Distribution Theory of Some Nonparametric Statistics via Finite Markov Chain Imbedding Technique

by

Wan-Chen Lee

A Thesis submitted to
the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

Abstract

The ranking method used for testing the equivalence of two distributions has been studied for decades and is widely adopted for its simplicity. However, due to the complexity of calculations, the power of the test is either estimated by normal approximation or found when an appropriate alternative is given. Here, via a Finite Markov chain imbedding (FMCI) technique, we are able to establish the marginal and joint distributions of the rank statistics considering the shift and scale parameters, respectively and simultaneously, under two continuous distribution functions. Furthermore, the procedures of distribution equivalence tests and their power functions are discussed. Numerical results of a joint distribution of two rank statistics under the standard normal distribution and the powers for a sequence of alternative normal distributions with mean from -20 to 20 and standard deviation from 1 to 9 and their reciprocals are presented. In addition, we discuss the powers of the rank statistics under the Lehmann alternatives.

Wallenstein et al. (1993, 1994) discussed power via combinatorial calculations for the scan statistic against a pulse alternative; however, unless certain proper conditions are given, computational difficulties exist. Our work extends their results and provides an alternative way to obtain the distribution of a scan statistic under various alternative conditions. An efficient and intuitive expression for the distribution as well as the power of the scan statistic are introduced via the FMCI. The numerical results of the exact power for a discrete scan statistic against various conditions are presented. Powers through the finite Markov chain imbedding method and a

combinatorial algorithm for a continuous scan statistic against a pulse alternative of a higher risk for a disease on a specified subinterval time are also discussed and compared.

Keywords: FMCI; hypothesis test; Lehmann alternative; rank statistic; rank-sum test; scan statistic; shift parameter, scale parameter, power

Acknowledgments

I would first like to gratefully express my appreciation to my advisor, Professor Emeritus James C. Fu, for his patience, guidance, and financial support through his Natural Sciences and Engineering Research Council of Canada research grant during my studies. This dissertation would not have been completed without his challenges and encouragements. In particular, I learned plenty about the enthusiasm and persistence in research from him. It is my privilege to have Dr. Fu as my mentor throughout my studies at the University of Manitoba.

I would also like to thank the other members of my committee: Drs. Benqi Guo, Alexandre Leblanc, and Xiqui Wang for their insightful suggestions and helpful discussions, thereby making my dissertation a better work. Thanks also go to my external examiner, Dr. Deli Li, for taking his time and patience to read this dissertation and provide meaningful suggestions.

Many thanks go to the Department of Statistics for awarding me a Faculty of Science Studentship and giving me the opportunities to teach an introductory level course in-person and online a few times. Thanks also go to Dr. Brad C. Johnson for the discussions and suggestions to better improve my teaching skills.

Last but not least, love and support of my friends are unforgettable, particularly Jal and Jenn who have been so supportive in my personal life over the past few years. This dissertation is dedicated to my family for their unconditional love and unending support who never stopped believing in me.

Dedication

To My Family

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Data express opinions through statistical analyses and the duties for statisticians are to dig out the information hidden in the numbers with appropriate statistical analysis methods. In the field of Statistics, descriptive statistics and inferential statistics are two well-known broad subdivisions. Descriptive statistics draws the basic features of the given data by simple summaries or graphical representations. Moreover, developing conclusions, making decisions, and having predictions are cases of statistical inference. Procedures, such as modeling, correlation analysis and hypotheses testing, are based on the population assumptions. When specific assumptions are known not to hold or when one is not willing to verify such assumptions, nonparametric statistical procedures fill the need. The advantage of using certain nonparametric statistics is its distribution-free property under the null hypothesis. However, one of the disadvantages is that the power function for the nonparametric test in most cases could only be found by either assuming an appropriate alternative distribution function or using the normality limiting distribution. The main difficulty of finding power for a nonparametric test is due to the complexity in mathematical calculations. Since nonparametric statistics are

widely applied to the studies of real-world phenomena, the method of the Finite Markov Chain Imbedding, brought out by the end of the twentieth century, is used to study the distributions of proposed nonparametric statistics and their power functions.

1.1 Finite Markov Chain Imbedding Technique

Suppose we have a index set $\Gamma = \{0, 1, \ldots\}$ and a state space of a finite number of possible values $\Omega_t = \{a_1, \ldots, a_{m_t}\}, t \in \Gamma$. A stochastic process $\{Y_t, t \in \Gamma\}$ is a set of random variables that takes values on Ω_t where Y_t is referred to as the state of the process at time t. Here, the stochastic process is called a discrete-time process since the index set is countable. The index set could also be an interval of the real line in which case the corresponding stochastic process is said to be a continuous-time process. Stochastic process is a branch of probability theory that offers sophisticated solutions to many practical questions. A first order Markov chain can be interpreted as the conditional distribution of any future state Y_{t+1} given the past states $Y_0, \ldots, Y_{t-1}$ and the present state Y_t is independent of the past states and depends only on the present state (Ross, 2000, Chap. 4). The Markov chain is said to be homogeneous if $\Omega_i = \Omega_j, \forall i, j \in \Gamma$ and heterogeneous if at least one $\Omega_i \neq \Omega_j$,, for any $i \neq j$.

The outcome of a trial which is either a success or a failure is called a Bernoulli or bistate trial. A sequence of consecutive successes or failures is defined as a successes or failures run. For example, a successes or failures run of size 3 represents the sequence of outcomes SSS or FFF, respectively. In the statistical model there are five commonly used statistics of success runs:

(i) the number of success runs of size exactly k;

(ii) the number of success runs of size greater than or equal to k;

(iii) the number of non-overlapping consecutive k successes;

(iv) the number of overlapping consecutive k successes;

(v) the size of the longest success run.

More generally, there are m possible outcomes for one trial. A simple or compound pattern, instead of runs, is of interest for multi-state cases. For instance the number of the DNA pattern "ACT" in the sequence "$ACGGTCACTGGTCACT$" is 2. Traditionally, the occurrence of runs and patterns in a sequence of n trials is studied through combinatorial calculations which are complicated and tedious. Fu and Koutras (1994) propose the finite Markov chain imbedding technique (FMCI), a simple unified approach, for the distribution theory of runs and the waiting time for the m^{th} occurrence of a specific run or pattern. Since then, there has been a good number of studies adopting the method of FMCI to the distribution theory of runs and patterns in a sequence of multi-state trials, see Koutras and Alexandrou (1995), Fu (1996), Lou (1996), Smit (1996), Lou (1997), Fu (2001), Fu et al. (2002), Fu and Lou (2008). This FMCI technique has also been applied in various studies, for instance distribution theory of waiting time (Koutras, 1996; Koutras and Alexandrou, 1997; Chang, 2005; Fu and Lou, 2006), reliability systems (Koutras, 1996; Fu, 2002; Fu et al., 2002; Koutras and Maravelakis, 1997; Zhao et al., 2007; Zhao and Cui, 2009; Antzoulakos et al., 2009), random permutations (Fu, 1995; Fu and Jonson, 2000; Fu and Lou, 2000; Johnson, 2001, 2002), DNA sequence analysis (Lou, 2003; Nuel,

2006), continuity of care measure (Lou, 1997, 1999; Fu and Lou, 2000), distributions of Eulerian and Simon Newcomb numbers (Fu et al., 1999), and boundary crossing probabilities for Brownian motion (Fu and Wu, 2010). A decade after the method was introduced, Fu and Lou (2003) summarize the studies in the book *Distribution of runs and patterns and its applications* and point out:

However, finding the appropriate combinatorial identities to derive the probability distribution can be difficult, if not impossible, for complex runs and patterns, and this perhaps is the reason why the exact distributions of many common statistics defined on runs and patterns remain unknown.

Recently, Fu and Lou (2007) show that the number of simple or compound patterns in a sequence of multi-states trials has a asymptotically normal distribution. Fu and Jonson (2009) and Fu et al. (2012) approximate the tail probability for the distribution of the number of patterns in a long sequence of independent and identically distributed (*i.i.d.*) multi-state or Markov-dependent trials through finite Markov chain imbedding.

1.2 Rank-sum tests

Suppose that after observing $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from the cumulative distribution functions F and G, respectively, we wish to test the hypothesis, for some $\theta \neq 0$:

$$Ho : F(x) = G(x) \quad \text{versus} \quad Ha : F(x) = G(x - \theta), \quad \text{for all} \quad x.$$

This is known as the shift alternative. Wilcoxon (1945) proposes a ranking method for testing the significance of the difference of the means of the two populations, which is also known as the Wilcoxon-Mann-Whitney (WMW) test or Wilcoxon rank-sum test. Wilcoxon defines a statistic W_Y which is the sum of the ranks of the $y's$ in the ordered sequence of $x's$ and $y's$, equivalent to

$$W_Y = \sum_{j=1}^{n} \{ \ \# \ \text{ of } \ x_i's \ < \ y_j \ \} + \frac{n(n+1)}{2}.$$

Mann and Whitney (1947) introduce an elaborate idea on ranking tests and proposed the statistic

$$U = \sum_{j=1}^{n} \{ \ \# \ \text{ of } \ x_i's \ > \ y_j \ \} = mn - W_Y + \frac{1}{2}n(n+1).$$

They also prove that the limiting distribution, in terms of the Wilcoxon form of the test, is normal

$$Z_W = \frac{W_Y - \mathbb{E}(W_Y)}{\sqrt{\mathbb{V}\text{ar}(W_Y)}} \xrightarrow{L} N(0,1)$$

as m and n tend to infinity in any arbitrary manner. Hence the normal approximation with continuity correction gives

$$\text{power} \approx 1 - \Phi \left(\frac{c - \frac{1}{2}mnp_1}{\sqrt{\mathbb{V}\text{ar}(W_Y)}} \middle| \ H_a \right)$$

where c is the value such that

$$\Phi\left(\frac{c-\frac{1}{2}mn}{\sqrt{\frac{1}{12}mn(m+n+1)}}\,\middle|\,H_o\right)\geq 1-\alpha,$$

$$p_1 = P(X < Y),$$

and the variance of the test statistic is defined by

$$\mathbb{V}\text{ar}(Z_W) = mnp_1(1-p_1) + mn(n-1)(p_2-p_1^2) + mn(m-1)(p_3-p_1^2),$$

with

$$p_2 = P(X < Y \text{ and } X < Y'),$$

X, Y and Y' being independently distributed, X with distribution F, and Y and Y' having the same distribution G, and

$$p_3 = P(X < Y \text{ and } X' < Y),$$

X, X' and Y being independently distributed, X and X' having the same distribution F, and Y with distribution G. Over the years, there have been studies on finding the exact or approximate power for the Wilcoxon rank-sum test. For instance, Shieh et al. (2006) derive the exact power for the uniform, normal, double exponential and exponential shift models by choosing an appropriate alternative distribution function in order to determine the values of p_1, p_2, and p_3. Rosner and Glynn (2009) discuss power against alternatives of the form

$$\Phi^{-1}(F_Y(y)) = \Phi^{-1}(F_X(y)) + \mu \text{ for some } \mu \neq 0,$$

where the underlying distributions F_X and F_Y are normal. Collings and Hamilton (1988) presented a bootstrap method to find the empirical distribution functions in order to approximate the power against the shift alternative. Lehmann (1953) drives the power function as

$$P(S_1 = s_1, S_2 = s_2, \cdots, S_n = s_n) = \frac{k^n}{\binom{m+n}{m}} \prod_{j=1}^{n} \frac{\Gamma(s_j + jk - j)}{\Gamma(s_j)} \frac{\Gamma(s_{j+1})}{\Gamma(s_{j+1} + jk - j)},$$

where s_j is the rank of y_j in the combined samples for the alternative hypothesis of

$$G_Y(x) = F_X(x)^k, \text{ for all } x,$$

where k is a positive integer. However Lehmann (1998) mentions that the power of the Wilcoxon rank-sum test obtained here was only qualitative due to the computation of numerical values of the power is considerably complicated when F and G are continuous distributions with $F \neq G$.

As the rank-sum test is widely adopted for testing the centre difference of two distributions, it is natural to study the efficiency of a rank-sum test for variability (Ansari and Bradley, 1960). For decades, studies have been focused on proposing new definitions of the rank statistic and using the methods of Chernoff and Savage to show the efficiency of the proposed statistic relative to the F-test. Four definitions of the rank statistics considering scale parameter are introduced below.

Let $D_i = i$ be the rank assigned to the i^{th} smallest value in the combined and ordered sample of $X's$ and $Y's$ and define $I(D_i)$ as the indicator variable such that

$$I(D_i) = \begin{cases} 1 & \text{if } D_i \text{ is from } X\text{-sample} \\ 0 & \text{otherwise.} \end{cases}$$

Mood (1954) proposes to use $D_i = i$ and the test statistic

$$\sum_{i=1}^{m+n} \left(D_i - \frac{m+n+1}{2} \right)^2 (1 - I(D_i)),$$

the sum of the squared deviations of each rank of the y observations to the mean rank. The larger/smaller the rank statistic, the more/less variability the distribution will be. Klotz (1962) proposes the statistic

$$\sum_{i=1}^{m+n} \left(\Phi^{-1} \left(\frac{D_i}{m+n+1} \right) \right)^2 I(D_i) \tag{1.1}$$

where Φ^{-1} is the inverse of the standard cumulative normal distribution and $D_i = i$. The idea of this definition is originated from giving more weight to the extreme ranks. Siegel and Tukey (1960) address assigning the rank D_i in the following sense: rank 1 to the smallest data point, rank 2 and 3 to the two largest numbers in the data, ranks 4 and 5 to the next two lowest, etc. This ranking procedure, given $m = 8$ and $n = 9$, is presented in the following example:

Data:	0	3	5	6	8	8	10	10	11	12	13	13	14	15	16	17
Sample:	X	Y	X	Y	X	Y	Y	Y	Y	Y	X	X	Y	X	X	X
Rank D_i:	1	4	5	8	9	12	13	16	15	14	11	10	7	6	3	2

Ansari and Bradley (1960) suggests to assign D_I starting from both ends beginning with unity and working towards the centre. Hence, if $m+n$ is even, the array of ranks D_i are given by

$$1, 2, 3, \ldots, (m+n)/2, \quad (m+n)/2, \ldots, 3, 2, 1;$$

and , if $m+n$ is odd, the array of ranks D_i becomes

$$1, 2, 3, \ldots, (m+n-1)/2, \quad (m+n+1)/2 \quad (m+n-1)/2, \ldots, 3, 2, 1.$$

Both Siegel and Tukey (1960) and Ansari and Bradley (1960) defined the associated rank statistic as

$$\sum_{i=1}^{m+n} D_i I(D_i),$$

except $D_i, i = 1, \ldots, m+n$ are defined accordingly. Ansari and Bradley (1960) mention that if the means of the X and Y samples cannot be considered equal, the difference in location has severe impact on all the tests of dispersion. Klotz (1962) shows that the power of a rank test can be found by integrating the joint density of X and Y samples over that part of the $m+n$ dimensional space defined by the alternative orderings which lie in the critical region of the test, for which conditions are very strict.

1.3 Scan Statistics

Let $X_1, X_2, \ldots, X_n$ be a sequence of bistate trials with outcomes 1 (or S) and 0 (or F). The scan statistic of window size r is defined as

$$S_n(r) = \max_{1 \le i \le n-r+1} S_n(r, i), \tag{1.2}$$

where

$$S_n(r, i) = \sum_{j=i}^{i+r-1} X_j.$$

The main idea of scan statistics is to detect whether there is an unusual number of specified outcomes in any r-consecutive trials among a sequence of n trials. Statistics can help identify that a change has taken place in the underlying process if any r-consecutive trials have plenty of specified outcomes relative to the null assumption, which is known as cluster analysis (Naus, 1974). Scan statistics have been one of the most widely used methods in many fields, such as astronomy, bioinformatics, electrical engineering, epidemiology, genetics, reliability and quality control, and telecommunication. For instance, one interesting application of scan statistic is to the problem of gene mapping. Hoh and Ott (2000) use a scan statistic with varying length to examine the genome for susceptibility genes to autism. In their work, they successfully detected a possible marker associated with the disease. This motivated authors to investigate power of tests based on scan statistics, an important issue to establish the credibility of a hypothesis test.

For bistate trials, it is common to assume the null distribution of a Bernoulli sequence of n independent trials with a constant probability of success

$$H_o : \pi_t = \pi \text{ for } t = 1, 2 \ldots, n.$$

Wallenstein et al. (1994) present approximations as well as bounds for power functions of a scan statistic against some contiguous sequence of Ir events with higher probability of observing a success which starts at unknown trial τ

$$H_a : \pi_t = \begin{cases} \phi & \tau \leq t \leq \tau + Ir - 1 \\ \pi & \text{otherwise,} \end{cases} \tag{1.3}$$

where $\phi > \pi$ and I is a positive integer. Let s_α be the critical value at the α level of significance, such that

$$P(S_n(r) \geq s_\alpha | H_o) \leq \alpha.$$

Then, the approximation for the power function performs well for $I = 1$ and is defined as

$$1 - Q(\phi, \pi, \pi) \left(\frac{Q(\pi, \pi, \pi)}{Q(\pi, \pi)} \right)^{n/r-3} \tag{1.4}$$

where

$$Q(p, q) = F(s_\alpha - 1, n, q)F(s_\alpha - 1, n, p) - b(s_\alpha, n, p) \sum_{j=1}^{s_\alpha - 1} \theta^{-j} F(s_\alpha - j - 1, n, q)$$

and

$$
\begin{aligned}
Q(p,q,v) &= F(s_\alpha - 1, n, p)F(s_\alpha - 1, n, q)F(s_\alpha - 1, n, v) \\
&- b(s_\alpha, n, q)F(s_\alpha - 1, n, v) \sum_{j=1}^{s_\alpha - 1} \left\{ \frac{p/(1-p)}{q/(1-q)} \right\}^j F(s_\alpha - j - 1, n, p) \\
&- b(s_\alpha, n, q)F(s_\alpha - 1, n, p) \sum_{j=1}^{s_\alpha - 1} \left\{ \frac{v/(1-v)}{q/(1-q)} \right\}^j F(s_\alpha - j - 1, n, v) \\
&+ \sum_y \sum_x \sum_z b(x, n, p)b(y, n, q)b(z, n, v)f(x, y, z) \left\{ g(x, y, z) + h(x, y, z) \right\},
\end{aligned}
$$

where

$$\theta = \frac{\phi(1-\pi)}{\pi(1-\phi)}$$

$$b(s_\alpha, n, q) = \binom{n}{s_\alpha} q^{s_\alpha}(1-q)^{n-s_\alpha}$$

$$F(s_\alpha, n, q) = \sum_{j=0}^{s_\alpha} b(j, n, q),$$

$$f(x, y, z) = x!(n-x)!y!(n-y)!z!(n-z)!,$$

$$g(x, y, z) = \left((x + y - s_\alpha)!(n + s_\alpha - x - y)!(y + z - s_\alpha)!(n + s_\alpha - y - z)!(2s_\alpha - y)!(n + y - 2s_\alpha)!\right)^{-1},$$

$$
\begin{aligned}
h(x,y,z) &= ((x+y+z-2s_\alpha)!(n+2s_\alpha-x-y-z)!)^{-1} \\
&\times \left((s_\alpha!(n-s_\alpha)!)^{-2} - (y!(n-y)!(2s_\alpha-y)!(n+y-2s_\alpha)!)^{-1}\right),
\end{aligned}
$$

where g and h are set to be 0 when any expression $w!$ for $w < 0$ and the summation is taken over $1 \leq x, y, z \leq s_\alpha - 1$ for τ equal to 1,

$$
1 - \frac{Q(\pi,\phi,\pi)Q(\phi,\pi,\pi)}{Q(\phi,\pi)} \left(\frac{Q(\pi,\pi,\pi)}{Q(\pi,\pi)}\right)^{n/r-4} \tag{1.5}
$$

for τ equal to $r+1$ or $n-2r+1$, and

$$
1 - Q(\pi,\phi,\pi)\frac{Q(\phi,\pi,\pi)^2}{Q(\phi,\pi)^2} \left(\frac{Q(\pi,\pi,\pi)}{Q(\pi,\pi)}\right)^{n/r-5} \tag{1.6}
$$

for $2r+1 \leq \tau \leq n-3r+1$. Moreover, the value of τ can only be 1, $r+1$, $2r+1$, ... Due to the tediousness of the equations, the power functions can be found in Wallenstein et al. (1994) when $1 < I \leq \lfloor \frac{n-\tau+1}{r} \rfloor$. The approximate power functions provided in their work not only require that the number of events with a higher chance of a success is a multiple of the window size but also the ratio of the length of the sequence to the size of the window has to be larger than 3, 4 or 5, depending on the value of τ.

Fu (2001) discusses that the distribution of a discrete scan statistic can be derived from a waiting time distribution of a special type of compound pattern which follows

$$
\begin{aligned}
P(S_n(r) < s) &= P(W(\Lambda_{r,s}) > n) \\
&= \xi_0 N_{r,s}^n \mathbf{1}', \qquad (1.7)
\end{aligned}
$$

where the variable $W(\Lambda_{r,s})$ is defined as the minimum number of trials required to obtain any one of the simple pattern which is associated with the scan statistic $S_n(r)$, ξ_0 is the initial distribution of the embedded Markov chain, $N_{r,s}$ is the transition probability matrix defined on the state space of the embedded Markov chain with the absorbing state excluded and $\mathbf{1}'$ is the transpose of the row vector $\mathbf{1} = (1, \ldots, 1)$.

