived. After a complicated calculation, we have

$$\begin{aligned} E(T^2 \mid n_1) &= \sum_{k=1}^{n_1}(2k)^2 \cdot P\{T = 2k \mid n_1\} + \sum_{k=1}^{n_1}(2k+1)^2 \cdot P\{T = 2k+1 \mid n_1\} \\ &= \frac{4n_1(n-n_1)(n_1-1)(n-n_1-1) + 2n_1(n-n_1)(4n-5)}{n(n-1)} + 1. \end{aligned}$$

To compute $E(T)$ and $\mathrm{Var}(T)$, we study the properties of the random variable n_1's the first to fourth moments. Since $n_1 = \sum_{i=1}^{n} I(u_i < 0)$, which follows binomial distribution $B(n, a)$, we have

$$\begin{aligned} E(n_1) &= na, \\ E(n_1^2) &= na + (n^2 - n)a^2, \\ E(n_1^3) &= na + 3n(n-1)a^2 + n(n-1)(n-2)a^3, \\ E(n_1^4) &= na + 7n(n-1)a^2 + 6n(n-1)(n-2)a^3 + n(n-1)(n-2)(n-3)a^4. \end{aligned}$$

Based on the results above and that in Equation (20), we have

$$\begin{aligned} E(T) &= E(E(T \mid n_1)) = E\left(\frac{2n_1(n-n_1)}{n} + 1\right) \\ &= 2a(1-a)(n-1) + 1, \\ E(T^2) &= E(E(T^2 \mid n_1)) \\ &= E\left(\frac{4n_1(n-n_1)(n_1-1)(n-n_1-1) + 2n_1(n-n_1)(4n-5)}{n(n-1)} + 1\right) \\ &= 4(n-2)(n-3)a^2(1-a)^2 + 2(4n-5)a(1-a) + 1. \end{aligned}$$

As a result,

$$\begin{aligned} \mathrm{Var}(T) &= E(T^2) - (E(T))^2 \\ &= 4(n-2)(n-3)a^2(1-a)^2 + 2(4n-5)a(1-a) \\ &\quad + 1 - (2a(1-a)(n-1) + 1)^2 \\ &= 4(5-3n)a^2(1-a)^2 + 2(2n-3)a(1-a). \end{aligned}$$

Below, we verify the asymptotical properties of the statistics T with $a = 0.5$. Recall the definition of the distribution of T in Equations (3) and (4).

Thus, based on Equation (12) in Lemma 5.1, when $a = 0.5$, we have

$$P\{T = 2k\} = \sum_{n_1=k}^{n-k} 2C_{n_1-1}^{k-1}C_{n-n_1-1}^{k-1}0.5^{n_1}(1-0.5)^{n-n_1} = C_{n-1}^{2k-1}0.5^{n-1},$$

and similarly,

$$
\begin{aligned}
&P\{T = 2k + 1\} \\
&\quad = \sum_{n_1=k+1}^{n-k} C_{n_1-1}^{k} C_{n-n_1-1}^{k-1} 0.5^{n_1} 0.5^{n-n_1} + \sum_{n_1=k}^{n-k-1} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k} 0.5^{n_1} 0.5^{n-n_1} \\
&\quad = C_{n-1}^{2k} 0.5^{n-1}.
\end{aligned}
$$

Hence, $T - 1 \sim B(n-1, 0.5)$. When $n \to \infty$, the binomial distribution $B(n-1, 0.5)$ asymptotically tends to normal distribution with the mean $0.5(n-1)$ and the variance $0.25(n-1)$. As a result, the statistics T asymptotically follows $N(0.5(n+1), 0.25(n-1)$.

When $a \neq 0.5$, denote $k_0 = \text{argmax}_k\{P(T = 2k), k = 1, 2, \ldots, \lfloor n/2 \rfloor\}$, and we have $P(T = 2k_0) - P(T = 2k_0 + 1)$ does not converge to 0 as $n \to \infty$. As a result, the statistic Tdoes not converge to any smooth continuous distribution when $a \neq 0.5$. ■