Let $N(t)$ be the total number of successes that have occurred up to time t, which is a Poisson process. Considering a time frame (0, 1], a continuous scan statistic is defined as the maximum number of successes in a subinterval of fixed length w moving over the entire time interval

$$S(w) = \sup_{0<t\leq 1-w} S(w,t) = \sup_{0<t\leq 1-w} \{N(t+w) - N(t)\}.$$

Through combinatorial calculations Wallenstein et al. (1993) derive the approximate power for a scan statistic to test

$$H_o : N(t) \sim \text{Poisson}(\lambda) \text{ for } 0 < t \leq 1 - w$$

against a pulse alternative of a higher risk of disease on a specified subinterval time

$$H_a : N(t) \sim \begin{array}{ll} \text{Poisson}(\theta\lambda) & \text{for } \tau - w \leq t < \tau + w \\ \text{Poisson}(\lambda) & \text{for } [0 < t < \tau - w) \bigcup [\tau + w < t \leq 1 - w], \end{array} \tag{1.8}$$

where $\theta > 1$ and $w \leq \tau \leq 1 - 2w$. That the number of successes should be considered unknown in designing trials to detect clustering was pointed out in their work, which implied that the unconditional scan statistic is more useful comparing

to the conditional one. Let s_α be the critical value at the α level of significance, that is

$$P(S(w) \geq s_\alpha | H_o) \leq \alpha,$$

and the approximations for finding the power is given by

$$1 - Q_{LH}(Q_3/Q_2)^{1/w-2}, \tag{1.9}$$

where

$$\begin{aligned} Q_{LH} &= F(s_\alpha - 1, \theta\lambda)F(s_\alpha - 1, \lambda) \\ &- p(s_\alpha, \theta\lambda)\left\{F(s_\alpha - 1, \lambda) - \theta^{1-s_\alpha}\exp(\theta\lambda - \lambda)F(s_\alpha - 1, \theta\lambda)\right\}/(\theta - 1), \end{aligned}$$

where

$$p(s_\alpha, \theta) = \frac{e^\theta \theta_\alpha^s}{s_\alpha!}, \quad \text{and} \quad F(s_\alpha, \theta) = \sum_{i=0}^{s_\alpha} \frac{e^\theta \theta^i}{i!}$$

for the case where the pulse begins at $\tau = w$, and by

$$1 - Q_{LHL}(Q_3/Q_2)^{1/w-3}, \tag{1.10}$$

where

$$
\begin{aligned}
Q_{LHL} &= F(s_\alpha - 1, \theta\lambda)F^2(s_\alpha - 1, \lambda) - 2F(s_\alpha - 1, \lambda)p(s_\alpha, \theta\lambda) \\
&\times \left\{F(s_\alpha - 1, \lambda) - \theta^{1-s_\alpha}\exp(\theta\lambda - \lambda)F(s_\alpha - 1, \theta\lambda)\right\}/(\theta - 1) \\
&+ \sum_{j=1}^{s_\alpha - 1} p(s_\alpha + j, \theta\lambda)\theta^{-2j}F^2(s_\alpha - j - 1, \lambda) + S(\lambda, \theta)
\end{aligned}
$$

where

$$
\begin{aligned}
S(\lambda, \theta) &= \sum_{j=1}^{s_\alpha - 2} \left(F(s_\alpha - j - 2, \lambda)(s_\alpha - j - 1 - \lambda) + \lambda p(s_\alpha - j - 2, \lambda)\right) \\
&\times \left(p(s_\alpha, \lambda)p(s_\alpha, \theta\lambda)\theta^{-j} - p(s_\alpha + j, \lambda)p(s_\alpha - j, \theta\lambda)\right)
\end{aligned}
$$

for the case where the pulse begins in the interval $w < \tau \leq 1 - 2w$. The expressions for Q_2 and Q_3 in Equations (1.6) and (1.10) are identical to Q_{LH} and Q_{LHL} evaluated under $\theta = 1$. The disadvantage of these approximations is that the length of the sub-interval time with a higher rate is fixed to be $2w$.

Fu et al. (2012) discretize the interval to n equal subintervals and proved that for a sufficient large n the limiting distribution can be found by

$$
\begin{aligned}
P(\sup_{0<t\leq 1-w} S(w, t) < s) &= \lim_{n\to\infty} P(S_n(\lfloor nw \rfloor + k) < s) \\
&= \lim_{n\to\infty} \xi_0 N_{r,s}^n(p_n)\mathbf{1}', \qquad (1.11)
\end{aligned}
$$

where k is an arbitrary positive integer, $p_n = \lambda/n$, $r = \lfloor nw \rfloor + k$, and $\lfloor \rfloor$ is the floor function mapping a real number to the largest previous integer.

Rothman (1967, 1969) have shown that the scan statistic is uniformly most powerful for testing the null hypotheses of randomness against the clustering alternatives. Power analysis is used not only for calculating the minimum requirement of the sample size, but also to make comparisons between different statistical testing procedures. Our understanding is that power functions through combinatorial calculations exhibit computation difficulties unless specific conditions are imposed, such as choosing a proper starting point or number of events with a higher chance of success. These conditions significantly limit the use of these tests in practice.

1.4 Summary

Nonparametric procedures assess the hypotheses based on the level of the preference or the rank of the numerical observations. The term of rank is used in most of the nonparametric studies. A full-scale development of ranking methods was sparked by a paper(Wilcoxon, 1945) in which a test for comparing two treatments is discussed. There has been a good number of publications ever since. The ranking method is widely used; however, in most cases, the power of associated tests has been approximated or evaluated only in very specific settings. Our approach aims at releasing some of the conditions for finding the distribution of the proposed rank statistic under the null hypothesis as well as alternative.

We first focus on applying the FMCI method to study the distribution of the rank statistic considering shift and scale parameters, respectively. A joint distribution of rank statistics considering shift and scale parameters simultaneously is induced throughout our work which, to the best of our knowledge, has not been studied in the literature. The main strength of using FMCI is to derive the distribution of the

rank statistic without giving any conditions. Therefore, under the null hypothesis of $F = G$, we are able to identify a proper critical region and, under the alternative assumption, the power of the test can be determined naturally. The distribution-free property of the statistic U_n is also demonstrated under the null hypothesis of the distribution equivalence.

Wallenstein et al. (1993, 1994) discuss the power, via combinatorial calculation, for scan statistics against a pulse alternative; however, there exist computational difficulties unless given certain proper conditions. An alternative method to retrieve the distribution of discrete and continuous scan statistics given an alternative distribution is proposed to eliminate the conditions for determining the power of the scan statistic.

This dissertation is organized in the following way. Chapter 2 introduces some preliminary results regarding the basic ideas and techniques of finite Markov chain imbedding.

Chapter 3 proposes procedures for deriving the distribution of the rank statistic considering either the shift or scale parameter, respectively. The procedures are general and can be applied to either two identical distribution functions of interest or two continuous density functions; therefore, their power function can be defined via FMCI. One significant contribution of this dissertation is to retrieve the joint distribution of the rank statistics considering location and scale parameters simultaneously as well as its power function due to the property of the proposed statistic. Numerical results of a joint distribution and some powers of the rank statistics against shift parameter and scale parameter, individually and simultaneously, and the powers of the rank statistics under a family of Lehmann alternatives are presented.

Chapter 4 first gives the expressions for both exact and approximate distributions of a scan statistic for a k-block independent Bernoulli trials and for Markov dependent trials. Using the proposed method, we present numerical results of power for a discrete scan statistic of a sequence of n homogeneous independent bistate trials against alternatives (i) 2-block independent trials, (ii) a certain number of trials with higher chance of success which starts at various positions in a sequence of n bistate trials, (iii) the chance of a success for n trials following a cyclic pattern, and (iv) a sequence of n Markov dependent trials; and powers for a continuous scan statistic against a pulse alternative stated in (1.8). We end the dissertation with a summary and a discussion of possible topics for future research in Chapter 5.

Chapter 2

Preliminary Results

2.1 Patterns

Let $\{X_t\}_{t=1}^n$ be a sequence of multi-state trials that takes value in a set $S = \{a_1, a_2, \ldots, a_d\}$ of d possible outcomes. If the value of d is 2, $\{X_t\}_{t=1}^n$ is reduced to a sequence of Bernoulli trials.

Definition 1. *Λ is called a simple pattern of length k if Λ is composed of a specified k symbols in S.*

Definition 2. *Patterns are said to be distinct if and only if each of the patterns is not part of the other patterns.*

Definition 3. *Λ is called a compound pattern if it is a union of l distinct simple patterns.*

For example, let $S = \{a, c, g, t\}$ and $\Lambda_1 = act$, $\Lambda_2 = tcc$, $\Lambda_3 = ata$, and $\Lambda_4 = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$. We have three simple patterns Λ_1, Λ_2, and Λ_3. Patterns Λ_1, Λ_2, and Λ_3 are distinct as each of the patterns is not part of the other two patterns. Finally, Λ_4 is a compound pattern.

2.2 Finite Markov Chain Imbedding

Let $\Gamma_n = \{0, 1, 2, \cdots, n\}$ be an index set, and let $\Omega_t = \{a_1, \ldots, a_{m_t}\}, t \in \Gamma_n$ be a finite state space where $m_i, \ldots, m_t$ are positive integers.

Definition 4. *The non-negative integer-valued random variable $R(\Lambda)$ is finite Markov chain imbeddable if:*

(a) there exists a finite Markov chain $\{Y_t : t \in \Gamma_n\}$ defined on a finite state space Ω_t with initial probability vector $\boldsymbol{\xi}_0$,

(b) there exists a finite partition $\{C_r : r = 1, \cdots, k_n\}$ on the state space Ω_n, and

(c) for every $r = 1, \cdots, k_n$, we have

$$P(R(\Lambda) = r) = P(Y_n \in C_r | \xi_0).$$

Definition 5. *A state α is called an absorbing state if and only if the system never leaves the state α once it enters α; i.e. $p_{\alpha\alpha} \equiv 1$.*

Let $\{\mathbf{M_t}\}_{t=1}^n$ be the sequence of $m_{t-1} \times m_t$ transition probability matrices of the finite Markov chain Y_t defined on the state space Ω_t and the initial probability distribution $\boldsymbol{\xi}_0 = (P(Y_0 = a_1), P(Y_0 = a_2), \ldots, P(Y_0 = a_{m_1}))$.

The following Theorems (1) to (5) can be found in (Fu and Lou, 2003). We omit the proofs.

Theorem 1. *$R(\Lambda)$ is finite Markov chain imbeddable, and*

$$P(R(\Lambda) = r) = \boldsymbol{\xi}_0(\prod_{t=1}^{n} \mathbf{M}_t)U'(C_r),$$

where $U(C_r) = \sum_{k:a_k \in C_r} e_k$, e_k *is a* $1 \times m$ *unit row vector corresponding to state* a_k, ξ_0 *is the initial probability vector, and* $\mathbf{M}_t, t = 1, 2, \ldots, n$, *are the transition probability matrices of the imbedded Markov chain.*

Let $\Omega_0 = \cdots = \Omega_n = \Omega$ and $A = \{\alpha_1, \alpha_2, \ldots, \alpha_a\}$ be the set of all absorbing states of a homogeneous Markov Chain $\{Y_t\}$ with transition probability matrix M. After some proper arrangement of the state space Ω, the transition probability matrix can be expressed in the form

$$\mathbf{M} = \left[\begin{array}{c|c} \mathbf{N}_{(m-a)\times(m-a)} & \mathbf{C}_{(m-a)\times a} \\ \hline \mathbf{O}_{a\times(m-a)} & \mathbf{I}_{a\times a} \end{array}\right], \tag{2.1}$$

where m and $a(m > a)$ are the numbers of states in Ω and A. Practically, the system always starts in a non-absorbing state. Consequently, we assume that the initial probability distribution is in the form

$$\boldsymbol{\xi}_0 = (\boldsymbol{\xi} : \mathbf{0})_{1\times m},$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_{m-a})$, $\mathbf{0} = (0, 0, \ldots, 0)_{1\times a}$ and $\sum_{i=1}^{m-a} \xi_i = 1$.

Theorem 2. *Given a transition probability matrix* $\mathbf{M}$ *of a homogeneous Markov Chain* $\{Y_t\}$ *in the form of Eq. (2.1), the probability that the the system first enters the set of absorbing states at time* n *can be obtained from*

$$P(Y_n \in A, Y_{n-1} \notin A, \cdots, Y_1 \notin A|\boldsymbol{\xi}_0) = \boldsymbol{\xi}\mathbf{N}^{n-1}(\mathbf{I}-\mathbf{N})\mathbf{1}'. \tag{2.2}$$

Theorem 3. *For any transitive state $i \in \Omega - A$, we have*

$$P(Y_n = i, Y_{n-1} \notin A, \cdots, Y_1 \notin A|\boldsymbol{\xi}_0) = \boldsymbol{\xi}\mathbf{N}^n e_i'. \tag{2.3}$$

Theorem 4. *For any absorbing state $j \in A$, the probability of the system first entering the absorbing state j at the n^{th} step is*

$$P(Y_n = j, Y_{n-1} \notin A, \cdots, Y_1 \notin A|\boldsymbol{\xi}_0) = \boldsymbol{\xi}\mathbf{N}^{n-1}\mathbf{C}_j', \tag{2.4}$$

where $\mathbf{C}_j'$ is the j^{th} column of matrix $\mathbf{C}$.

2.3 Waiting-Time Distribution

Let $\Lambda = b_1 b_2 \ldots b_l$ be a pattern of interest of size l of interest and the random variable $W(\Lambda)$ be defined as the waiting time for the first occurrence of the pattern Λ, *i.e.*

$$W(\Lambda) = \inf\{n : X_{n-l+1} = b_1, X_{n-l+2} = b_2, \cdots, X_n = b_l\},$$

in a sequence of *i.i.d.* multistate trials $\{X_t\}_{t=1}^n$. For instance, given an observed DNA sequence $ACGGTCACTGGTCACT$, $W(ACT) = 9$ means the pattern of interest ACT occurs for the first time at the 9^{th} trial. It is known that the waiting-time random variable $W(\Lambda)$ is homogeneous finite Markov chain imbeddable (Fu and Lou, 2003).

Theorem 5. *Given a pattern Λ, there exists a homogeneous Markov chain $\{Y_t\}$ on a state space Ω with transition probability matrix*

$$\mathbf{M} = \begin{array}{c} \Omega\backslash\alpha \\ \alpha \end{array} \left[\begin{array}{c:c} \mathbf{N} & \mathbf{C} \\ \hdashline \mathbf{O} & \mathbf{I} \end{array}\right], \tag{2.5}$$

such that

(i)

$$P(W(\Lambda) = n) = \boldsymbol{\xi}\mathbf{N}^{n-1}(\mathbf{I} - \mathbf{N})\mathbf{1}', n = 1, 2, \cdots,$$

where α is the set of all absorbing states in the state space Ω, $\boldsymbol{\xi}$ is the initial distribution, $\mathbf{I}$ denotes an identity matrix, and $\mathbf{1}$ is a row vector with each entry equal to one;

(ii) the probability generating function of $W(\Lambda)$ is given by

$$\varphi_W(s) = 1 + (s-1)\boldsymbol{\xi}(\mathbf{I} - s\mathbf{N})^{-1}\mathbf{1}'; \tag{2.6}$$

(iii) the mean of $W(\Lambda)$ is

$$\mu_W = \varphi_W^{(1)}(1), \tag{2.7}$$

and the variance of $W(\Lambda)$ is given by

$$\sigma_W^2 = \varphi_W^{(2)}(1) + \varphi_W^{(1)}(1) - (\varphi_W^{(1)}(1))^2 \tag{2.8}$$

where $\varphi_W^{(i)} = (\partial/\partial s)^i \varphi_W(s)|_{s=1}$, for $i = 1, 2$.

Let $X_n(\Lambda)$ be the number of non-overlapping patterns occurring in the sequence $\{X_i\}_{i=1}^n$ and $W_i(\Lambda)$ be the inter-arrival times of the pattern Λ, which are *i.i.d.* random variables having the same distribution as $W(\Lambda)$ given in the previous theorem. There exists a duality relationship between the non-overlapping counts and inter-arrival times, which is given by

$$P(X_n(\Lambda) < k) = P(W_1(\Lambda) + W_2(\Lambda) + \cdots + W_k(\Lambda) > n), \tag{2.9}$$

for all $k \in \mathbb{N}$. As $n \to \infty$, Feller (1968) showed that

$$EX_n(\Lambda) = \frac{n}{\mu_W}(1 + o(n^{-\frac{1}{2}})). \tag{2.10}$$

Theorem 6. *Under non-overlap counting, the random variable $X_n(\Lambda)$ is asymptotically normally distributed in the sense*

$$\frac{X_n(\Lambda) - \frac{n}{\mu_W}}{\sqrt{\sigma_W^2 \mu_W^{-3} n}} \xrightarrow{L} N(0,1)$$

as $n \to \infty$, where $\xrightarrow{L}$ stands for convergence in law and μ_W and σ_W^2 are given in Theorem 5.

The proof of this theorem can be found in Feller (1968) and Fu and Lou (2007).

Chapter 3

Distributions of Rank Statistics

The Mann-Whitney-Wilcoxon(MWW) test was first addressed by Wilcoxon (1945) and used for testing the equalites of two population location parameters. Here we apply FMCI on this ranking method for determining the distribution of the rank statistics considering the location and scale parameters respectively and the power of the test. Through the procedure, we find the joint distribution of the rank statistics against location and scale parameters, simultaneously.

3.1 Distributions of the rank statistic in the location case

Let $\{X_1, \ldots, X_m\}$ and $\{Y_1, \ldots, Y_n\}$ be two independent samples from the continuous cumulative density distributions $F(x)$ and $G(x-\theta)$, respectively. Given $\boldsymbol{x} = \{x_1, \ldots, x_m\}$ and $x_{[i]}$ is the i^{th} smallest number in the sample, we have

$$p_i = P(x_{[i-1]} < Y < x_{[i]}) = \int_{x_{[i-1]}}^{x_{[i]}} g(y)dy = G(x_{[i]}) - G(x_{[i-1]}),$$

for $i = 1, 2, \ldots, m+1$ where $x_{[0]} = -\infty$ and $x_{[m+1]} = \infty$. Therefore, we define the sampling distribution of Y in the $(m+1)$ intervals as

$$\begin{aligned} \boldsymbol{p} &= \Big(G(x_{[1]}) - G(x_{[0]}), \ldots, G(x_{[m+1]}) - G(x_{[m]})\Big) \\ &= (p_1, p_2, \ldots, p_{m+1})\,. \end{aligned} \tag{3.1}$$

Given m, for $t = 1, 2, \ldots, n$, define the state space

$$\Omega_t = \{\mathbf{U}_t = (u_1(t), \cdots, u_{m+1}(t)) : \sum_{i=1}^{m+1} u_i(t) = t \text{ and } u_i(t) \geq 0, \;\; i = 1, \ldots, m+1\},$$

where $u_i(t)$ is the number of $y's$ in the interval $[x_{[i-1]}, x_{[i]})$ among $y_1, \ldots, y_t$. For each $\mathbf{u}_n = (u_1(n), \cdots, u_{m+1}(n))$, we have a corresponding rank-sum R_l of y's in the combined sequence

$$R_l(\mathbf{U}_n \mid \boldsymbol{X}) = \frac{\sum_{i=1}^{m+1} u_i^2(n) + \sum_{i=1}^{m+1} u_i(n)}{2} + \sum_{i=1}^{m}(u_i(n)+1)\left(\sum_{j=i+1}^{m+1} u_j(n)\right). \tag{3.2}$$

Theorem 7. *The statistic R_l is equivalent to the statistic W_Y, which is addressed by Wilcoxon in 1945.*

Proof. Let

$$I(x_i, y_j) = \begin{cases} 1 & \text{if } x_i < y_j \\ 0 & \text{otherwise.} \end{cases}$$

Then $\sum_{i=1}^{m} I(x_i, y_j)$ is the rank of $y_{[j]}$ within the x observations and $\sum_{i=1}^{m} I(x_i, y_j) + j$ is the rank of $y_{[j]}$ in the combined sample. The rank statistic W_Y, sum of the ranks of y observations, can be determined by

$$\begin{aligned}\sum_{j=1}^{n}\left(\sum_{i=1}^{m} I(x_i, y_j) + j\right) &= \sum_{j=1}^{n}\sum_{i=1}^{m} I(x_i, y_j) + \sum_{j=1}^{n} j \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} I(x_i, y_j) + \frac{n(n+1)}{2}. \qquad (3.3)\end{aligned}$$

The first summation of the first term in Equation (3.3) can be interpreted as the number of y observations larger than x_i which is $\sum_{j=i+1}^{m+1} u_j(n)$ in our expression. It is not difficult to see that $\sum_{i=1}^{m+1} u_i(n)$ equals n the size of y sample. Therefore, Equation (3.3) can be rewritten as

$$\sum_{i=1}^{m}\left(\sum_{j=i+1}^{m+1} u_j(n)\right) + \frac{(\sum_{i=1}^{m+1} u_i(n))(\sum_{i=1}^{m+1} u_i(n) + 1)}{2}$$

$$= \sum_{i=1}^{m}\left(\sum_{j=i+1}^{m+1} u_j(n)\right) + \frac{\sum_{i=1}^{m+1} u_i(n)^2 + 2\sum_{i=1}^{m} u_i(n)(\sum_{j=i+1}^{m+1} u_j(n)) + \sum_{i=1}^{m+1} u_i(n)}{2}.$$

It is then easy to see that

$$\sum_{i=1}^{m}(u_i(n) + 1)\left(\sum_{j=i+1}^{m+1} u_j(n)\right) + \frac{\sum_{i=1}^{m+1} u_i(n)^2 + \sum_{i=1}^{m+1} u_i(n)}{2} = R_l.$$

□

Next, we demonstrate that for two random samples having the same distribution function, the distribution of the random vector $\mathbf{U}_n$ is independent of the form of the distribution function. The distribution of the random vector $\mathbf{U}_n$ is discrete uniform with the mass function one over the number of possible outcomes of the random vector $\mathbf{U}_n$ only when assuming $F = G$. In other words, the distribution of the random variable $\mathbf{U}_n$ can be found by the traditional combinatorial analysis when $F = G$.

Theorem 8. *Distribution-free property of* $\mathbf{U}_n$

$$P(\mathbf{U}_n = \mathbf{u}_n | H_o) = \frac{1}{Card(\Omega_n)} = \frac{1}{\binom{m+n}{n}}. \tag{3.4}$$

Proof. We know the joint PDF of the ordered sample of $x's$ is given by

$$f(x_{[1]}, \ldots, x_{[m]}) = m! \prod_{i=1}^{m} f(x_i)$$

and, when $F = G$, the conditional probability of the random variable $\mathbf{U}_n$ given $\boldsymbol{x} = (x_1, \ldots, x_m)$ is

$$\begin{aligned} P(\mathbf{U}_n = \mathbf{u}_n |\, x_1, \ldots, x_m\,) &= \frac{n!}{\prod_{i=1}^{m+1} u_i(n)!} \prod_{i=1}^{m+1} \left(\int_{x_{[i-1]}}^{x_{[i]}} f(y)dy \right)^{u_i(n)} \\ &= \frac{n!}{\prod_{i=1}^{m+1} u_i(n)!} \prod_{i=1}^{m+1} \Big(F(x_{[i]}) - F(x_{[i-1]}) \Big)^{u_i(n)} \end{aligned} \tag{3.5}$$

where $x_{[0]} = -\infty$ and $x_{[m+1]} = \infty$. By taking the expected value of the conditional probability, we have

$$
\begin{aligned}
& P(\mathbf{U}_n = \mathbf{u}_n | H_o) \\
&= \int_{-\infty \le x_{[1]} \le \cdots \le x_{[m]} \le \infty} \cdots \int P(\mathbf{U}_n |\, x_1, \ldots, x_m,\ H_o) f(x_{[1]}, \ldots, x_{[m]})\, dx_{[1]}\ \cdots\ dx_{[m]} \\
&= \int_{-\infty \le x_{[1]} \le \cdots \le x_{[m]} \le \infty} \cdots \int \frac{n!}{\prod_{i=1}^{m+1} u_i(n)!} (F(x_{[1]}))^{u_1(n)} (F(x_{[2]}) - F(x_{[1]}))^{u_2(n)} \\
&\qquad \cdots (1 - F(x_{[m]}))^{u_{m+1}(n)} m!\ \prod_{i=1}^{m} f(x_i)\, dx_{[1]}\ \cdots\ dx_{[m]} \\
&= \int_{-\infty}^{\infty} \int_{x_{[1]}}^{\infty} \cdots \int_{x_{[m-1]}}^{\infty} \frac{n!}{\prod_{i=1}^{m+1} u_i(n)!} (F(x_{[1]}))^{u_1(n)} (F(x_{[2]}) - F(x_{[1]}))^{u_2(n)} \\
&\qquad \cdots (1 - F(x_{[m]}))^{u_{m+1}(n)} m! dF(x_{[1]}) \cdots dF(x_{[m]}) \\
&= \frac{n!m!}{(n+m)!} \int_{-\infty}^{\infty} \int_{x_{[1]}}^{\infty} \cdots \int_{x_{[m-1]}}^{\infty} \frac{(n+m)!}{\prod_{i=1}^{m+1} u_i(n)!} (F(x_{[1]}))^{u_1(n)} (F(x_{[2]}) - F(x_{[1]}))^{u_2(n)} \\
&\qquad \cdots (1 - F(x_{[m]}))^{u_{m+1}(n)} dF(x_{[1]}) \cdots dF(x_{[m]}). \qquad (3.6)
\end{aligned}
$$

Using variable transformation, it is clear to see that the random variables $F(x_{[1]}), \ldots, F(x_{[m]})$ have a Dirichlet distribution with parameters $u_1(n) + 1, u_2(n) + 1, \ldots\ u_{m+1}(n) + 1$. Therefore, we have

$$
P(\mathbf{U}_n = \mathbf{u}_n | H_o) = \frac{n!m!}{(n+m)!} = \frac{1}{Card(\Omega_n)},
$$

which is independent of the distribution function. □

This is the reason that the distribution of the rank statistic $\mathbf{U}_n$ is said to be distribution-free under the null hypothesis. Unfortunately, when $F \neq G$, we will not be able to establish the distribution of $\mathbf{U}_n$ through Equation (3.5) as solving the multiple integral in Equation (3.6) is either tedious or difficult given some appropriate alternative distribution function. To overcome this difficulty, we bring in the FMCI approach.

Let Ω_t, $t = 0, 1, \ldots, n$, be the state space which has

$$\binom{m+t}{t} \tag{3.7}$$

possible states, $\Gamma_n = \{0, 1, \ldots, n\}$ be an index set, and $\{Z_t : t \in \Gamma_n\}$ be a non-homogeneous Markov chain on the state space Ω_t. As a transition probability matrix M_t for this chain, $t = 1, \ldots, n$, consider

$$M_t = \begin{array}{c} \\ \Omega_{t-1} \end{array} \begin{array}{c} \begin{array}{c} \Omega_t \end{array} \\ \left[\begin{array}{c} p_{\boldsymbol{u}_{t-1},\boldsymbol{u}_t} \end{array} \right]_{\binom{m+t-1}{t-1} \times \binom{m+t}{t}} \end{array},$$

where

$$\begin{aligned} p_{\boldsymbol{u}_{t-1},\boldsymbol{u}_t} &= P(Z_t = \boldsymbol{u}_t | Z_{t-1} = \boldsymbol{u}_{t-1}) \\ &= \begin{cases} p_i & \text{if } u_i(t-1) + 1 = u_i(t) \text{ and } u_j(t-1) = u_j(t) \ \forall\, j \neq i \\ 0 & \text{otherwise} \end{cases}, \end{aligned}$$

and p_i is defined in Equation (3.1).

Theorem 9. *$R_l(\mathbf{U}_n|\boldsymbol{X})$ is finite Markov chain imbeddable, and*

$$P(R_l(\mathbf{U}_n) = r|\boldsymbol{X}) = \boldsymbol{\xi}(\prod_{t=1}^{n} M_t)\boldsymbol{B}'(C_r), \tag{3.8}$$

where $\boldsymbol{B}(C_r) = \sum_{k:R_l(\mathbf{U}_n)=r} e_k$, e_k is a $1 \times \binom{m+n}{n}$ unit row vector corresponding to state $\mathbf{u}_n$, $\boldsymbol{\xi}(P(Z_0 = 1) = 1)$ is the initial probability and M_t, $t = 1, \ldots, n$, are the transition probability matrices of the imbedded Markov chain defined on the state space Ω_t.

Proof. For each $\mathbf{u}_n = (u_1(n), \cdots, u_{m+1}(n))$ in the state space Ω_n, we have a corresponding rank R_l as shown in Equation (3.2). Intuitively, the minimum rank r_{ls} is $n(n+1)/2$ and the maximum rank r_{lb} is $n(2m+n+1)/2$. In accordance with the possible values of the rank R_l, we define a finite partition $\{C_r : r = r_{ls}, \ldots, r_{lb}\}$ such that

$$P(Z_n \in C_r|\boldsymbol{p}) = \boldsymbol{\xi}(\prod_{t=1}^{n} M_t)\boldsymbol{B}'(C_r)$$

where $\boldsymbol{B}(C_r) = \sum_{r:R_l(\mathbf{U}_n)=r} e_k$, e_k is a $1 \times \binom{m+n}{n}$ unit row vector corresponding to state $\mathbf{U}_n$, we then obtain the conditional probability of the rank R_l. □

Then, the Law of Large Numbers (LLN) is used to determine the probability of $\mathbf{U}_n$ for two continuous F and G

$$\frac{1}{N}\sum_{i=1}^{N} P(\mathbf{U}_n = \mathbf{u}_n|\ \boldsymbol{x}_i) \quad \xrightarrow{p} \quad P(\mathbf{U}_n = \mathbf{u}_n) \tag{3.9}$$

where $\boldsymbol{x}_i$ is the i^{th} sample of size m from the distribution function F. It is not difficult to see that

$$P(R_l(\mathbf{U}_n) = r) = \sum_{\mathbf{U}_n : R(\mathbf{U}_n)=r} P(\mathbf{U}_n = \mathbf{u}_n). \tag{3.10}$$

Example 1. *Suppose we have two samples with size $m = 2$ and $n = 3$ selected from $N(0, 1)$ which implies $F = G$. Let $\mathbf{x} = \{x_1, x_2\} = \{-1.25, 0.8\}$ be a random sample. The sampling distribution of y observations in the $m + 1 = 3$ intervals is*

$$\mathbf{p} = (.1056,\ .6825,\ .2119),$$

and the state spaces for the Markov chain are

$$\begin{aligned}
\Omega_0 &= \{(0,0,0)\} \\
\Omega_1 &= \{(1,0,0),(0,1,0),(0,0,1)\} \\
\Omega_2 &= \{(2,0,0),(1,1,0),(1,0,1),(0,2,0),(0,1,1),(0,0,2)\} \\
\Omega_3 &= \{(3,0,0),(2,1,0),(2,0,1),(1,2,0),(1,1,1),(1,0,2) \\
&\qquad (0,3,0),(0,2,1),(0,1,2),(0,0,3)\}.
\end{aligned}$$

The transition probability matrices, then, are as follows

Table 3.1: For $m = 2$ and $n = 3$ the conditional distribution of $\mathbf{U}_n$ given $\mathbf{x} = \{-1.25, 0.8\}$

$\mathbf{U}_n$	R_l	$P(\mathbf{U}_n\|\mathbf{x})$
(3, 0, 0)	6	0.0012
(2, 1, 0)	7	0.0228
(2, 0, 1)	8	0.0071
(1, 2, 0)	8	0.1476
(1, 1, 1)	9	0.0916
(1, 0, 2)	10	0.0142
(0, 3, 0)	9	0.3179
(0, 2, 1)	10	0.2961
(0, 1, 2)	11	0.0919
(0, 0, 3)	12	0.0095

Table 3.2: For $m = 2$ and $n = 3$ the conditional distribution of R_l given $\mathbf{x} = \{-1.25, 0.8\}$

R_l	$\mathbf{U}_n$'s	$P(R_l\|\mathbf{x})$
6	(3, 0, 0)	0.0012
7	(2, 1, 0)	0.0228
8	(2, 0, 1), (1, 2, 0)	0.1547
9	(1, 1, 1), (0, 3, 0)	0.4095
10	(1, 0, 2), (0, 2, 1)	0.3103
11	(0, 1, 2)	0.0919
12	(0, 0, 3)	0.0095

$$M_1 = \begin{matrix}(0,0,0)\end{matrix} \begin{bmatrix} .1056 & .6825 & .2119 \end{bmatrix}$$

$$M_2 = \begin{matrix}(1,0,0)\\(0,1,0)\\(0,0,1)\end{matrix} \begin{bmatrix} .1056 & .6825 & .2119 & 0 & 0 & 0 \\ 0 & .1056 & 0 & .6825 & .2119 & 0 \\ 0 & 0 & .1056 & 0 & .6825 & .2119 \end{bmatrix}$$

$$M_3 = \begin{matrix}(2,0,0)\\(1,1,0)\\(1,0,1)\\(0,2,0)\\(0,1,1)\\(0,0,2)\end{matrix} \begin{bmatrix} .1056 & .6825 & .2119 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .1056 & 0 & .6825 & .2119 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .1056 & 0 & .6825 & .2119 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .1056 & 0 & 0 & .6825 & .2119 & 0 & 0 \\ 0 & 0 & 0 & 0 & .10560 & 0 & 0 & .6825 & .2119 & 0 \\ 0 & 0 & 0 & 0 & 0 & .10560 & 0 & 0 & .6825 & .2119 \end{bmatrix}$$

From the proposed Theorem (9), the conditional distributions of $\mathbf{U}_n$ *and* R_l *are given in Tables 3.1 and 3.2, respectively.*

Then, with ten million simulations and applying Equations (3.9) and (3.10), the distributions of the random variables $\mathbf{U}_n$ *and* R_l *can be found, as shown in Tables 3.3 and 3.4.*

Table 3.3: For $m = 2$ and $n = 3$ the distribution of $\mathbf{U}_n$

$\mathbf{U}_n$	R_l	$P(\mathbf{U}_n)$
(3, 0, 0)	6	0.1000
(2, 1, 0)	7	0.1000
(2, 0, 1)	8	0.1000
(1, 2, 0)	8	0.1000
(1, 1, 1)	9	0.1000
(1, 0, 2)	10	0.1000
(0, 3, 0)	9	0.1000
(0, 2, 1)	10	0.1000
(0, 1, 2)	11	0.1000
(0, 0, 3)	12	0.1000

Table 3.4: For $m = 2$ and $n = 3$ the distribution of R_l

R_l	$\mathbf{U}_n$'s	$P(R_l)$
6	(3, 0, 0)	0.1000
7	(2, 1, 0)	0.1000
8	(2, 0, 1), (1, 2, 0)	0.2000
9	(1, 1, 1), (0, 3, 0)	0.2000
10	(1, 0, 2), (0, 2, 1)	0.2000
11	(0, 1, 2)	0.1000
12	(0, 0, 3)	0.1000

It is worth mentioning that using a simulation size of ten million in order to have the distribution of $\mathbf{U}_n$ or R_l converging (accurate to four decimal places) to the theoretical values. Why do we need such a large number of samples in order to have the sampling distribution converge to the theoretical one? One conjecture is that the cardinal of the random vector $\mathbf{U}_n$ is small. To illustrate our conjecture, we constructed the distribution of $\mathbf{U}_n$ or R_l given $m = 3$ and $n = 4$ for which less than two million simulations were needed to have the distribution of $\mathbf{U}_n$ or R_l converging (accurate to four decimal places) to the theoretical values. And it only took 30,000 simulations to converge for $m = 5$ and $n = 7$. We also want to clarify that there is no doubt that using Theorem 8 is easy and efficient to find the distribution of R_l under the null hypothesis. Since from Equation (3.7), we know that the number of possible states for the random vector $\mathbf{U}_n$ is 10. Therefore, the distribution of $\mathbf{U}_n$ is discrete uniform with probability $\frac{1}{10}$.

To test

$$H_o : F(x) = G(x) \quad \text{versus} \quad H_a : F(x) = G(x-\theta),$$

for some $\theta \neq 0$, the power function is approximated by

$$\begin{aligned}
& P(R_l(\mathbf{U}_n) \leq r_{1\alpha}|H_a) + P(R_l(\mathbf{U}_n) \geq r_{2\alpha}|H_a) \\
= & \sum_{r=r_{ls}}^{r_{1\alpha}} P(R_l(\mathbf{U}_n) = r|H_a) + \sum_{r=r_{2\alpha}}^{r_{lb}} P(R_l(\mathbf{U}_n) = r|H_a) \\
= & \sum_{r=r_{ls}}^{r_{1\alpha}} \sum_{\mathbf{U}_n:R(\mathbf{U}_n)=r} P(\mathbf{U}_n = \mathbf{u}_n|H_a) + \sum_{r=r_{2\alpha}}^{r_{lb}} \sum_{\mathbf{U}_n:R(\mathbf{U}_n)=r} P(\mathbf{U}_n = \mathbf{u}_n|H_a) \\
\approx & \sum_{r=r_{ls}}^{r_{1\alpha}} \sum_{\mathbf{U}_n:R(\mathbf{U}_n)=r} \frac{1}{N} \sum_{i=1}^{N} P(\mathbf{U}_n|H_a;\ \mathbf{X}_i) + \sum_{r=r_{2\alpha}}^{r_{lb}} \sum_{\mathbf{U}_n:R(\mathbf{U}_n)=r} \frac{1}{N} \sum_{i=1}^{N} P(\mathbf{U}_n|H_a;\ \mathbf{X}_i) \\
= & \frac{1}{N} \left(\sum_{r=r_{ls}}^{r_{1\alpha}} \sum_{i=1}^{N} \sum_{\mathbf{U}_n:R(\mathbf{U}_n)=r} P(\mathbf{U}_n|H_a;\ \mathbf{X}_i) + \sum_{r=r_{2\alpha}}^{r_{lb}} \sum_{i=1}^{N} \sum_{\mathbf{U}_n:R(\mathbf{U}_n)=r} P(\mathbf{U}_n|H_a;\ \mathbf{X}_i) \right) \\
= & \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{r=r_{ls}}^{r_{1\alpha}} P(R_l(\mathbf{U}_n) = r|H_a;\ \mathbf{X}_i) + \sum_{r=r_{2\alpha}}^{r_{lb}} P(R_l(\mathbf{U}_n) = r|H_a;\ \mathbf{X}_i) \right),
\end{aligned}$$

where

$$P(R_l(\mathbf{U}_n) \leq r_{1\alpha}|H_o) + P(R_l(\mathbf{U}_n) \geq r_{2\alpha}|H_o) \leq \alpha.$$

Note that the alternative hypothesis is subject to the purpose of the test. This simply needs to slightly modified if a one-sided test is adopted. For example Table 3.4, we can choose a critical region as $\{R_l(\mathbf{U}_n) \leq 6\} \cup \{R_l(\mathbf{U}_n) \geq 12\}$ at the level of

significance 20% for the alternative hypothesis of $\theta \neq 0$. Note that in our example the probability of the smallest and largest value of rank statistic R_l under the null hypothesis is 10%, respectively. The minimum level of significance for a two-sided test can only be set at 20% in this case.

3.2 Distributions of the rank statistic in the scale case

We studied the distribution and the power function of the rank statistic R_l considering a shift in location. Now, the distribution and the power function of the rank statistic considering the scale parameter will be addressed. For this purpose, we consider $F(x) = G(x\sigma^{-1})$ and state the null and alternative hypotheses as

$$H_o : \sigma = 1 \quad \text{versus} \quad H_a : \sigma \neq 1.$$

To do so, we begin with the procedure of finding the distribution of the rank statistic, denoted R_s, considering the scale parameter through the random vector $\mathbf{U}_n$. The array of ranks are given by

$$(m+n)/2, \ldots, 3, 2, 1, \qquad 1, 2, 3, \ldots, (m+n)/2;$$

if $m+n$ is even, and

$$(m+n-1)/2, \ldots, 3, 2, 1, \qquad 0 \qquad 1, 2, 3, \ldots, (m+n-1)/2$$

if $m+n$ is odd. We first introduce how to determine the rank-sum of $y's$ observations in the combined samples, R_s, with respect to

$$\Omega_n = \{\mathbf{u}_n = (u_1(n), \ldots, u_{m+1}(n)) : \sum_{i=1}^{m+1} u_i(n) = n\}$$

where $u_i(n)$ means the number of y observations belonging to $[x_{[i-1]}, x_{[i]})$. Let $med(x, y)$ be the median among $x's$ and $y's$ and belongs to $[x_{[i]}, x_{[i+1]})$ which will then break $\mathbf{U}_n$ into two parts $\boldsymbol{U}_n^-$ and $\boldsymbol{U}_n^+$. If $m + n$ is odd and $med(x, y) = x_{[i]}$, then

$$\boldsymbol{U}_n^- = (u_1^- = u_i(n)\ ,\ u_2^- = u_{i-1}(n)\ ,\ \cdots\ ,\ u_i^- = u_1(n))$$

is a $1 \times i$ vector and

$$\boldsymbol{U}_n^+ = (u_1^+ = u_{i+1}(n)\ ,\ u_2^+ = u_{i+2}(n)\ ,\ \cdots\ ,\ u_{m+1-i}^+ = u_{m+1}(n))$$

is a $1 \times (m + 1 - i)$ vector. The second possible case is, if $m + n$ is odd and $med(x, y) = y_{[\sum_{k=1}^{i} u_k(n)+j]}$, then $\boldsymbol{U}_n^-$, a row vector with length $i + 1$, has the form

$$(u_1^- = j - 1\ ,\ u_2^- = u_i(n)\ ,\ \cdots\ ,\ u_{i+1}^- = u_1(n))$$

and $\boldsymbol{U}_n^+$, a row vector with length $m + 1 - i$, is given by

$$(u_1^+ = u_{i+1}(n) - j\ ,\ u_2^+ = u_{i+2}(n)\ ,\ \cdots\ ,\ u_{m+1-i}^+ = u_{m+1}(n)).$$

The third possible case is, if $m + n$ is even and $x_{[i]}$ is the smallest number larger than $med(x, y)$, the vectors are now defined as

$$\boldsymbol{U}_n^- = (u_1^- = u_i(n)\ ,\ u_2^- = u_{i-1}(n)\ ,\ \cdots\ ,\ u_i^- = u_1(n))$$

and

$$\boldsymbol{U}_n^+ = (u_1^+ = 0\ ,\ u_2^+ = u_{i+1}(n)\ ,\ \cdots\ ,\ u_{m+2-i}^+ = u_{m+1}(n)).$$

The last possibility is, if $m+n$ is even, $y_{\left[\sum_{k=1}^{i} u_k(n)+j\right]}$ is the smallest number larger than $med(x,y)$. The vectors are now defined as

$$\boldsymbol{U}_n^- = (u_1^- = j-1\ ,\ u_2^- = u_i(n)\ ,\ \cdots\ ,\ u_{i+1}^- = u_1(n))$$

and

$$\boldsymbol{U}_n^+ = (u_1^+ = u_{i+1}(n) - j + 1\ ,\ u_2^+ = u_{i+2}(n)\ ,\ \cdots\ ,\ u_{m+1-i}^+ = u_{m+1}(n)).$$

Let n^- be the length of the vector $\boldsymbol{U}_n^-$ and n^+ be the length of the vector $\boldsymbol{U}_n^+$.

Theorem 10. *$R_s(\mathbf{U}_n|\boldsymbol{X})$ is finite Markov chain imbeddable, and*

$$P(R_s(\mathbf{U}_n) = r|\boldsymbol{X}) = \boldsymbol{\xi}(\prod_{t=1}^{n} M_t)\boldsymbol{B}'(C_r),$$

where $\boldsymbol{B}(C_r) = \sum_{k:R_s(\mathbf{U}_n)=r} e_k$, e_k is a $1 \times \binom{m+n}{n}$ unit row vector corresponding to state $\mathbf{U}_n$, $\boldsymbol{\xi}(= P(Z_0 = 1) = 1)$ is the initial probability and M_t, $t = 1, \ldots, n$ are the transition probability matrices of the imbedded Markov chain defined on the state space Ω_t.