***Proof for Theorem* 2.3:** In order to prove that T_1 is asymptotical normal distribution with mean μ_1 and variance σ_1^2, we only need to derive that

$$
f_k(a) := P\{T_1 = 2k\} \cong g_k(a) =: \frac{2}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(2k - \mu_1)^2}{2\sigma_1^2}\right),
$$

where $\mu_1 = 2(n-3)a(1-a) + 2, \sigma_1^2 = 4a(1-a)(3na^2 - 3na + n - 11a^2 + 11a - 3)$. Based on the Taylor expansion, we have

$$
f_k(a) = \sum_{m=0}^{n-2} f_k^{(m)}(0.5) \frac{(a-0.5)^m}{m!}, \quad g_k(a) = \sum_{m=0}^{+\infty} g_k^{(m)}(0.5) \frac{(a-0.5)^m}{m!}.
$$

In the following, we prove that $f_k^{(m)}(0.5) \cong g_k^{(m)}(0.5)$ for $m = 0, 1, 2, \ldots, n-2$ as $n \to \infty$. For m being odd number, based on Lemma 5.1, it can be concluded that $f_k^{(m)}(0.5) = g_k^{(m)}(0.5) = 0$. For $m = 0$, we have $f_k^{(0)}(0.5) = C_{n-1}^{2k-1} 0.5^{n-2} = 2C_{n-1}^{2k-1} 0.5^{n-1}$ and

$$
g_k^{(0)}(0.5) = \frac{2}{\sqrt{2\pi \cdot 0.25(n-1)}} \exp\left(-\frac{(2k - 1 - 0.5(n-1))^2}{2 \cdot 0.25(n-1)}\right).
$$

According to the fact that $f_k^{(0)}(0.5) = 2P\{X = 2k - 1\}$ with $X \sim B(n-1, 0.5)$, and that X is asymptotical normal distribution $N(0.5(n-1), 0.25(n-1))$, we have $f_k^{(0)}(0.5) \cong g_k^{(0)}(0.5)$. For $m = 2$, by Lemma 5.1, it can be derived that

$$
\begin{aligned}
f_k^{(2)}(0.5) &= \sum_{n_1=k}^{n-k} C_{n_1-1}^{k-1} C_{n-n_1-1}^{k-1} 0.5^{n-2} \cdot [4(n^2 - n + 2) - 16nn_1 + 16n_1^2] \\
&= C_{n-1}^{2k-1} 0.5^{n-2} \cdot \left(4(-n^2 - n + 2) - \frac{8k}{(2k+1)} n + \frac{8(k+1)}{(2k+1)} n^2\right) \\
&=: f_k^{(0)}(0.5) \times J_1
\end{aligned}
$$

and

$$\begin{aligned} g_k^{(2)}(0.5) &= \frac{2\sqrt{2}}{\sqrt{\pi(n-1)}} \exp\left\{-\frac{2(2k-\frac{n+1}{2})^2}{n-1}\right\} \\ &\quad \times \left(-24+16n-96k+\frac{128k^2+192k-96}{n-1}-\frac{128(2k-1)^2}{(n-1)^2}\right) \\ &=: g_k^{(0)}(0.5) \times J_2. \end{aligned}$$

Here

$$\begin{aligned} J_1 &= 4(-n^2-n+2) - \frac{8k}{(2k+1)}n + \frac{8(k+1)}{(2k+1)}n^2, \\ J_2 &= -24+16n-96k+\frac{128k^2+192k-96}{n-1}-\frac{128(2k-1)^2}{(n-1)^2}. \end{aligned}$$

Under the condition that $\lim_{n\to\infty}(2k-(n+1)/2)^2/(n-1) = c \geq 0$, we have $k = (n+1)/4 \pm \frac{1}{2}\sqrt{c(n-1)}$ and $\lim_{n\to\infty} J_1/J_2 = 1$. Note that $f_k^{(0)}(0.5) \simeq g_k^{(0)}(0.5)$, we have $f_k^{(2)}(0.5) \simeq g_k^{(2)}(0.5)$.

When $\lim_{n\to\infty}[(2k-(n+1)/2)^2]/(n-1) = \infty$, $J_1 = O(n^2)$, $J_2 = O(n)$ and $D_2 = o(n^{-5/2})$. According to the De Movire–Laplace theorem, $D_1/D_2 \to 1$ as $n \to \infty$. Hence,

$$\begin{aligned} f_k^{(2)}(0.5) &= O(n^2) \cdot o(n^{-5/2}) = o(n^{-1/2}) \to 0, \\ g_k^{(2)}(0.5) &= O(n) \cdot o(n^{-5/2}) = o(n^{-3/2}) \to 0. \end{aligned}$$

That is $f_k^{(2)}(0.5) \simeq g_k^{(2)}(0.5)$. Therefore, $f_k^{(2)}(0.5) \simeq g_k^{(2)}(0.5)$. For m being even number and larger than 2, based on Lemma 5.1, we can derive that $f_k^{(m)}(0.5) \simeq g_k^{(m)}(0.5)$ by following the derivation for $m=2$.