Proof. For each $\mathbf{U}_n$ in the state space Ω_n, we have a corresponding

$$\begin{aligned} R_s(\mathbf{U}_n|\boldsymbol{X}) &= R_s(\boldsymbol{U}_n^-|\boldsymbol{X}) + R_s(\boldsymbol{U}_n^+|\boldsymbol{X}) \\ &= \frac{\sum_{k=1}^{n^-}(u_k^-)^2 + \sum_{k=1}^{n^-} u_k^-}{2} + \sum_{k=1}^{n^- -1}(u_k^- + 1)(\sum_{j=k+1}^{n^-} u_j^-) \\ &+ \frac{\sum_{k=1}^{n^+}(u_k^+)^2 + \sum_{k=1}^{n^+} u_k^+}{2} + \sum_{k=1}^{n^+ -1}(u_k^+ + 1)(\sum_{j=k+1}^{n^+} u_j^+). \end{aligned} \tag{3.11}$$

The smallest possible value of $R_s(\mathbf{U}_n)$ is

$$r_{ss} = \begin{cases} \frac{n(n+2)}{4} & \text{if } m+n \text{ is even and } n \text{ is even} \\ \frac{(n+1)(n+3)}{4} & \text{if } m+n \text{ is even and } n \text{ is odd} \\ \frac{n^2}{4} & \text{if } m+n \text{ is odd and } n \text{ is even} \\ \frac{(n+1)(n-1)}{4} & \text{if } m+n \text{ is odd and } n \text{ is odd} \end{cases} \tag{3.12}$$

and the largest possible value is

$$r_{sb} = \begin{cases} \frac{n(2m+n+2)}{4} & \text{if } m+n \text{ is even and } n \text{ is even} \\ \frac{n(2m+n+2)-1}{4} & \text{if } m+n \text{ is even and } n \text{ is odd} \\ \frac{n(2m+n-1)}{4} & \text{if } m+n \text{ is odd and } n \text{ is even} \\ \frac{n(2m+n)-1}{4} & \text{if } m+n \text{ is odd and } n \text{ is odd} \end{cases} \tag{3.13}$$

In accordance with Equation (3.11), we use the possible value of R_s as a rule of the partition. The rest of the proof follows along the same line as that of Theorem 9, and here, is omitted.

□

Similarly, we apply the LLN to conclude that

$$\frac{1}{N}\sum_{i=1}^{N} P(R_s|\ \mathbf{X}_i\ ) \quad \xrightarrow{p} \quad P(R_s)$$

which establishes the distribution of R_s.

Through FMCI we, again, successfully retrieved the distribution of R_s under selected alternative distributions, for which the procedures are similar to those in the previous section. In addition, it is quite intuitive to approximate the power function by

$$\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{s=r_{ss}}^{s_{1\alpha}} P(R_s(\mathbf{U}_n)=s|\ \mathbf{X}_i) + \sum_{s=s_{2\alpha}}^{r_{sb}} P(R_s(\mathbf{U}_n)=s|\ \mathbf{X}_i)\right),$$

where

$$P(R_s(\mathbf{U}_n) \le s_{1\alpha}|H_o) + P(R_s(\mathbf{U}_n) \ge s_{2\alpha}|H_o) \le \alpha.$$

The following example is to illustrate our procedure.

Example 2. *Following the previous example, suppose that we have two samples with sizes $m = 2$ and $n = 3$ from a standard Normal distribution. By Equations (3.12) and (3.13), we calculate $r_{ss} = \frac{(3+1)(3-1)}{4} = 2$ and $r_{sb} = \frac{3((2)(2)+3)-1}{4} = 5$. Given a sample of $\mathbf{x} = \{-1.25, 0.8\}$, Table 3.6 shows the median for each possible $\mathbf{U}_n$, how $\mathbf{U}_n$ is split into $\mathbf{U}_{s-}$ and $\mathbf{U}_{s+}$, the corresponding value of $R_s(\mathbf{U}_n)$, and the conditional distribution of R_s. Once again, by applying the Law of Large numbers, we got the distribution of R_s in Table 3.7.*

Table 3.5: For $m = 2$ and $n = 3$ the list of $\mathbf{U}_n$, median, $\{\mathbf{U}_{s-}, \mathbf{U}_{s+}\}$, R_s and its conditional probability given $\mathbf{x} = \{-1.25, 0.8\}$

$\mathbf{U}_n$	$med(x, y)$	$\mathbf{U}_{s-}$	$\mathbf{U}_{s+}$	R_s	$P(\mathbf{U}_s\|\mathbf{x})$
(3, 0, 0)	$y_{[3]}$	(2)	(0, 0, 0)	3	0.0012
(2, 1, 0)	$x_{[1]}$	(2)	(1, 0)	4	0.0228
(2, 0, 1)	$x_{[1]}$	(2)	(0, 1)	5	0.0071
(1, 2, 0)	$y_{[2]}$	(0, 1)	(1, 0)	3	0.1476
(1, 1, 1)	$y_{[2]}$	(0, 1)	(0, 1)	4	0.0916
(1, 0, 2)	$x_{[2]}$	(0, 1)	(2)	5	0.0142
(0, 3, 0)	$y_{[2]}$	(1, 0)	(1, 0)	2	0.3179
(0, 2, 1)	$y_{[2]}$	(1, 0)	(0, 1)	3	0.2961
(0, 1, 2)	$x_{[2]}$	(1, 0)	(2)	4	0.0919
(0, 0, 3)	$y_{[1]}$	(0, 0, 0)	(2)	3	0.0095

Table 3.6: For $m = 2$ and $n = 3$ the conditional distribution of R_s given $\mathbf{x} = \{-1.25, 0.8\}$

R_s	$\mathbf{U}_n$'s	$P(R_s\|\mathbf{x})$
2	(0, 3, 0)	0.3179
3	(3, 0, 0), (0, 0, 3) , (1, 2, 0), (0, 2, 1)	0.5544
4	(2, 1, 0), (1, 1, 1) , (0, 1, 2)	0.2063
5	(2, 0, 1), (1, 0, 2)	0.0213

Table 3.7: For $m = 2$ and $n = 3$ the distribution of R_s

R_s	$\mathbf{U}_n$'s	$P(R_s)$
2	(0, 3, 0)	0.1000
3	(3, 0, 0), (0, 0, 3) , (1, 2, 0), (0, 2, 1)	0.4000
4	(2, 1, 0), (1, 1, 1) , (0, 1, 2)	0.3000
5	(2, 0, 1), (1, 0, 2)	0.2000

Note that the distribution of R_s when $F = G$ is not symmetric and can be obtained efficiently by Theorem 8. Because of this, one should be careful choosing the critical region when conducting a hypothesis testing.

3.3 Joint distributions of the rank statistics in the shift and scale case

We have derived the marginal distributions of R_l and R_s in terms of $\mathbf{U}_n$, respectively, which yield the following theorem.

Theorem 11. *$(R_l(\mathbf{U}_n|\boldsymbol{X}), R_s(\mathbf{U}_n|\boldsymbol{X}))$ is finite Markov chain imbeddable, and*

$$P(R_l(\mathbf{U}_n) = r_1; R_s(\mathbf{U}_n) = r_2|\boldsymbol{X}) = \boldsymbol{\xi}(\prod_{t=1}^{n} M_t)\boldsymbol{B}'(C_r)$$

where $\boldsymbol{B}(C_r) = \sum_{k:R_l(\mathbf{U}_n)=r_1 \ \& \ R_s(\mathbf{U}_n)=r_2} e_k$, e_k is a $1 \times \binom{m+n}{n}$ unit row vector corresponding to state $\mathbf{U}_n$, $\boldsymbol{\xi}(P(Z_0 = 1) = 1)$ is the initial probability and M_t, $t = 1, \ldots, n$ are the transition probability matrices of the imbedded Markov chain defined on the state space Ω_t.

Proof. By Equations (3.2) and (3.11), we know each $\mathbf{u}_n$ in the state space Ω_n has corresponding values of R_l and R_s. The combinations of the values R_l and R_s are used to be the standard of the partition. The rest of the proof follows along the same line as that of Theorem 9. □

Example 3. *Using the same setting, we show the conditional joint distribution of R_l and R_s given a sample of $\mathbf{x} = \{x_1, x_2\} = \{-1.25, 0.8\}$ from $N(0, 1)$ in Table 3.8*

Table 3.8: For $m = 2$ and $n = 3$ the conditional joint distribution of R_l and R_s given $\mathbf{x} = \{-1.25, 0.8\}$

$\mathbf{U}_n$'s	R_l	R_s	$P(R_l, R_s\|\mathbf{x})$
(3, 0, 0)	6	3	0.0012
(2, 1, 0)	7	4	0.0228
(1, 2, 0)	8	3	0.1476
(2, 0, 1)	8	5	0.0071
(0, 3, 0)	9	2	0.3179
(1, 1, 1)	9	4	0.0916
(1, 0, 2)	10	5	0.0142
(0, 2, 1)	10	3	0.2961
(0, 1, 2)	11	4	0.0919
(0, 0, 3)	12	3	0.0095

Table 3.9: For $m = 2$ and $n = 3$ the joint distribution of R_l and R_s

$\mathbf{U}_n$'s	R_l	R_s	$P(R_l, R_s)$
(3, 0, 0)	6	3	0.1000
(2, 1, 0)	7	4	0.1001
(1, 2, 0)	8	3	0.1001
(2, 0, 1)	8	5	0.0999
(0, 3, 0)	9	2	0.0999
(1, 1, 1)	9	4	0.1001
(1, 0, 2)	10	5	0.1000
(0, 2, 1)	10	3	0.1001
(0, 1, 2)	11	4	0.1000
(0, 0, 3)	12	3	0.1000

and the joint distribution of R_l and R_s in Table 3.9 approximated using a simulation of size ten million. Again, these can be obtained by Theorem 8.

The joint distribution of the ranks considering both the location and scale parameters which can be determined through our algorithm is yet to be studied in the literature. Our result allows us to test the homogeneity of the distribution functions $F(x) = G((x - \theta)\sigma^{-1})$. We state the hypotheses as follows

$$H_o : \theta = 0 \ \text{ and } \ \sigma = 1 \ \text{ v.s. } \ H_a : \theta \neq 0 \ \text{ or } \ \sigma \neq 1. \tag{3.14}$$

Also we are able to decide a critical region under the null hypothesis and discuss its power when $F \neq G$. For example, a rectangular critical region can be

$$C_\alpha = \{R_l \leq r_{1l}, \ R_l \geq r_{2l}, \ R_s \leq r_{1s} \quad \text{or} \quad R_s \geq r_{2s}\}$$

where r_{1l}, r_{2l}, r_{1s} and r_{2s} are the critical values such that

$$P(R_l \leq r_{1l}|H_o) + P(R_l \geq r_{2l}|H_o) + P(r_{1l} < R_l < r_{2l}, R_s \leq r_{1s}|H_o) \\ + P(r_{1l} < R_l < r_{2l}, R_s \geq r_{2s}|H_o) \leq \alpha$$

or an elliptic critical region

$$C'_\alpha = \left\{ \frac{R_l^2}{a} + \frac{R_s^2}{b} > C \right\}$$

for some positive constants a and b such that

$$P\left(\frac{R_l^2}{a} + \frac{R_s^2}{b} > C|H_o \right) \leq \alpha.$$

According to the above defined rejection region, the power of the test can be found as

$$P(R_l \leq r_{1l}|H_a) + P(R_l \geq r_{2l}|H_a) + P(r_{1l} < R_l < r_{2l}, R_s \leq r_{1s}|H_a) \\ + P(r_{1l} < R_l < r_{2l}, R_s \geq r_{2s}|H_a) \qquad (3.15)$$

or

$$P\left(\frac{R_l^2}{a} + \frac{R_s^2}{b} > C|H_a \right). \qquad (3.16)$$

Note that unless having a conjecture about the values of θ and σ, we tend to use a two-sided test. However, with the knowledge of the centre and shape of the distribution of interest, deciding a sectorial critical region is a better choice, for which an example is demonstrated in the numerical studies.

3.4 Numerical Studies

3.4.1 A joint distribution of R_l and R_s

Let $\{X_1, \ldots, X_5\} \sim N(0, 1)$ and $\{Y_1, \ldots, Y_7\} \sim N(\theta, \sigma)$. Figure 3.1 gives the joint distribution of the random variables R_l and R_s under the null hypothesis of $\theta = 0$ and $\sigma = 1$. The marginal distributions of R_l and R_s can be easily established from their joint distribution, see Figures 3.2 and 3.3. Figure 3.1 also shows that, given the largest possible value of R_l, there is only one possible value of R_s; given R_l a value close to the centre of the distribution, there is a higher probability of having R_s closer to the mean value and less probability of being an extreme value; and given the smallest possible value of R_l, there is also one possible value of R_s, which implies that the two random variables R_l and R_s are dependent. We construct two critical regions as shown in Figures 3.4 and 3.5, according to their joint distribution. The rows are the possible values of R_l from 28 to 63; the columns are the possible values of R_s from 16 to 33; and the numbers in the cell are the probabilities. Outside the yellow area in Figure 3.4 is the selected rectangular critical region $C_{0.1738}$ and outside the red shadow in Figure 3.5 is the elliptic one $C'_{0.1738}$.

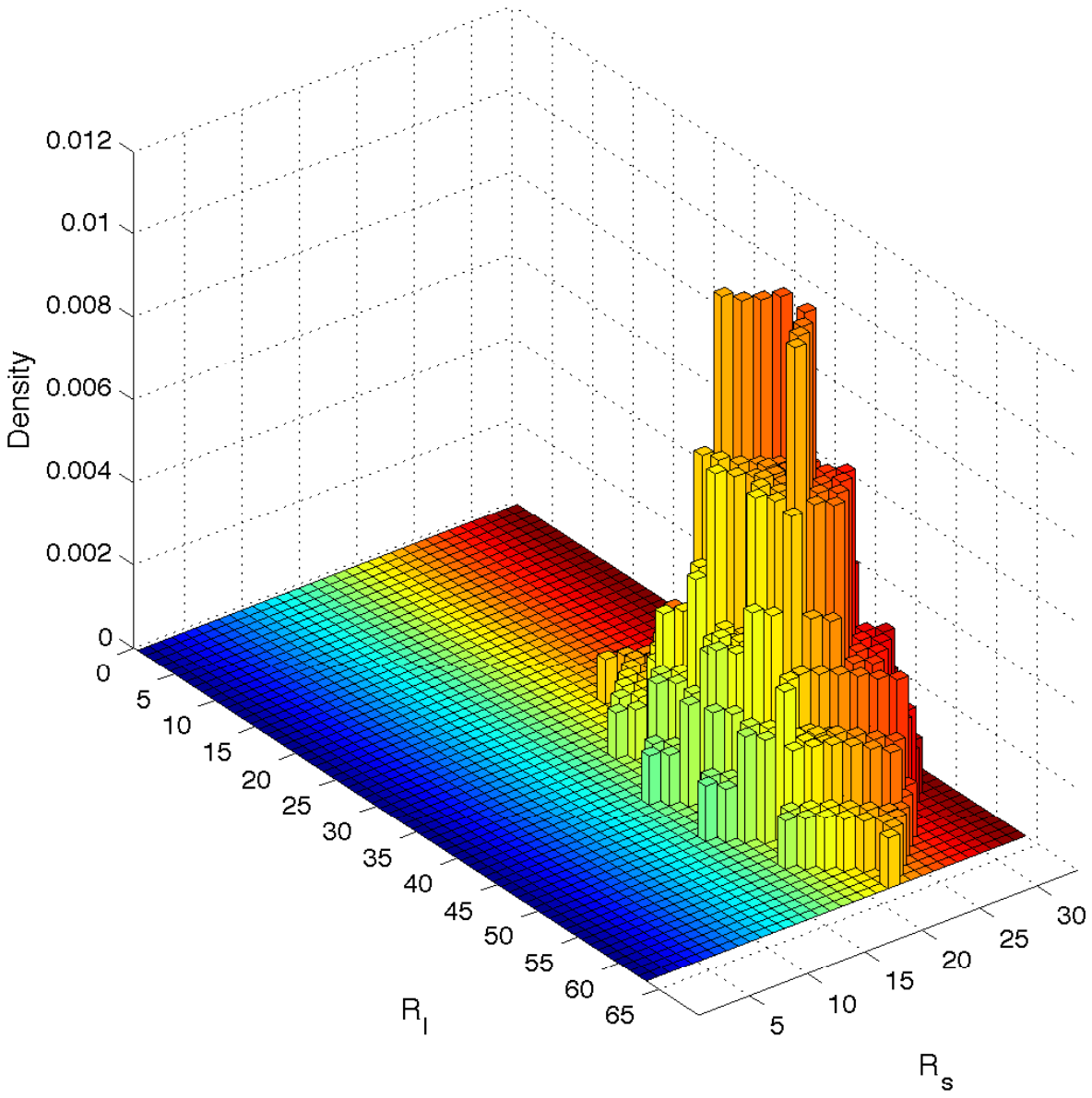

Figure 3.1: Joint distribution of R_l and R_s in the case where $m = 5$, $n = 7$ and $F = G \sim N(0, 1)$.

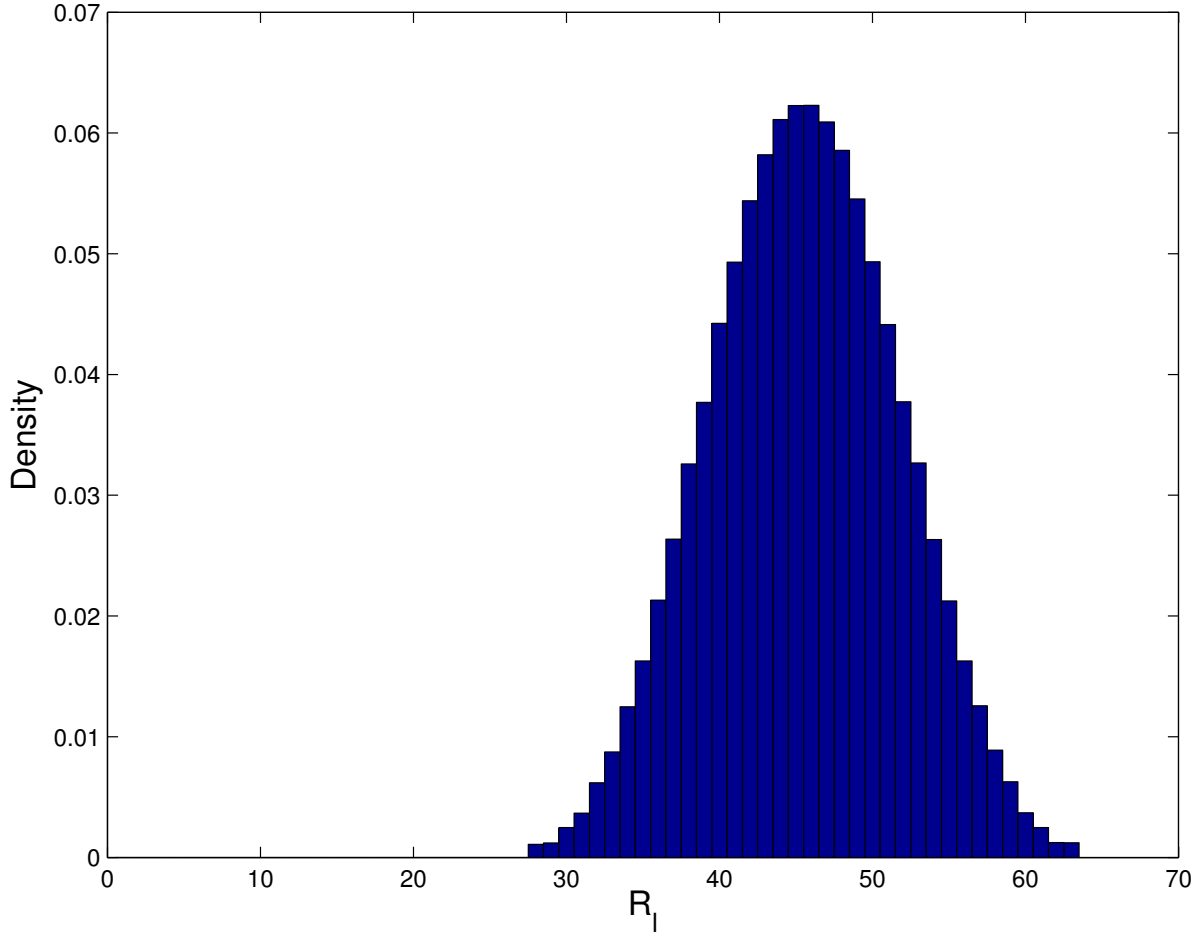

Figure 3.2: Marginal distribution of R_l in the case where $m = 5$, $n = 7$ and $F = G \sim N(0, 1)$.

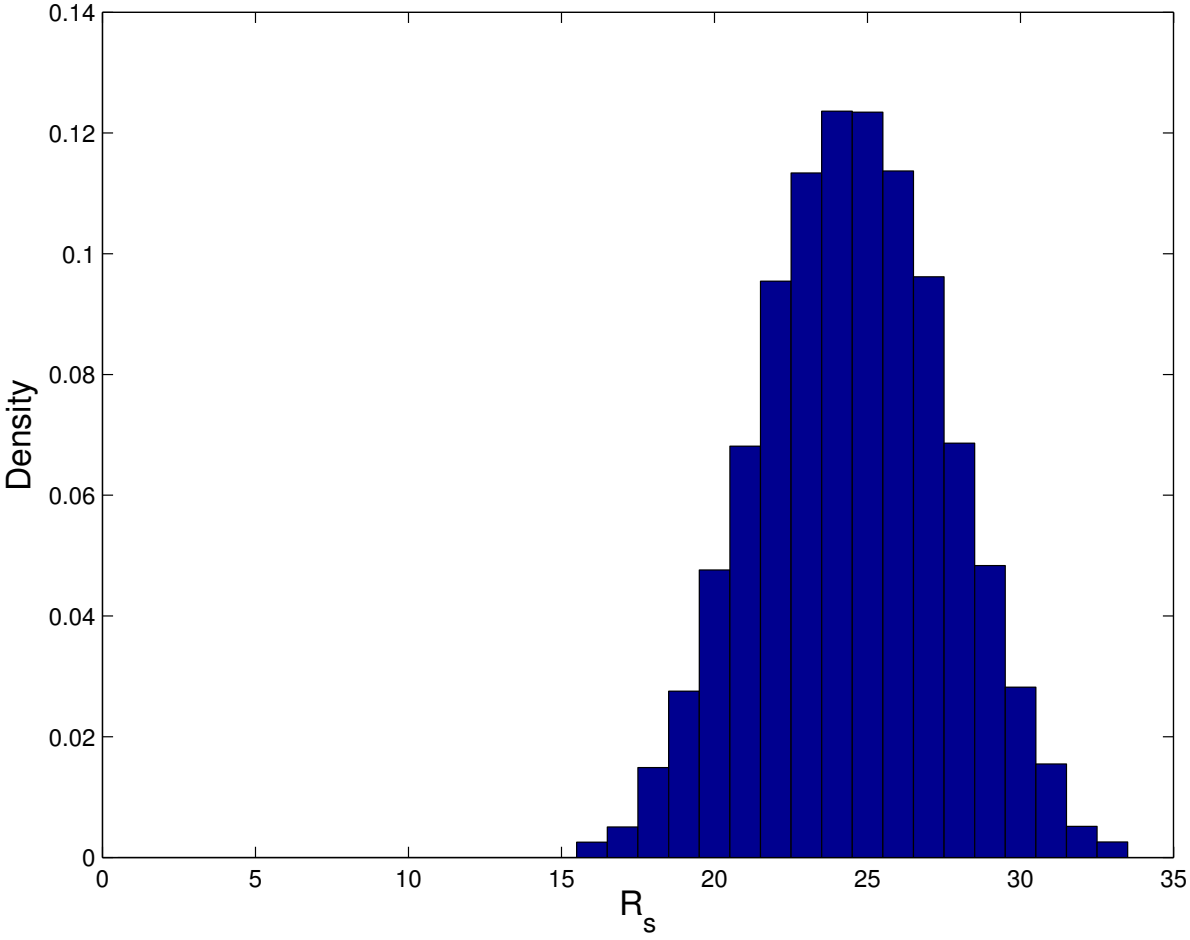

Figure 3.3: Marginal distribution of R_s in the case where $m = 5$, $n = 7$ and $F = G \sim N(0, 1)$.

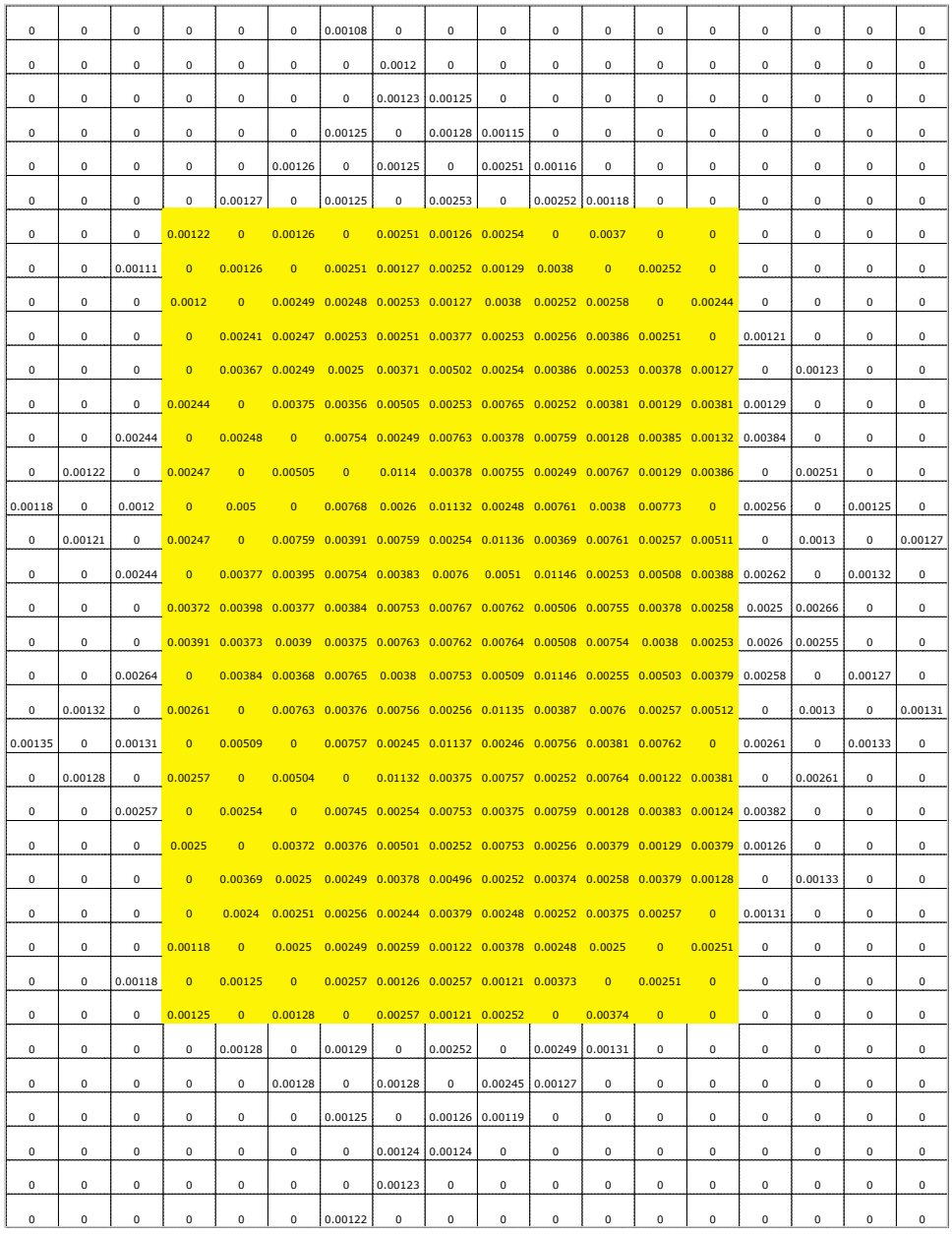

Figure 3.4: Rectangular critical region at size 17.38% for R_l and R_s in the case where $m = 5$ and $n = 7$.

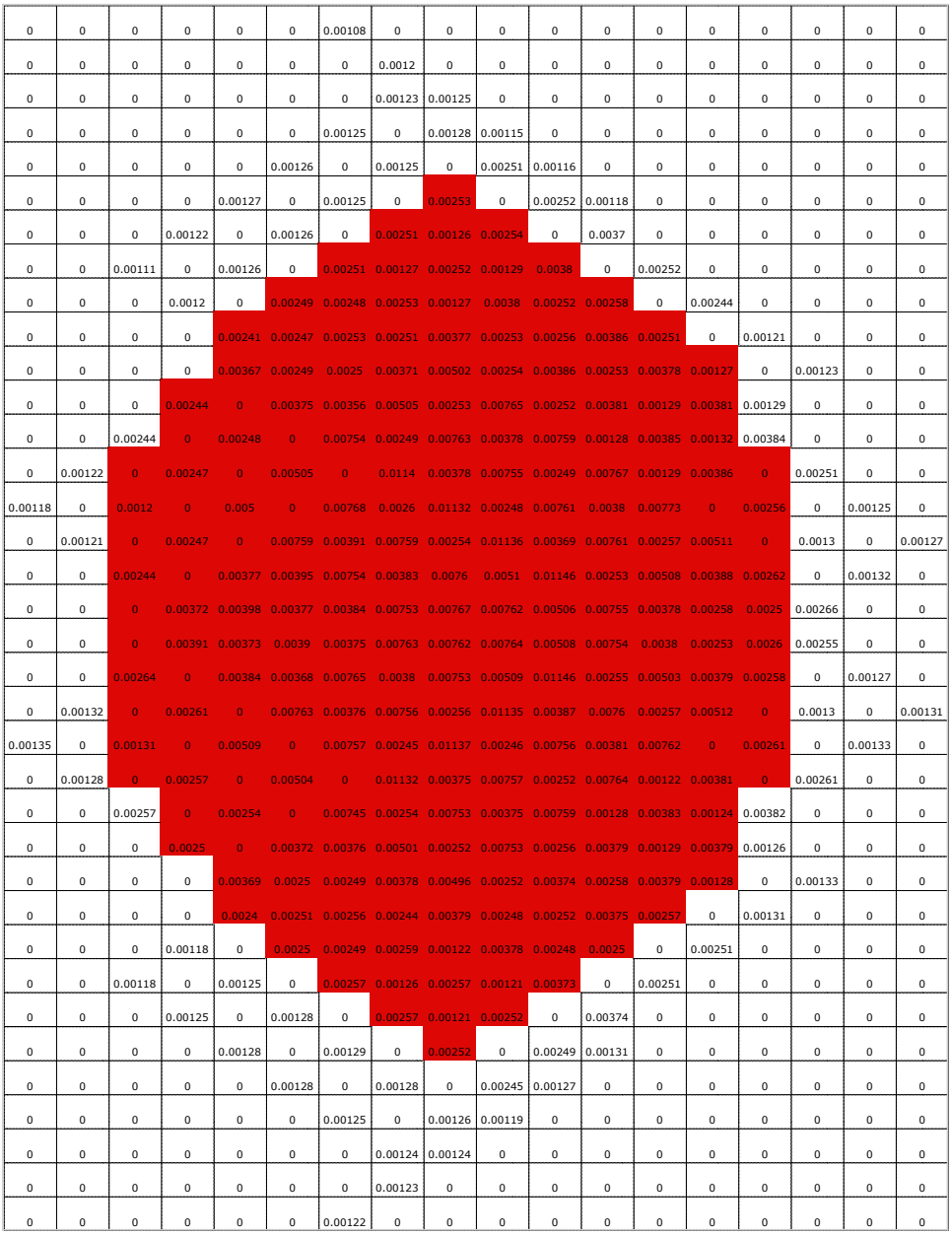

Figure 3.5: Elliptic critical region at size 17.38% for R_l and R_s in the case where $m = 5$ and $n = 7$.

3.4.2 Powers for a joint test using R_l and R_s

The alternative of interest is stated in the preceding section (see Equation (3.14)). The power functions of the test statistics R_l and R_s for a sequence of normally distributed populations with θ from -20 to 20 with an increment of 0.5 and σ from 1 to 10 with an increment of 1, and its reciprocal under two types of critical regions are provided in Figures 3.6 and 3.7. We adopt a two-sided test because of the selected values of the parameters. It should be slightly modified the critical region in the previous step in order to calculate the powers if a one-sided test is adopted. Both critical regions roughly perform equally well as shown in Figures 3.6 and 3.7. Figure 3.8 presents the performance of the two critical regions for given various parameter settings. Figures 3.8 and 3.9 show that given a standard deviation of 1 or a mean of 0, the powers of the two critical regions, rectangular and elliptic, are high and similar. However, when the variation of the alternative population reduces ($\sigma = 1/10$) or increases ($\sigma = 10$), the elliptic critical region performs better than the rectangular one as shown in Figures 3.10 and 3.11. Therefore, we suggest that when conducting a test for the equivalence of two distributions, an elliptic rejection area should be used.

Next, we consider the problem of determining an optimum rank test. To conduct a test of distributions equivalency, we can use either R_l or R_s as the test statistic. As mentioned earlier, the marginal distribution R_l or R_s can be easily established from their joint distribution. Figures 3.12 and 3.13 provide the power functions for the test statistics R_l and R_s at the level of significance 17.38%, respectively. Figure 3.13 shows that the rank test against scale parameter is badly effected by the centre of the alternative population. This was seen before by Ansari and Bradley (1960).

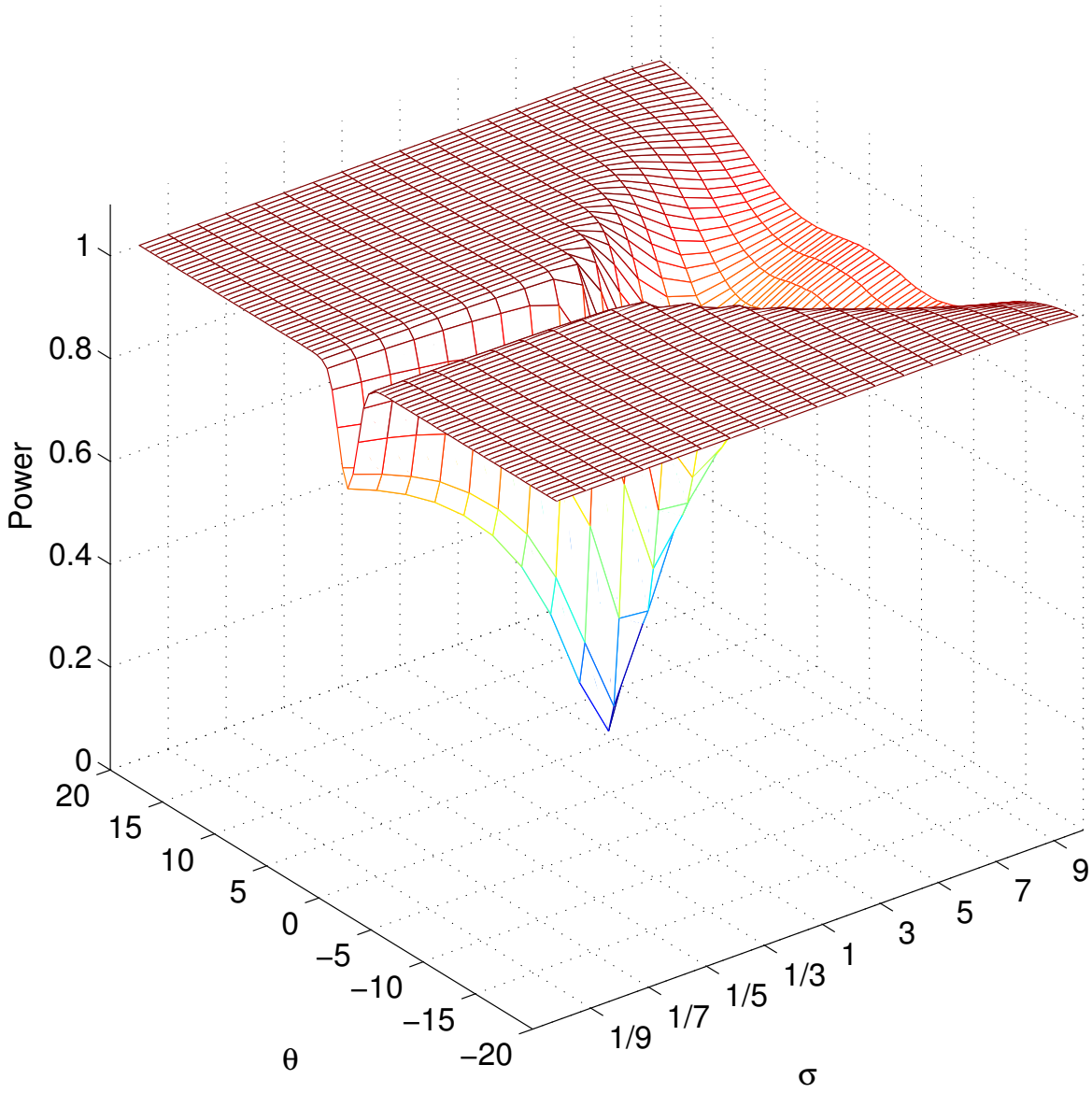

Figure 3.6: Power functions of R_l and R_s in the case where $m = 5$ and $n = 7$ under C_α.

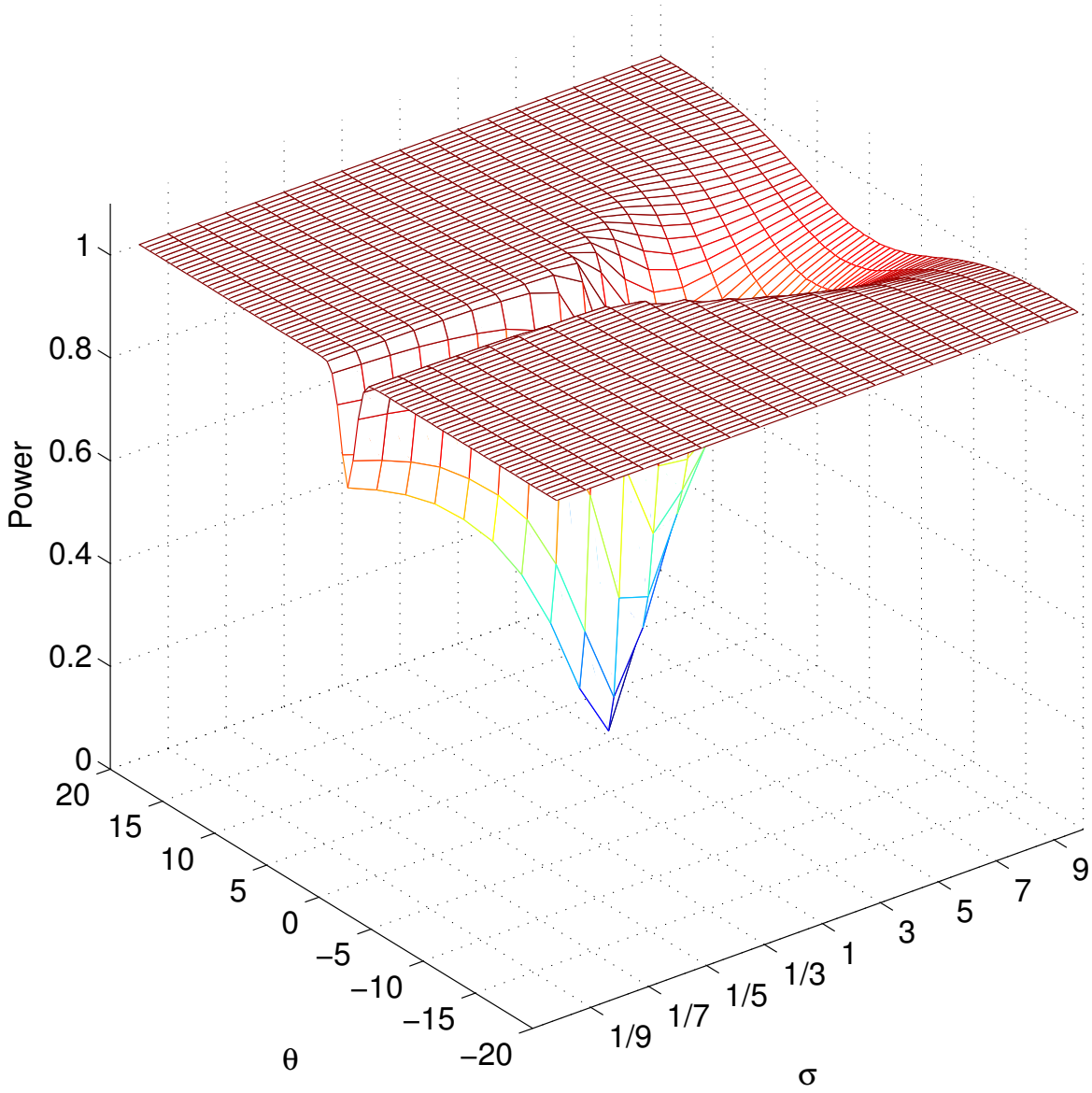

Figure 3.7: Power functions of R_l and R_s in the case where $m = 5$ and $n = 7$ under C'_α.

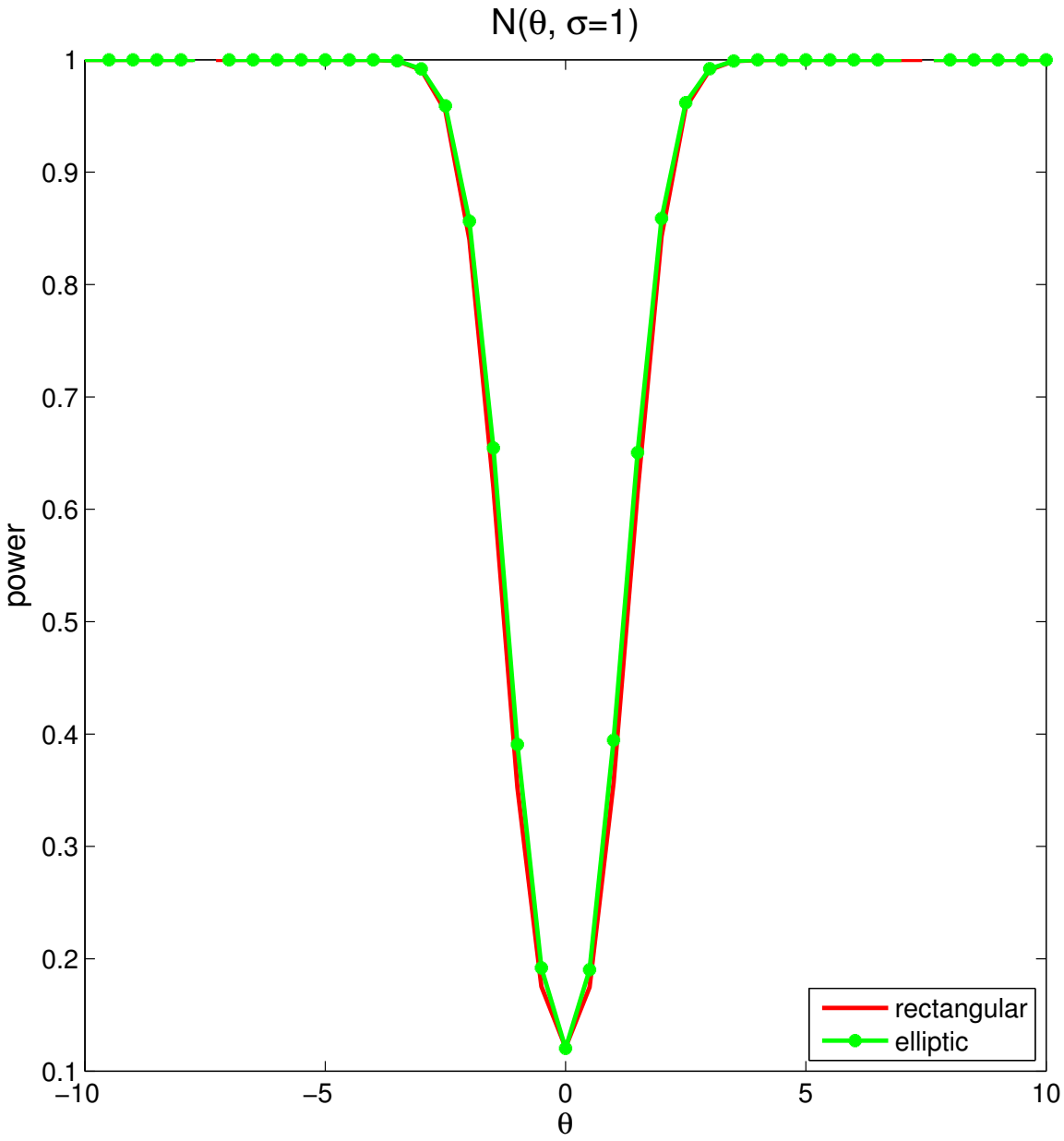

Figure 3.8: Power comparisons of the joint test R_l and R_s in the case where $m = 5$, $n = 7$ and $\sigma = 1$.