Hence, T_1 is asymptotical normal distribution with mean μ_1 and variance σ_1^2. By a similar method, it can be proved that T_2 is asymptotical normal distribution with mean μ_2 and variance σ_2^2.

In the following, we compute critical values t_1, t_2, which makes $P(T < t_1) = \alpha/2$ and $P(T > t_2) = \alpha/2$. Note that

$$\begin{aligned} &P(T < t_1) \\ &\quad = P(T < t_1, T \text{ is even}) + P(T < t_1, T \text{ is odd}) \\ &\quad = P(T < t_1 \mid T \text{ is even}) \cdot P(T \text{ is even}) + P(T < t_1 \mid \text{T is odd}) \cdot P(\text{T is odd}) \\ &\quad = P(T_1 < t_1) \cdot \theta_1 + P(T_2 < t_1) \cdot \theta_2 \\ &\quad = P\left(\frac{T_1-\mu_1}{\sigma_1} < \frac{t_1-\mu_1}{\sigma_1}\right) \cdot \theta_1 + P\left(\frac{T_2-\mu_2}{\sigma_2} < \frac{t_1-\mu_2}{\sigma_2}\right) \cdot \theta_2 \\ &\quad = \Phi\left(\frac{t_1-\mu_1}{\sigma_1}\right) \cdot \theta_1 + \Phi\left(\frac{t_1-\mu_2}{\sigma_2}\right) \cdot \theta_2, \end{aligned}$$

we let $\Phi((t_1 - \mu_1)/\sigma_1) \cdot \theta_1 + \Phi((t_1 - \mu_2)/\sigma_2) \cdot \theta_2 = \alpha/2$. Solve this equation and we have

$$
\begin{aligned}
t_1 &\approx -1.96\sqrt{-4(3n-5)a^4 + 8(3n-5)a^3 + 2(4n-7)a^2 + 2(2n-3)a} \\
&\quad - 2(n-1)a^2 + 2(n-1)a + 1 \\
&= -1.96\sqrt{4(5-3n)a^2(1-a)^2 + 2(2n-3)a(1-a)} + 2a(1-a)(n-1) + 1 \\
&= -\sigma z_{\alpha/2} + \mu.
\end{aligned}
$$

Similarly, $t_2 \approx \sigma z_{\alpha/2} + \mu$. ■

Proof for Theorem* 2.4*: For the asymptotical properties of the variables T_1 and T_2 under alternative hypothesis, the proof is very similar to that in Theorem 2.3, and we do not detail it here. Under the condition that statistics T_1 and T_2 follow normal distributions $N(\tilde{\mu}_1, \tilde{\sigma}_1^2)$ and $N(\tilde{\mu}_2, \tilde{\sigma}_2^2)$, respectively, we discuss the choice of optimal a in the following. For significant level α, the test power is