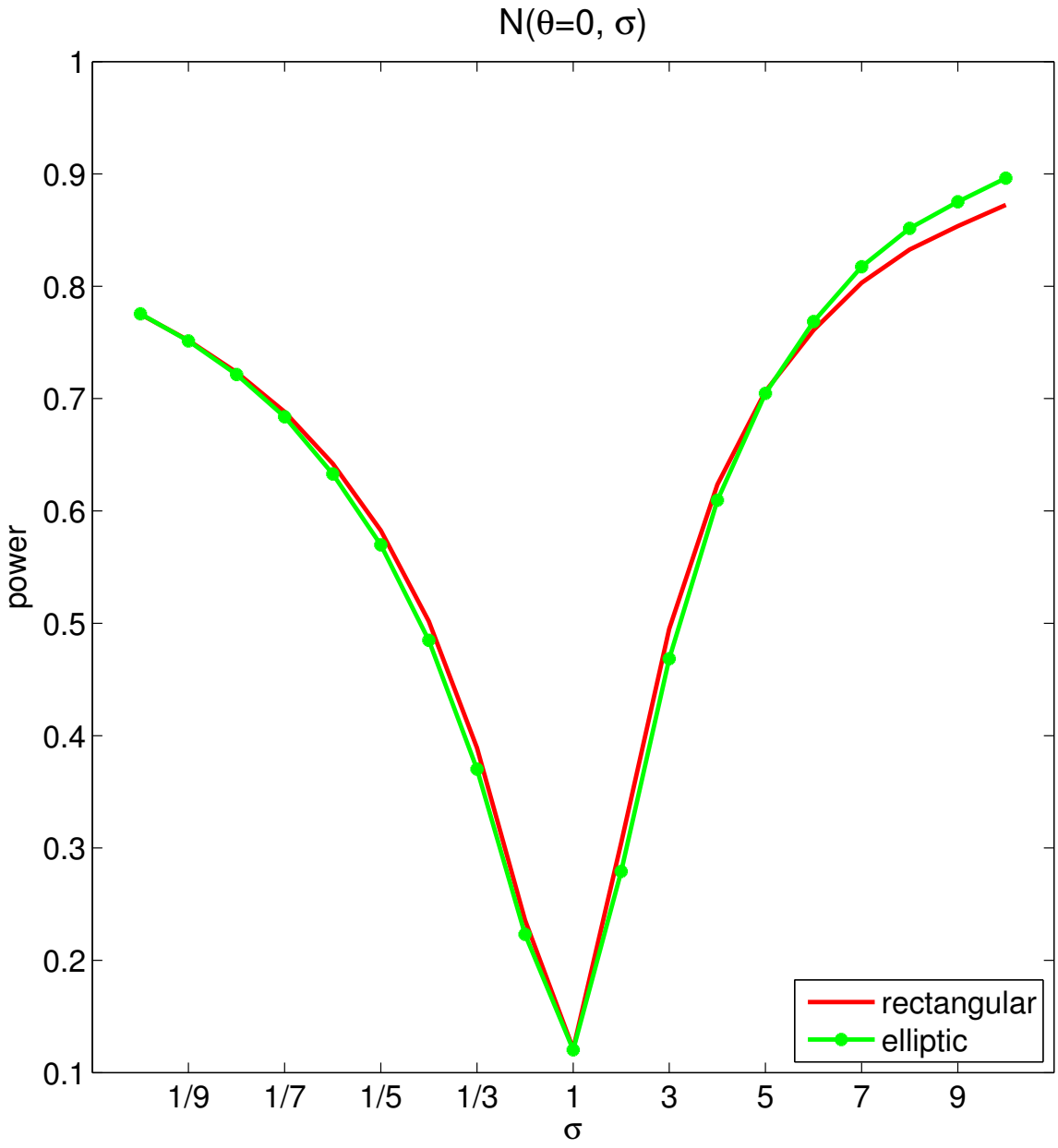

Figure 3.9: Power comparisons of the joint test R_l and R_s in the case where $m = 5$, $n = 7$ and $\theta = 0$.

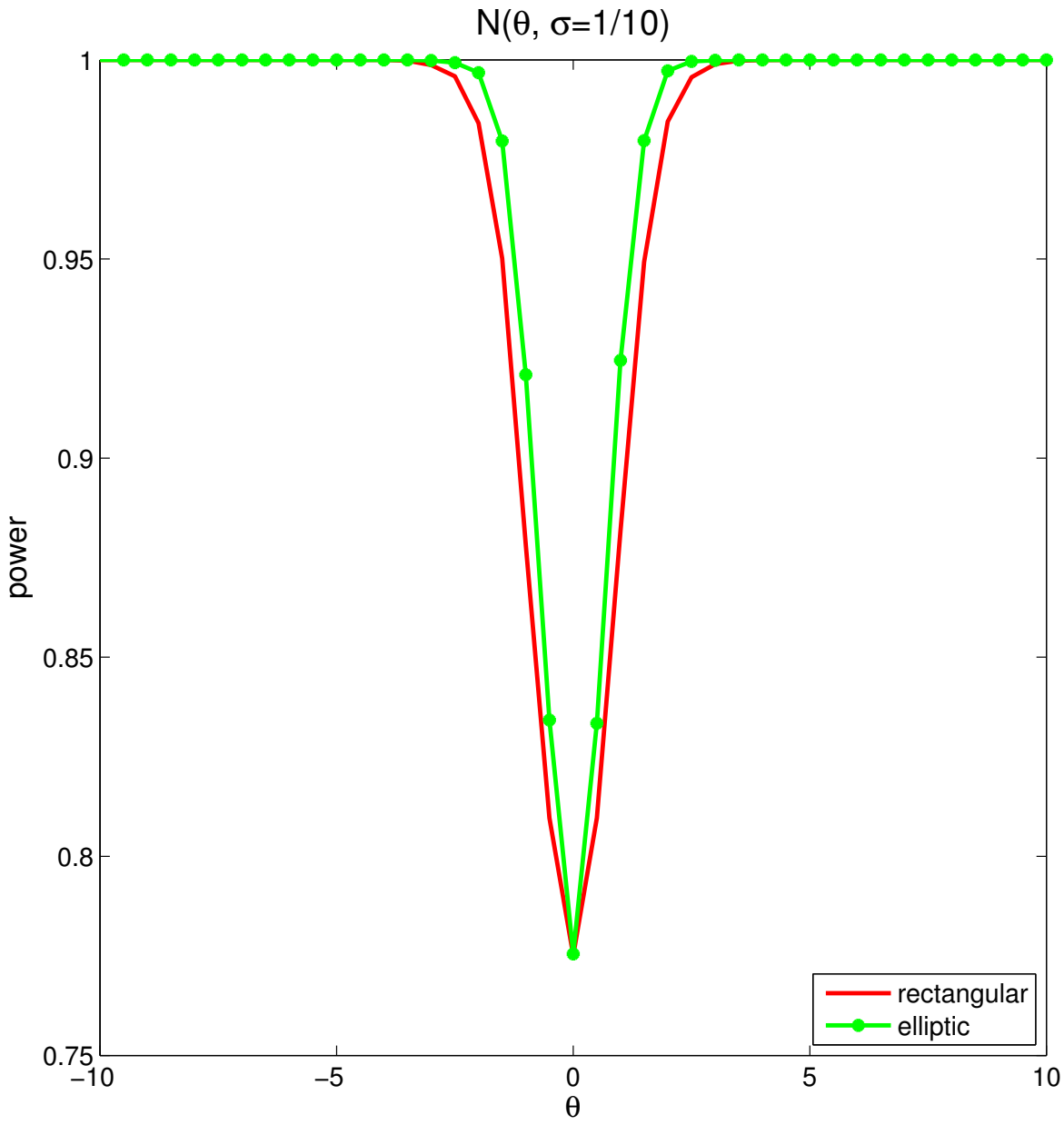

Figure 3.10: Power comparisons of the joint test R_l and R_s in the case where $m = 5$, $n = 7$ and $\sigma = 1/10$.

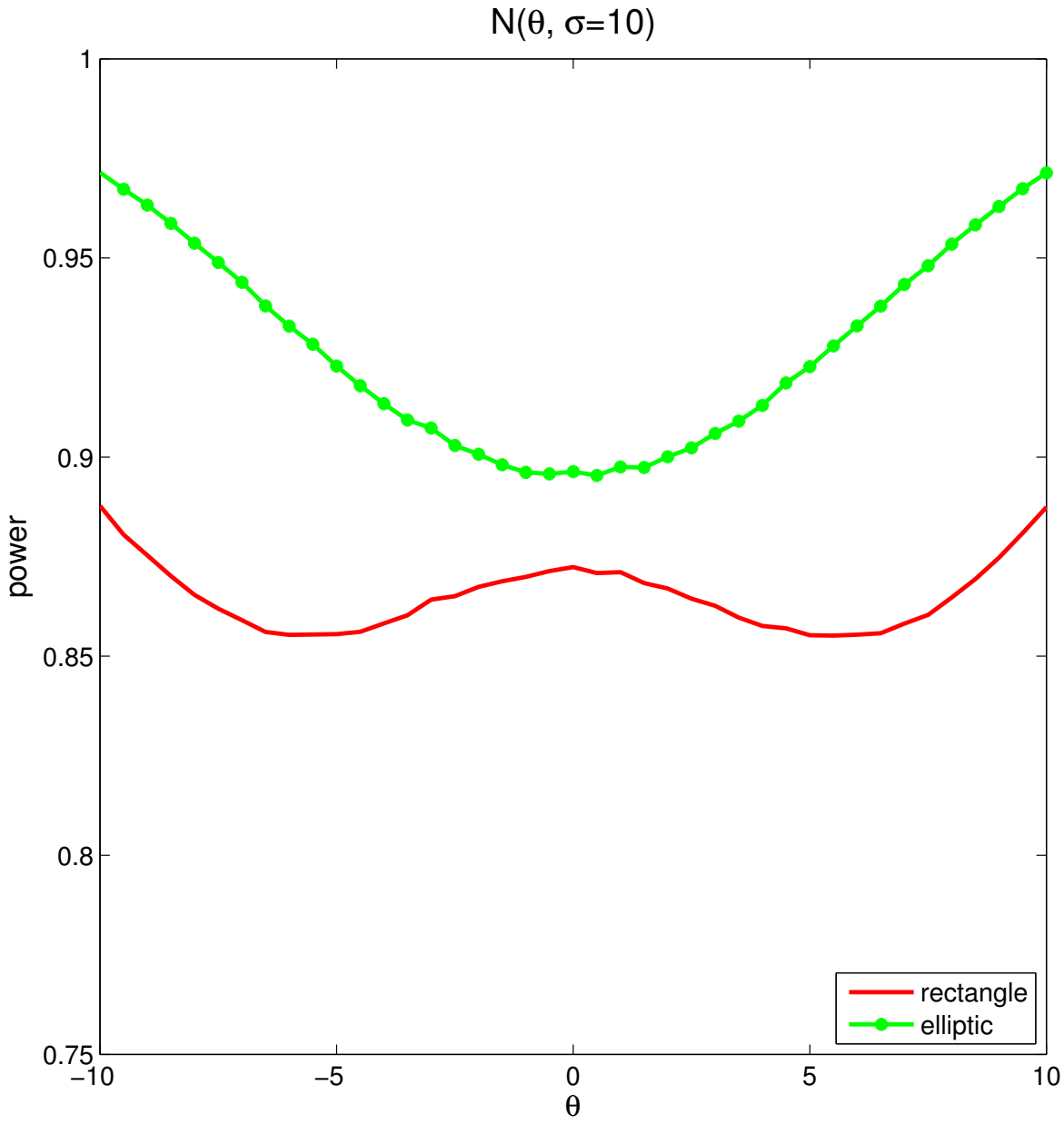

Figure 3.11: Power comparisons of the joint test R_l and R_s in the case where $m = 5$, $n = 7$ and $\sigma = 10$.

By comparing Figures 3.12 and 3.13 with Figure 3.7, it seems that the joint test would be much more reliable than either R_l or R_s alone for distributions equivalence tests. A joint test for distributions equivalency would like a better option under most circumstances.

3.4.3 Lehmann alternatives

Consider the one-sided alternative $F(x;\theta,\sigma) > G(x;\theta,\sigma)$. Lehmann (1953) proposed a test of $H_o : F(x;\theta,\sigma) = G(x;\theta,\sigma)$ against $H_a : F(x;\theta,\sigma)^k = G(x;\theta,\sigma)$ which is known as the family of Lehmann alternative. Note $F(x;\theta,\sigma)^k$ is the cumulative distribution of $\max_{1\le i\le k}(x_i)$ when $X_i \sim F$ and, under the alternative hypothesis, $G(x;\theta,\sigma)$ is stochastically larger than $F(x;\theta,\sigma)$. First of all, we know

$$\begin{aligned}\mathbb{E}_k(X) &= \int_{-\infty}^{0}[-G(x)]\,dx + \int_{0}^{\infty}[1-G(x)]\,dx \\ &> \int_{-\infty}^{0}[-F(x)]\,dx + \int_{0}^{\infty}[1-F(x)]\,dx = \mathbb{E}(X). \end{aligned} \tag{3.17}$$

Therefore, the larger the R_l is, the stronger the evidence against the null hypothesis will be. For the variation of the distribution per se, the codomain of the density function is compressed to larger numbers; therefore, in most cases, we have $\mathbb{V}\mathrm{ar}(X_k) < \mathbb{V}\mathrm{ar}(X)$. We then propose to reject the null hypothesis when R_s is large. For example, given $F \sim U(0,1)$ and $G = F^k$, it is easy to see

$$\frac{\mathbb{E}_{k+1}(X)}{\mathbb{E}_k(X)} = \frac{(k+1)^2}{k(k+2)} > 1 \tag{3.18}$$

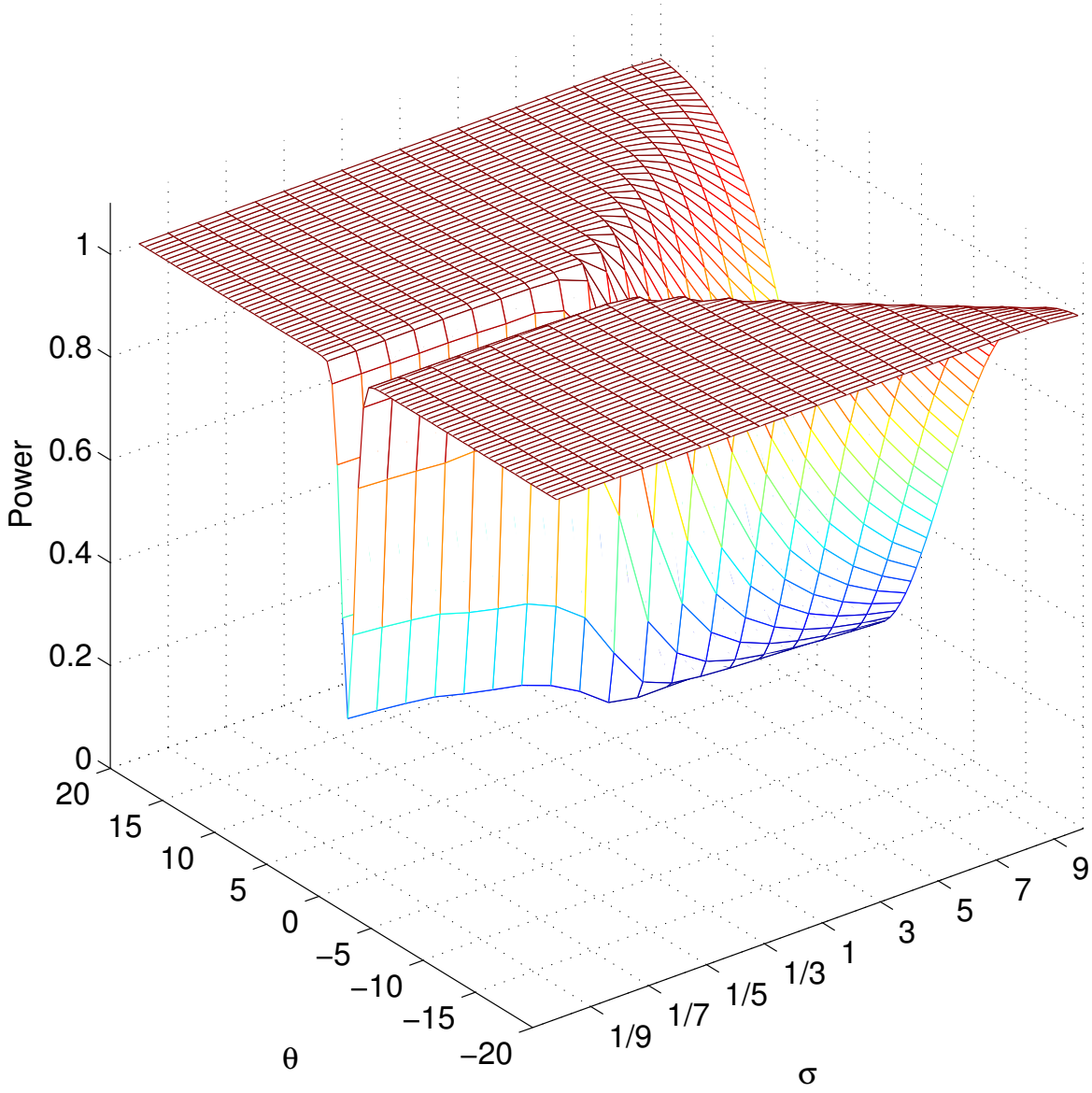

Figure 3.12: Power functions of R_l in the case where $m = 5$ and $n = 7$.

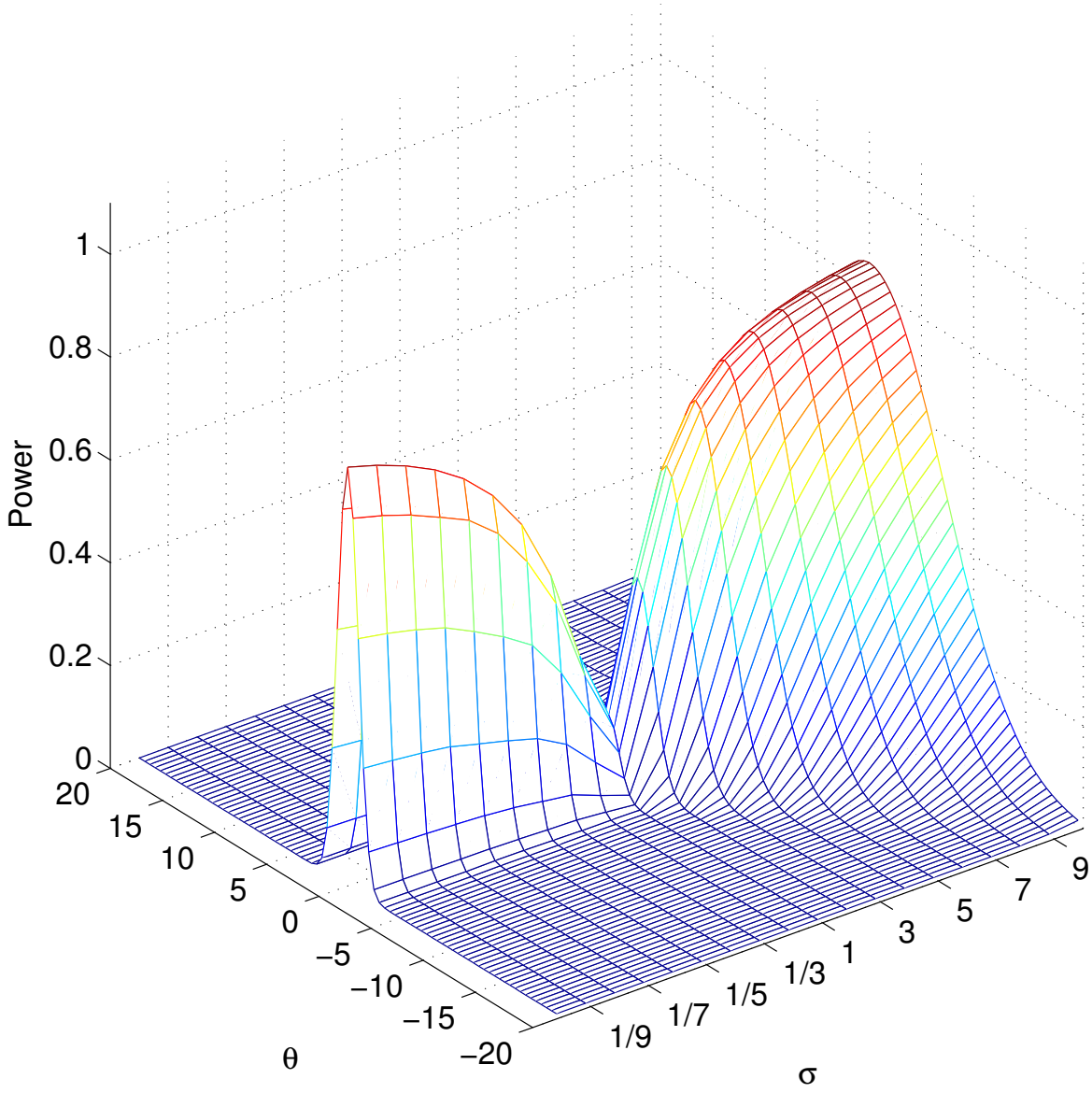

Figure 3.13: Power functions of R_s in the case where $m = 5$ and $n = 7$.

and

$$\frac{\mathbb{V}\mathrm{ar}_{k+1}(X)}{\mathbb{V}\mathrm{ar}_{k}(X)} = \frac{(k+1)^3}{k(k+2)(k+3)} < 1 \tag{3.19}$$

for all k. We first find the marginal and joint distributions of the ranks R_l and R_s in order to define critical regions for R_l and R_s individually and simultaneously. Due to the properties of the mean and variance of the alternative distribution, as shown in Equations (3.17), (3.18) and (3.19), we are cautious to define the critical regions. Table 3.10 provides powers for the tests as we choose uniform, standard Normal, student-t with 3 degrees of freedom, exponential distributions for the hypothesized distribution, a couple of different settings for sample sizes m and n, and 2, 3, 6 for k. Clearly, a joint test considering both R_l and R_s for the equality of distributions is best suited in comparison with tests considering only one of the rank statistics.

Table 3.10: Power comparisons for a one-sided rank test
$H_0 : F(x;\theta_o,\sigma_o) = G(x;\theta_a,\sigma_a)$ v.s. $H_a : F^k(x;\theta_o,\sigma_o) = G(x;\theta_a,\sigma_a)$.

		$m=6 \quad n=10$				$m=10 \quad n=10$				$m=10 \quad n=20$			
F	Test	$\beta(F)$	$\beta(F^2)$	$\beta(F^3)$	$\beta(F^6)$	$\beta(F)$	$\beta(F^2)$	$\beta(F^3)$	$\beta(F^6)$	$\beta(F)$	$\beta(F^2)$	$\beta(F^3)$	$\beta(F^6)$
$U(0,1)$	R_l	.090	.411	.647	.900	.096	.496	.761	.967	.099	.591	.845	.984
	R_s	.080	.152	.193	.218	.076	.137	.149	.123	.100	.236	.370	.638
	$R_l \& R_s$	.100	.452	.699	.934	.100	.531	.799	.981	.100	.622	.878	.992
$N(0,1)$	R_l	0.090	.404	.647	.897	0.096	.494	.760	.968	0.099	.585	.839	.985
	R_s	0.080	.152	.190	.220	0.076	.137	.154	.120	0.100	.236	.367	.634
	$R_l \& R_s$	0.100	.447	.702	.933	0.100	.527	.801	.983	0.100	.616	.870	.991
$t(3)$	R_l	0.090	.412	.639	.897	0.096	.493	.756	.965	0.099	.574	.841	.987
	R_s	0.080	.150	.197	.217	0.076	.137	.152	.121	0.100	.234	.367	.634
	$R_l \& R_s$	0.100	.453	.696	.932	0.100	.528	.798	.980	0.100	.606	.874	.993
$Exp(1)$	R_l	0.090	.411	.650	.899	0.096	.490	.764	.967	0.099	.579	.841	.987
	R_s	0.080	.149	.195	.217	0.076	.140	.152	.122	0.100	.232	.376	.641
	$R_l \& R_s$	0.100	.451	.702	.933	0.100	.525	.805	.982	0.100	.607	.875	.993

Note: A sectorial critical region is chosen for a simultaneous testing.

Chapter 4

Scan Statistics

4.1 Distributions of the scan statistics

We first use a simple example to show how the distribution of a scan statistic $S_n(r)$ can be derived from a waiting time random variable of a compound pattern. Given the window size $r = 4$ and number of successes $s = 3$, the event $S_n(4) < 3$ does not occur in a sequence of n bistate trials if and only if zero occurrence of the three simple patterns $\Lambda_1 = SSS$, $\Lambda_2 = SFSS$ and $\Lambda_3 = SSFS$ is observed in the sequence. Let $\Lambda_{4,3} = \bigcup_{i=1}^{3} \Lambda_i$ be the compound pattern and $W(\Lambda_{4,3})$ be the waiting time of the compound pattern, then we have

$$P(S_n(4) < 3) = P(W(\Lambda_{4,3}) > n).$$

One key point to apply the dual relationship between the events $\{S_n(r) < s\}$ and $\{W(\Lambda_{r,s}) > n\}$ is to find all the simple patterns with a length less than or equal to r that starts and ends with a success, and associates with the event $\{S_n(r) = s\}$. Fu (2001) showed that the number of simple patterns associated with the event $\{S_n(r) = s\}$ is

$$l = \sum_{i=0}^{r-s} \binom{s-2+i}{i}.$$

Therefore, given r and s, we are able to define l distinct simple patterns associated with the scan statistic. Let

$$\boldsymbol{F}_{r,s} = \{\Lambda_i, i = 1, \ldots, l\}$$

be a collection of simple patterns and

$$\Lambda_{r,s} = \bigcup_{i}^{l} \Lambda_i$$

be the compound pattern; $S(\Lambda_i)$ be a collection of all sub-patterns in Λ_i and $\Lambda_i \in \boldsymbol{F}_{r,s}$. For instance, given $r = 4$ and $s = 3$,

$$\boldsymbol{F}_{4,3} = \{SSS, SFSS, SSFS\}$$

and

$$\bigcup_{i=1}^{3} S(\Lambda_i) = \{SS, SF, SFS, SSF, SSS, SFSS, SSFS\}.$$

4.1.1 Sequence of k-block trials

Suppose there are some consecutive bistate trials, each of which has the same probability of observing a success event. These consecutive bistate trials are said

to form a block. In this setting, the sequence of n identical and independent trials forms only one block. If the chance of success changes after the i^{th} trial, we have the chance of success p_1 for the first i trials and p_2 for the remaining $n-i$ trials. We then call $\{X_i\}$ a sequence of 2-block trials.

Theorem 12. *Let $X_1, X_2, \ldots, X_n$ be a sequence of k-block independent bistate trials, m_t trials with chance of success $\pi(t)$ in the t^{th} block for $t = 1, 2, \ldots, k$ and $\sum_{t=1}^{k} m_t = n$. Then, the waiting time random variable $W(\Lambda_{r,s})$ for the compound pattern $\Lambda_{r,s}$ associated with the scan statistic $S_n(r)$ is finite heterogeneous Markov chain imbeddable on the state space*

$$\Omega = \{\emptyset, S, F\} \bigcup_{i=1}^{l} S(\Lambda_i) \ \ for \ \ \Lambda_i \in \boldsymbol{F}_{r,s} \tag{4.1}$$

with a transition probability matrix

$$\boldsymbol{M}_{r,s}(\pi(t)) = \left[\begin{array}{c|c} \boldsymbol{N}_{r,s}(\pi(t)) & \boldsymbol{C}_{r,s}(\pi(t)) \\ \hline \boldsymbol{O} & \boldsymbol{I} \end{array}\right]$$

where $\boldsymbol{F}_{r,s}$ is a collection of simple patterns associated with the event $S_n(r) = s$, $S(\Lambda_i)$ is a collection of all sub-patterns in Λ_i and $\Lambda_i \in \boldsymbol{F}_{r,s}$ and $\pi(t)$ is the probability of observing a success in the t^{th} block. For $k = n$,

$$P(W(\Lambda_{r,s}) > n) = \xi_0 \prod_{t=1}^{k} \boldsymbol{N}_{r,s}(\pi(t))^{m_t} \mathbf{1}' \tag{4.2}$$

where $\xi_0 = (1, 0, \ldots, 0)$ *is one row vector and* $\mathbf{1}'$ *is the transpose of the row vector* $\mathbf{1} = (1, 1, \ldots, 1)$.

Proof. As Fu (2001) pointed out,

$$S_n(r) < s \overset{1-1}{\Longleftrightarrow} W(\Lambda_{r,s}) > n \quad \text{for any} \quad 1 \leq s \leq r \leq n.$$

Suppose the chance of observing a success at block t is $\pi(t)$, the transition probability matrix $\boldsymbol{M}(\pi(t))$ for a random variable of waiting time of a compound pattern can be arranged in the form

$$\boldsymbol{M}(\pi(t)) = \begin{matrix} \Omega - \boldsymbol{F}_{r,s} \\ \boldsymbol{F}_{r,s} \end{matrix} \left[\begin{array}{c:c} \boldsymbol{N}_{r,s}(\pi(t)) & \boldsymbol{C}_{r,s}(\pi(t)) \\ \hdashline \boldsymbol{O} & \boldsymbol{I} \end{array}\right].$$

Equation (4.2) follows directly. □

Let π_i be the chance of observing a success for i^{th} trial. To test the null hypothesis of a constant probability of a success against the alternative of k-block independent trials, we thus state

$$H_o : \pi_i = \pi \ \text{ for } \ i = 1, 2, \ldots, n$$

versus

$$H_a : \pi_i = \pi(t) \quad \text{if} \quad \sum_{i=1}^{t-1} m_i + 1 \leq i \leq \sum_{i=1}^{t} m_i.$$

By Equation (4.2) we are able to determine a critical region s_α such that

$$P(S_n(r) \geq s_\alpha | H_o) = P(W(\Lambda_{r,s_\alpha}) \leq n | H_o)$$
$$= 1 - \xi_0 \boldsymbol{N}_{r,s_\alpha}(\pi)^n \mathbf{1}' \quad \leq \alpha$$

and the power function can be found as

$$P(S_n(r) \geq s_\alpha | H_a) \qquad = P(W(\Lambda_{r,s_\alpha}) \leq n | H_a)$$
$$= 1 - \xi_0 \prod_{t=1}^{k} \boldsymbol{N}_{r,s_\alpha}(\pi(t))^{m_t} \mathbf{1}'. \tag{4.3}$$

Remark 1. *The scan statistic discussed by Fu (2001) is a special case of this theorem when $k = 1$.*

Remark 2. *When $k = 2$, the distribution of $S_n(r)$ for $m_1 = m$, $\pi(1) = p_1$, and $\pi(2) = p_2$ will be the same as the distribution of $S_n(r)$ for $m_1 = n - m$, $\pi(1) = p_2$ and $\pi(2) = p_1$.*

Remark 3. *The power, addressed by Wallenstein et al. (1994), for a scan statistic to detect a changed segment in a Bernoulli sequence is also a special case of this theorem when (i) $k = 2$ and $m_1 = Ir$, (ii) $k = 2$ and $m_2 = Ir$, or (iii) $k = 3$ and $m_2 = Ir$.*

4.1.2 Sequence of Markov dependent trials

Theorem 13. *Suppose $\{X_i\}_{i=1}^n$ is a sequence of homogeneous Markov dependent bistate trials with transition probability matrix*

$$\boldsymbol{A} = \begin{bmatrix} p_{SS} & p_{SF} \\ p_{FS} & p_{FF} \end{bmatrix}.$$

Then, the waiting time random variable $W(\Lambda_{r,s})$ for the compound pattern $\Lambda_{r,s}$ associated with the scan statistic $S_n(r)$ is finite homogeneous Markov chain imbeddable on the state space Ω, presented in (4.1), with transition probability matrix

$$\boldsymbol{M}_{r,s}(\boldsymbol{A}) = \left[\begin{array}{c|c} \boldsymbol{N}_{r,s}(\boldsymbol{A}) & \boldsymbol{C}_{r,s}(\boldsymbol{A}) \\ \hline \boldsymbol{O} & \boldsymbol{I} \end{array}\right],$$

where $\boldsymbol{F}_{r,s}$ is a collection of simple patterns associated with the event $S_n(r) = s$, $S(\Lambda_i)$ is a collection of all sub-patterns in Λ_i and $\Lambda_i \in \boldsymbol{F}_{r,s}$. Also,

$$P(W(\Lambda_{r,s}) > n) = \xi_0 \boldsymbol{N}_{r,s}(\boldsymbol{A})^n \mathbf{1}', \tag{4.4}$$

where $\xi_0 = (1, 0, \ldots, 0)$ is one row vector and $\mathbf{1}'$ is the transpose of the row vector $\mathbf{1} = (1, 1, \ldots, 1)$.

Proof. As mentioned earlier, the probability of observing a success for a trial only affects the transition probability matrix and the theorem follows immediately. □

Let $\pi_{i|i-1}$ be the transition probability matrix of i^{th} trial. To test for

$$H_o : \pi_{i|i-1} \sim \boldsymbol{A}_0 \quad \text{v.s.} \quad H_a : \pi_{i|i-1} \sim \boldsymbol{A}_a \ \text{ for all } \ \boldsymbol{A}_0 \neq \boldsymbol{A}_a, \ \ i = 1, \ldots, n$$

we have the power function of

$$\begin{aligned} P(S_n(r) \geq s_\alpha | H_a) &= P(W(\Lambda_{r,s_\alpha}) \leq n | H_a) \\ &= 1 - \xi_0 \boldsymbol{N}_{r,s_\alpha}(\boldsymbol{A}_a)^n \mathbf{1}', \end{aligned}$$

where s_α is the critical region such that

$$\begin{aligned} P(S_n(r) \geq s_\alpha | H_o) &= P(W(\Lambda_{r,s_\alpha}) \leq n | H_o) \\ &= 1 - \xi_0 \boldsymbol{N}_{r,s_\alpha}(\boldsymbol{A}_0)^n \mathbf{1}' \quad \leq \alpha \end{aligned}$$

through Equation (4.4).

Theorem 14. *Let $\{X_i\}_{i=1}^n$ be a sequence of first-order homogeneous Markov dependent bistate trials taking values in $\boldsymbol{A}$, let $\Lambda_{r,s}$ be a compound pattern, associated with the scan statistic $S_n(r)$, with transition probability matrix $\boldsymbol{M}_{r,s}(\boldsymbol{A})$. If*

(i) λ_1 has algebraic multiplicity m and $\lambda_1 > |\lambda_j|$ for all $j > m$, and

(ii) there exist constants $a_1, a_2, \ldots, a_d$ such that $\mathbf{1}' = \sum_{i=1}^d a_i \eta_i'$ and $a_1(\xi_0 \eta_1') > 0$, then

$$P(W(\Lambda_{r,s}) > n) = \sum_{i=1}^{m} a_i \xi_1' \eta_i' \lambda_1^{n-1} \left[1 + o(e^{-n(\log \lambda_1 - \log \lambda_j)})\right]$$

where $\eta_i, i = 1, 2, \ldots, d$, is the eigenvector associated with λ_i for $\boldsymbol{N}_{r,s}^s(\boldsymbol{A})$, a submatrix of $\boldsymbol{N}_{r,s}(\boldsymbol{A})$.

Proof. It is not difficult to see that Equation (4.4) can be expressed as

$$P(W(\Lambda_{r,s}) > n) = \xi_1 \boldsymbol{N}_{r,s}(\boldsymbol{A})^{n-1} \mathbf{1}' \tag{4.5}$$

where $\xi_1 = (0, p, q, \ldots, 0)$ is a row vector, $\mathbf{1}'$ is the transpose of the row vector $\mathbf{1} = (1, 1, \ldots, 1)$, and the matrix $\boldsymbol{N}_{r,s}(\boldsymbol{A})$ can be shown to be of the form

$$\boldsymbol{N}_{r,s}(\boldsymbol{A}) = \begin{bmatrix} 0 & p & q & 0 & \cdots \\ 0 & & \boldsymbol{N}^s_{r,s}(\boldsymbol{A}) & & \end{bmatrix}.$$

As the first entry in ξ_1 and the components of the first column in $\boldsymbol{N}_{r,s}(\boldsymbol{A})$ are zero, Equation (4.5) can be easily rewritten as

$$P(W(\Lambda_{r,s}) > n) = \xi_1' \boldsymbol{N}^s_{r,s}(\boldsymbol{A})^{n-1} \mathbf{1}',$$

where $\xi_1' = (p, q, \ldots, 0)$ is one row vector and $\mathbf{1}'$ is the transpose of the row vector $\mathbf{1} = (1, 1, \ldots, 1)$.

Inspired by Fu and Jonson (2009), since $\lambda_1 = \cdots = \lambda_m$, $\eta_i \eta_i' = 1$, and $\eta_i \eta_j' = 0$, we have

$$\begin{aligned} P(W(\Lambda_{r,s}) > n) &= \xi_1' \boldsymbol{N}^s_{r,s}(\boldsymbol{A})^{n-1} \mathbf{1}' \\ &= \xi_1' \left[\sum_{i=1}^{d} \lambda_i^{n-1} \eta_i' \eta_i \right] \left[\sum_{i=1}^{d} a_i \eta_i' \right] \\ &= \sum_{i=1}^{m} a_i \xi_1' \eta_i' \lambda_1^{n-1} \left[1 + \sum_{i=m+1}^{d} e^{\log a_i + log \xi_1' + log \eta_i' - \log(\sum_{i=1}^{m} a_i \xi_1' \eta_i') + (n-1)(\log \lambda_i - \log \lambda_1)} \right]. \end{aligned}$$

As $n \to \infty$. This equation can be rewritten as

$$\sum_{i=1}^{m} a_i \xi_1' \eta_i' \lambda_1^{n-1} \Big[1 + o(e^{-n(\log \lambda_1 - \log \lambda_j)}) \Big].$$

□

4.2 Numerical Results

4.2.1 Powers for a discrete scan statistic in a bistate sequence

4.2.1.1 The *i.i.d.* case

We first consider a common case which has only two blocks in a sequence of n bistate trials with different chance of success within each block. We state the hypotheses as

$$H_o : \pi(1) = \pi(2) = \pi \text{ v.s. } H_a : \pi(1) = \phi, \pi(2) = \pi \ , \phi > \pi.$$

Given the value of the type I error α, Table 4.1 provides the exact power for the discrete scan statistic under various settings of the sample size n and the number of trials with a higher success chance m_1. Notice that we are testing against the entire sequence of binomial trials with a higher chance of a success ϕ when $m_1 = n$. We can see in Table 4.1 that the power increases from 6.44% to 12.54% when m_1 increases from 10 to 20, and from 6.44% to 24.50% when ϕ increases from 0.6 to 0.9. Therefore, we conclude that there might be a positive association between power and m_1, and between power and ϕ in this case. However, ϕ has a stronger effect on the power than m_1. Also the scan statistic test can easily detect two patterns in the sequence as long as there is a certain number of events with a higher chance of a success. For example, given that m_1 is 20, power would be 96.85%, 89.33%, and 72.20% when n is 50, 100, and 500, respectively.

In order to compare the performance between approximate power (see Equations (3) and (4) in Wallenstein et al., 1994) and exact power calculated by Equation

Table 4.1: Evaluations of power of the discrete scan statistic for $k = 2$

				$\pi = 0.5$	m_1				
n	r	s_α	α	ϕ	10	20	30	40	50
50	20	16	0.0487	0.6	0.0644	0.1254	0.1879	0.2427	0.2733
				0.7	0.0951	0.3423	0.5149	0.6419	0.7111
				0.8	0.1509	0.7060	0.8699	0.9474	0.9736
				0.9	0.2450	0.9685	0.9950	0.9996	0.9999
100	20	17	0.0317	0.6	0.0367	0.0605	0.0883	0.1149	0.1407
				0.7	0.0477	0.1792	0.2996	0.4021	0.4900
				0.8	0.0705	0.4915	0.6948	0.8245	0.8986
				0.9	0.1143	0.8933	0.9709	0.9949	0.9989
500	20	18	0.0345	0.6	0.0356	0.0420	0.0504	0.0587	0.0669
				0.7	0.0383	0.0872	0.1445	0.1976	0.2475
				0.8	0.0447	0.2762	0.4488	0.5857	0.6886
				0.9	0.0584	0.7220	0.8789	0.9582	0.9845

(4.2), our second example considers a segment of $Ir = 20$ consecutive trials with a higher chance of a success ϕ starting at 1^{st}, 11^{th}, and 21^{st} trial in the sequence. Apparently, the combinatorial method of Wallenstein et al. (1994) performs quite well as powers found through FMCI and combinatorial calculation are similar. In Table 4.2 the blank cells in the 6^{th} and 9^{th} columns are the results of the violation of the constraint: the length of the sequence must be a multiple of the window size. Moreover, the approximations can only be applied when the $\tau = 1, r + 1, \cdots$; therefore, the equation is not applicable when the 20 consecutive trials with a high chance of a success starting at the 11^{th} trial. Note that the number of blocks reduces to 2 when the starting point is at the beginning of the sequence. It is seen that power increases from 12.54% to 15.37% when we move the 20 consecutive trials with a higher chance of a success from starting at the 1^{st} trial to starting at the 11^{th} trial;

Table 4.2: Power comparisons of the discrete scan statistic given 20 contiguous events with a higher chance of a success that starts at the 1^{st}, 11^{th}, and 21^{st} trial

				$k=2, \pi(1)=\phi$ $m_1=r$ $\pi(2)=0.5$		$k=3, \pi(2)=\phi$ $m_1=10, m_2=r$ $\pi(1)=\pi(3)=0.5$	$k=3, \pi(2)=\phi$ $m_1=m_2=r$ $\pi(1)=\pi(3)=0.5$	
n	r	s_α	ϕ	(4.3)	W1	(4.3)	(4.3)	W2
50	20	16	0.6	0.1254		0.1537	0.1537	
			0.7	0.3423		0.4023	0.4023	
			0.8	0.7060		0.7553	0.7553	
			0.9	0.9685		0.9763	0.9763	
100	20	17	0.6	0.0605	0.0606	0.0729	0.0741	0.0742
			0.7	0.1792	0.1793	0.2178	0.2197	0.2198
			0.8	0.4915	0.4916	0.5445	0.5460	0.5461
			0.9	0.8933	0.8933	0.9120	0.9123	0.9123
500	20	18	0.6	0.0420	0.0420	0.0456	0.0460	0.0460
			0.7	0.0872	0.0873	0.1040	0.1047	0.1048
			0.8	0.2762	0.2763	0.3149	0.3158	0.3158
			0.9	0.7220	0.7220	0.7525	0.7528	0.7528

Note: W1 and W2 were determined by Equations (4) and (5a) in Wallenstein et al., 1994.

however, power remains similar to 15.37% when once again moving these trials with a higher chance of a success to starting at the 21^{st} trial. In general, it seems that a different pattern in the sequence is more likely to be detected if it does not occur early.