$$
\begin{aligned}
Q(a) &= P(T < t_1, \text{ or } T > t_2) \\
&= P(T < t_1) + P(T > t_2), \\
P(T < t_1) &= P(T_1 < t_1) \cdot P(T_1 \text{ is even}) + P(T_2 < t_1) \cdot P(T_2 \text{ is odd}) \\
&= P\left(\frac{T_1 - \tilde{\mu}_1}{\tilde{\sigma}_1} < \frac{t_1 - \tilde{\mu}_1}{\tilde{\sigma}_1}\right) \cdot \tilde{\theta}_1 + P\left(\frac{T_2 - \tilde{\mu}_2}{\tilde{\sigma}_2} < \frac{t_1 - \tilde{\mu}_2}{\tilde{\sigma}_2}\right) \cdot \tilde{\theta}_2 \\
&= \Phi\left(\frac{t_1 - \tilde{\mu}_1}{\tilde{\sigma}_1}\right) \cdot \tilde{\theta}_1 + \Phi\left(\frac{t_1 - \tilde{\mu}_2}{\tilde{\sigma}_2}\right) \cdot \tilde{\theta}_2 \\
&= \Phi\left(\frac{-\sigma z_{\alpha/2} + \mu - \tilde{\mu}_1}{\tilde{\sigma}_1}\right) \cdot \tilde{\theta}_1 + \Phi\left(\frac{-\sigma z_{\alpha/2} + \mu - \tilde{\mu}_2}{\tilde{\sigma}_2}\right) \cdot \tilde{\theta}_2, \\
P(T > t_2) &= 1 - \Phi\left(\frac{\sigma z_{\alpha/2} + \mu - \tilde{\mu}_1}{\tilde{\sigma}_1}\right) \cdot \tilde{\theta}_1 - \Phi\left(\frac{\sigma z_{\alpha/2} + \mu - \tilde{\mu}_2}{\tilde{\sigma}_2}\right) \cdot \tilde{\theta}_2, \\
Q(a) &= 1 + \Phi\left(\frac{-\sigma z_{\alpha/2} + \mu - \tilde{\mu}_1}{\tilde{\sigma}_1}\right) \cdot \tilde{\theta}_1 + \Phi\left(\frac{-\sigma z_{\alpha/2} + \mu - \tilde{\mu}_2}{\tilde{\sigma}_2}\right) \cdot \tilde{\theta}_2 \\
&\quad - \Phi\left(\frac{\sigma z_{\alpha/2} + \mu - \tilde{\mu}_1}{\tilde{\sigma}_1}\right) \cdot \tilde{\theta}_1 - \Phi\left(\frac{\sigma z_{\alpha/2} + \mu - \tilde{\mu}_2}{\tilde{\sigma}_2}\right) \cdot \tilde{\theta}_2.
\end{aligned}
$$

Here, $\Phi(\cdot)$ stands for the standard normal distribution function. ■

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research was supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China [No. 12XNI004].

References

[1] Taussky O, Todd J. Generation and testing of pseudo-random numbers. In: Meyer HA, editor. Symposium on Monte Carlo methods. New York: John Wiley and Sons; 1956. p. 15–28.

[2] Kossack CF. A million random digits with 100,000 normal deviates by the rand corporation. J R Stat Soc. 1955;3(4):568–571.

[3] Wald A, Wolfowitz J. On a test whether two samples are from the same population. Cancer Res.. 1940;51(2):612–618.

[4] Mood AM. The distribution theory of runs. Ann Math Statist. 1940;11(209):367–392.

[5] Cowles A, Jones HE. Some a posteriori probabilities in stock market action. Econometrica. 1937;5(3):280–294.

[6] Weiler H. A new type of control chart limits for means, ranges, and sequential runs. J Am Stat Assoc. 1954;49(266):298–314.

[7] Moore PG. A test for randomness in a sequence of two alternatives involving a 2×2 table. Biometrika. 1949;36(3–4):305–316.

[8] Stuart A. The efficiencies of tests of randomness against normal regression. J Am Stat Assoc. 1956;51(274):285–287.

[9] Henze N. A quick omnibus test for the proportional hazards model of random censorship. Stat J Theoret Appl Statist. 1993;24(3):253–263.

[10] Gibbons JD, Chakraborti S. Nonparametric statistical inference. 3rd ed. J R Stat Soc. 1986;149(3):93–113.

[11] Gibbons JD. Nonparametric methods for quantitative analysis. American Sci. 1985;41(1): 835–853.

[12] Cohen JP, Menjoge SS. One-sample run tests of symmetry. J Stat Plann Inference. 1988;18(1):93–100.

[13] McWilliams TP. A distribution-free test for symmetry based on a runs statistic. J Am Stat Assoc. 1990;85(412):1130–1133.

[14] Friedman JH, Rafsky LC. Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. Ann Stat. 1979;7(4):697–717.

[15] Marden JI Multivariate rank tests. In: Ghosh S, editor. Multivariate analysis design of experiments & survey sampling. New York: Marcel Dekker; 1999. p. 401–432.

[16] Paindaveine D. On multivariate runs tests for randomness. J Am Stat Assoc. 2009;104(488): 1525–1538.

[17] Ley C. Univariate and multivariate symmetry: statistical inference and distributional aspects [unpublished thesis]. Brussels: Univ. Libre de Bruxelles; 2010.

[18] Dyckerhoff R, Ley C, Paindaveine D. Depth-based runs tests for bivariate central symmetry. Ann Inst Stat Math. 2015;67:917–941.

[19] Ley C, Paindaveine D Runs tests. In: Encyclopedia of environmetrics. 2nd ed. New York: John Wiley; 2012. p. 2474–2481.