Next, we consider an example of a cyclic pattern of success probabilities in the Bernoulli sequence to show that the proposed method has no constraints to derive the power function for a discrete scan statistic. We consider three cyclic paths: (i)0.5-0.6-0.7-0.8-0.9, (ii)0.50-0.55-0.60-0.65-0.70, and (iii)0.50-0.51-0.52-0.53-0.54 in the sequence of n bistate trials, and the exact powers are provided in Table 4.3. This suggests that the longer the sequence is, the larger the power would be.

Table 4.3: Power comparisons of the discrete scan statistic given a cyclic pattern of success chances

				cyclic path		
n	r	s_α	α	(i)	(ii)	(iii)
50	20	16	0.0487	0.6839	0.2650	0.0732
100	20	17	0.0317	0.7129	0.2371	0.0512
500	20	18	0.0345	0.9248	0.3459	0.0601

4.2.1.2 Markov dependent case

As our next example, consider a sequence of n first-order Markov dependent trials with transition probability matrix

$$\pi_{t|t-1} \sim \boldsymbol{A}_a = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \text{ for } t = 2, 3, \ldots, n.$$

We test the null hypothesis of $p = 0.5$ against the alternative hypothesis of $p > 0.5$, which means that the chance of having the same outcome as in the previous trial is higher than 50%, given the window size of 20. It is no surprise to see that the larger the value of p is, the higher the power would be, as seen in Figure 4.1.

4.2.2 Powers for a continuous scan statistic in detecting change segments in a time interval

By adapting Equation(1.11), we have for sufficiently large n,

$$P\left(\sup_{0<t\leq 1-w} S(w,t) < s\right) \approx \xi \prod_{i=1}^{n} N_{r,s}(p_i)\mathbf{1}', \tag{4.6}$$

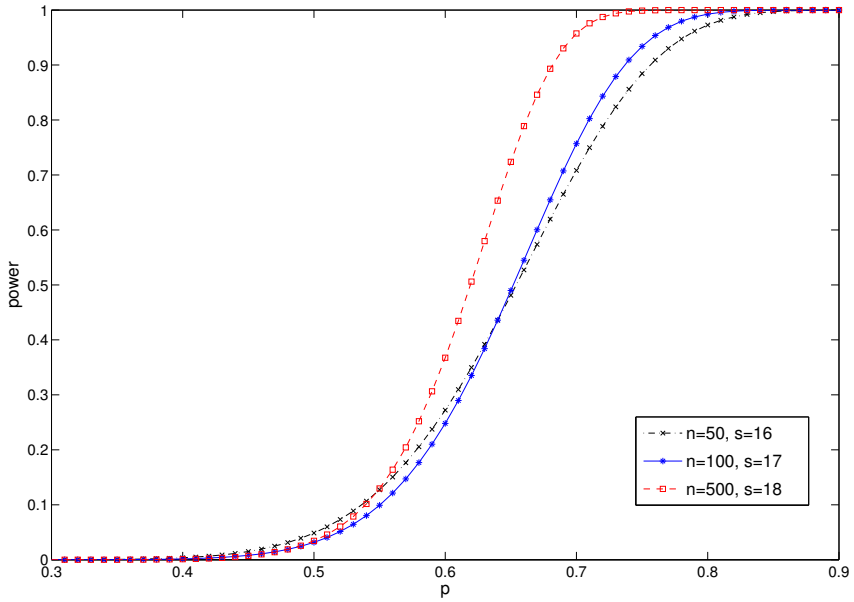

Figure 4.1: Probability curves of observing at least s successes in a window size 20 over a sequence of n Markov dependent trials

where $r = \lfloor nw \rfloor + k$ and

$$p_i = \begin{cases} \theta\lambda/n & \text{if} \quad i \in [\tau - w, \tau + w] \\ \lambda/n & \text{if} \quad i \in (0, \tau - w) \bigcup (\tau + w, 1 - w]. \end{cases}$$

Given $\lambda > 0$, we want to test the following hypotheses about the counting process $N(t)$

$$H_o : \theta = 1 \quad \text{v.s.} \quad H_a : \theta > 1.$$

Applying Equation (4.6), the critical value s_α of the test is determined from

$$\begin{aligned} P(S(w) \geq s_\alpha | H_o) \quad &\approx P(W(\Lambda_{r,s_\alpha}) \leq n | H_o) \\ &= 1 - \xi \prod_{i=1}^{n} N_{r,s}\left(\tfrac{\lambda}{n}\right) \mathbf{1}' \quad \leq \alpha, \end{aligned}$$

and the approximate power function can be derived as

$$\begin{aligned} P(S(w) \geq s_\alpha | H_a) \quad &\approx P(W(\Lambda_{r,s_\alpha}) \leq n | H_a) \\ &= 1 - \xi \prod_{i=1}^{n} N_{r,s}(p_i) \mathbf{1}'. \end{aligned}$$

There is no constraint when using FMCI to determine the power for a continuous scan statistic, such as the length of the subinterval time with a higher rate. In order to investigate the performance of approximate power equations, we choose $\tau = w$ and $\tau = 0.25$ to meet the conditions of approximate power equations and show the powers along with the simulation results of size 200,000 accordingly in Table 4.4. The numerical results suggest, by comparing columns 6, 7 and 8, that the value of

Table 4.4: Power comparison for the continuous scan statistic

					$\tau = w$			$\tau = 0.25$		
n	w	λ	s_α	θ	(4.6)	(1.9)	simulation	(4.6)	(1.10)	simulation
250	0.08	10	5	1	0.0579	0.0580	0.0581			
				1.4	0.0811	0.0672	0.0813	0.0845	0.0705	0.0816
				1.8	0.1257	0.0813	0.1245	0.1331	0.0888	0.1269
				2.2	0.1935	0.1032	0.1927	0.2049	0.1154	0.1930
				2.6	0.2814	0.1335	0.2789	0.2958	0.1504	0.2802
				3	0.3823	0.1719	0.3822	0.3986	0.1932	0.3816
500	0.04	24	6	1	0.0443	0.0498	0.0440			
				1.4	0.0565	0.0539	0.0564	0.0584	0.0559	0.0565
				1.8	0.0859	0.0628	0.0864	0.0906	0.0678	0.0863
				2.2	0.1390	0.0788	0.1396	0.1470	0.0875	0.1400
				2.6	0.2171	0.1033	0.2169	0.2282	0.1163	0.2167
				3	0.3161	0.1372	0.3147	0.3295	0.1544	0.3167

power through FMCI is closer to the simulation results than the ones through the combinatorial method.

Chapter 5

Conclusion and Future Work

Since Fu and Koutras (1994) proposed the FMCI approach to study the distribution theory for the runs in a sequence of Bernoulli trials, the method has been applied to various topics as we introduced in Chapter 1. In this thesis, we defined the random vector $\mathbf{U}_n$ and systematically imbed the random vector $\mathbf{U}_n$ into a Markov chain to induce the marginal and joint distributions of the rank statistics R_l and R_s for the location and scale parameters. Our approach provides not only a simple and clear matrix equation to derive the distribution function but also a solution for finding power for distributions equivalence tests considering the shift and scale parameters, separately and simultaneously. Secondly, numerical studies showed that a joint test should always be adopted for the test homogeneity of distributions as well as under Lehmann alternatives by comparing Figures 3.6 and 3.5 with Figures 3.12 and 3.13 and Table 3.10. Figures 3.8 to 3.11 also show that an elliptic critical region is always a better choice rather than a rectangular one for a joint test. Note that Figures 3.4 and 3.5 show how we decide a rectangular and elliptic critical region, respectively. In practice, it is reasonable to have neither the normality assumption nor equal mean/variance of the interested distributions when conducting a test for

distributions equivalence. The development of the joint distribution of the rank statistics R_l and R_s along with the discussion of its power is considered the main contribution of this dissertation.

Thirdly, we define blocks in a sequence of Bernoulli trials and apply the FMCI method to study the distribution of the scan statistic against (i) k-block independent trials with m_t events in t^{th} block and $\sum_{t=1}^{k} m_t = n$ in which the probability of success is $\pi(t)$ and (ii) a sequence of n Markov dependent trials. Our approach releases the conditions of τ and Ir in the alternative hypothesis (1.3). Numerical studies show that the power is dependent on τ and Ir and high. Fourthly, through modifying Equation (1.11), we are able to find the distribution of the continuous scan statistic considering a pulse alternative stated in (1.8) as well as the power of the scan statistic in detecting clustering. Wallenstein et al. (1993) used Poisson process to approximate the power under a condition which requires the length of the subinterval time τ with a higher success rate to be a multiple of the scan window size w. In our approach, the scan window size and the length of the subinterval time are independent. Table 4.4 showed that the power determined through our method is quite accurate even if τ is not a multiple of w.

The numerical results have been done using the software MATLAB®. Our program takes 1.6 hours to run a simulation of size one million for $m = 2$ and $n = 3$, 7 hours for $m = 3$ and $n = 4$, and 24 hours for $m = 5$ and $n = 7$ using a Macintosh desktop for finding the joint distribution of R_l and R_s. There is no problem for finding the joint distribution of R_l and R_s for up to sample sizes $m = 20$ and $n = 30$ using a computer with 8 GB of random-access memory (RAM); however, the process is time consuming. The main issue that affects the computer's performance is that

the size of the transition probability matrix Ω_n grows rapidly. For example, Ω_n is a 6-by-10 matrix when $m = 2$ and $n = 3$ and 462-by-792 when $m = 5$ and $n = 7$ which are two examples used in this thesis. Increasing the size of random-access memory can solve the issue of out of available memory but the speed of the calculations is a unsolvable issue in this thesis. One possible solutions could be to reduce the dimension of Ω_n, to upgrade the central processor unit (CPU) of the computer, or to use parallel computation. For example, to use a laptop with Intel Core i-7 2.90 GHz CPU can reduce the above times by roughly half. Regarding the numerical results for scan statistics, as long as the window size is smaller than or equal to 21, the program can determine the power in less than one minute using a Macintosh desktop. When the window size increases to 24, it takes around 20 hours to find the power. Even though the concept of FMCI is straightforward, the size of the transition probability matrix is always the main issue.

For nearly two decades the FMCI method has been applied to various topics, for example distribution theory of runs and patterns, waiting time distributions, boundary crossing, reliability, nonparametric tests, etc. However, most studies still focus on finding equations and simulating data to strengthen the theoretical results. Practically, this approach has not been recognized widely. One of the possible reasons is that this approach is not accessible for researchers from other fields due to the complexity in computations. Therefore, it is necessary to integrate programs in a suitable platform.

Given $\boldsymbol{x} = (x_1, \ldots, x_m)$, the conditional distribution of $\mathbf{U}_n = (u_1(n), \ldots, u_{m+1}(n))$ is multinomial with parameters $\boldsymbol{p} = (G(x_{[1]}) - G(x_{[0]}), \ldots, G(x_{[m+1]}) - G(x_{[m]}))$. Finding the limiting distribution of R_l or R_s in terms of the random vector $\mathbf{U}_n$ may be

another possible solution to deal with the problem of large sample sizes m and n. To extend this method to more than two samples might become possible.

Responses might be expensive or time consuming to collect in certain cases even though a larger sample can yield more accurate results. That makes the minimal sample size to achieve a desired power an important feature in inferential statistics. The relative efficiency of our proposed joint rank statistic to parametric tests has not been investigated yet.

Bibliography

Ansari, A.R. and Bradley, R.A. (1960), Rank-Sum Tests for Dispersions, *The Annals of Mathematical Statistics*, **31**, 1174-1189. (Cited on pages 7, 9 and 52.)

Antzoulakos, D.L., Koutras, M.V. and Rakitzis, A.C. (2009), Start-Up demonstration tests based on run and scan statistics, *Journal of Quality Technology*, **41**, 48-59. (Cited on page 3.)

Chang, Y.M. (2005), Distribution of waiting time until rth occurrence of a compound pattern, *Statistics and Probability Letters*, **75**, 29-38. (Cited on page 3.)

Collings, B.J. and Hamilton, M.A. (1988), Estimating the power of the two-sample Wilcoxon Test for location shift, *Biometrics*, **44**, 847-860. (Cited on page 7.)

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, 3rd ed., John Wiley & Sons, New York. (Cited on page 25.)

Fu, J.C. (1995) Exact and limiting distributions of the number of successions in a random permutation, *Annals of the Institute of Statistical Mathematics*, **47**, 435-446. (Cited on page 3.)

Fu, J.C. (1996) Distribution theory of runs and patterns associated with a sequence of multi-state trails, *Statistica Sinica*, **6**, 957-974. (Cited on page 3.)

Fu, J.C. (2001) Distribution of the scan statistic for a sequence of bistate trails, *Journal of Applied Probability*, **38**, 908-916. (Cited on pages 3, 13, 64 and 68.)

Fu, J.C. and Chang, Y.M. (2001) On probability generating functions for waiting time distributions of compound patterns in a sequence of multistate trials, *Journal of Applied Probability*, **39**, 70-80. (Cited on page 3.)

Fu, J.C. and Johnson, B.C. (2000), The distribution of increasing ℓ-sequences in random permutations: a Markov chain approach, *Statistics and Probability Letters*, **49**, 337-344. (Cited on page 3.)

Fu, J.C. and Johnson, B.C. (2009), Approximate probabilities for runs and patterns in *i.i.d.* and Markov-dependent multistage trials, *Advances in Applied Probability*, **41**, 292-308. (Cited on pages 4 and 71.)

Fu, J.C., Johnson, B.C. and Chang, Y.M. (2012), Approximating the extreme right-hand tail probability for the distribution of the number of patterns in a sequence of multi-state trials, *Journal of Statistical Planning and Inference*, **142**, 473-480. (Cited on page 4.)

Fu, J.C. and Koutras, M.V. (1994), Distribution theory of runs: a Markov chain approach, *Journal of the American Statistical Association*, **89**, 1050-1058. (Cited on pages 3 and 79.)

Fu, J.C. and Lou, W.Y.W. (2000), On the exact distribution of SECON and its application, *Statistica Sinica*, **10**, 999-1010. (Cited on page 4.)

Fu, J.C. and Lou, W.Y.W. (2000), Joint distribution of rises and falls, *Annals of the Institute of Statistical Mathematics*, **52**, 415-425. (Cited on page 3.)

Fu, J.C. and Lou, W.Y.W. (2003), *Distribution Theory of Runs and Patterns and Its Applications*, 1st edn. World Scientific, Singapore. (Cited on pages 4, 21 and 23.)

Fu, J.C. and Lou, W.Y.W. (2006), Waiting time distributions of simple and compound patterns in a sequence of $r-$th order Markov dependent multi-state trials, *Annals of the Institute of Statistical Mathematics*, **58**, 291-310. (Cited on page 3.)

Fu, J.C. and Lou, W.Y.W. (2007), On the Normal Approximation for the Distribution of the Number of Simple or Compound Patterns in a Random Sequence of Multi-state Trials, *Methodology and Computing in Applied Probability*, **9**, 195-205. (Cited on pages 4 and 25.)

Fu, J.C. and Lou, W.Y.W. (2008), Distribution of the length of the longest common subsequence of two multi-state biological sequences, *Journal of Statistical Planning and Inference*, **138**, 3605-3615. (Cited on page 3.)

Fu, J.C., Lou, W.Y.W. and Wang, Y.J.(1999), On the exact distributions of Eulerian and Simon Newcomb numbers associated with random permutations, *Statistics and Probability Letters*, **42**, 115-125. (Cited on page 4.)

Fu, J.C., Spring, F.A. and Xie, H. (2002), On the average run lengths of quality

control schemes using a Markov chain approach, *Statistics and Probability Letters*, **56**, 369-380. (Cited on page 3.)

Fu, J.C. and Wu, T.L. (2010), Linear and nonlinear boundary crossing probabilities for Brownian motion and related processes, *Journal of Applied Probability*, **47**, 1058-1071. (Cited on page 4.)

Fu, J.C., Wu, T.L. and Lou, W.Y.W. (2012), Continuous, discrete and conditional scan statistics, *Journal of Applied Probability*, **49**, 199-209. (Cited on page 16.)

Hoh, J. and Ott, J. (2000), Scan Statistics to Scan Markers for Susceptibility Genes, *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 9615-9617. (Cited on page 10.)

Johnson, B.C. (2001), The distribution of increasing ℓ-sequences in random permutations of arbitrary multi-state, *Methodology and Computing in Applied Probability*, **3**, 35-49. (Cited on page 3.)

Johnson, B.C. (2002), The distribution of increasing 2-sequences in random permutations of arbitrary multi-state, *Statistics and Probability Letters*, **59**, 67-74. (Cited on page 3.)

Klotz, J. (1962), Nonparametric test for scale, *The Annals of Mathematical Statistics*, **33**, 498-512. (Cited on pages 8 and 9.)

Koutras, M.V. (1996), On a waiting time distribution in a sequence of Bernoulli trials, *Annals of the Institute of Statistical Mathematics*, **48**, 789-806. (Cited on page 3.)

Koutras, M.V. (1996), On a Markov Chain Approach for the Study of Reliability Structures, *Journal of Applied Probability*, **33**, 357-367. (Cited on page 3.)

Koutras, M.V. and Alexandrou, V.A.(1995), Runs, scans and urn model distributions: A unified Markov chain approach, *Annals of the Institute of Statistical Mathematics*, **47**, 743-766. (Cited on page 3.)

Koutras, M.V. and Alexandrou, V.A.(1997), Sooner Waiting Time Problems in a Sequence of Trinary Trials, *Journal of Applied Probability*, **34**, 593-609. (Cited on page 3.)

Koutras, M.V. and Maravelakis, P.E. (2007), On the IFR property preservation for Markov chain imbeddable systems, *Probability in the Engineering and Informational Sciences*, **21**, 201-216. (Cited on page 3.)

Lehmann, E.L. (1953), The power for rank tests, *The Annals of Mathematical Statistics*, **24**, 23-43. (Cited on pages 7 and 59.)

Lehmann, E.L. (1998), *Nonprametrics: Statistical Methods Based on Ranks*, Revised ed., Prentice-Hall, New Jersey. (Cited on page 7.)

Lou, W.Y.W. (1996), On runs and longest run tests: a method of finite Markov chain imbedding, *Journal of the American Statistical Association*, **91**, 1595-1601. (Cited on page 3.)

Lou, W.Y.W. (1997), An application of the method of finite Markov chain imbedding to runs tests, *Statistics and Probability Letters*, **31**, 155-161. (Cited on page 3.)

Lou, W.Y.W. (1997), The exact distribution of overlapping counting patterns associated with a sequence of homogeneous Markov-dependent multi-state trials: an application to SECON, *Nonlinear Analysis, Theory, Methods & Applications* , **30**, 3349-3357. (Cited on page 4.)

Lou, W.Y.W. (1999), The exact distribution of the continuity of care measure NOP, *Statistics and Probability Letters*, **48**, 361-368. (Cited on page 4.)

Lou, W.Y.W. (2003), The exact distribution of the k-tuple statistic for sequence homology *Statistics and Probability Letters*, **61**, 51-59. (Cited on page 3.)

Mann, H.B. and Whitney, D.R.(1947), On a test of whether one of two random variables is stochastically larger than the other, *The Annals of Mathematical Statistics*, **18**, 50-60. (Cited on page 5.)

Mood, A.M.(1954), On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests, *The Annals of Mathematical Statistics*, **25**, 514-522. (Cited on page 8.)

Naus, J.I. (1974), Probabilities for a generalized birthday problem, *Journal of the American Statistical Association*, **69**, 810-815. (Cited on page 10.)

Nuel, G. (2006), Effective p-value computations using Finite Markov Chain Imbedding (FMCI): application to local score and to pattern statistics, *Algorithms for Molecular Biology*, **1**, :5. (Cited on page 3.)

Rosner, B. and Glynn, R.J.(2009), Power and sample size estimation for the Wilcoxon

rank sum test with application to comparisons of C statistics from alternative prediction models, *Biometrics*, **65**, 188-197. (Cited on page 6.)

Ross, M. S. (2000), *Introduction to Probability Models*, 7th edn. Academic Press, USA. (Cited on page 2.)

Rothman, E. (1967), Tests for clusters in a Poisson process, *The Annals of Mathematical Statistics*, **38**, 965. (Cited on page 16.)

Rothman, E. (1969), Tests for uniformity against regularly spaced alternatives, *Technical Report Number*, **119**, John Hopkins University. (Cited on page 17.)

Shieh, G., Jan, S.L. and Randles, R.H.(2006), On power and sample size determinations for the Wilcoxon-Mann-Whitney test, *Nonparametric Statistics*, **18**, 33-43. (Cited on page 6.)

Siegel, S. and Tukey, J.W.(1960), Distribution of runs tests under various models, *South African Statistical Journal* , **30**, 63-79. (Cited on pages 8 and 9.)

Smit, C.F.(1996), A nonparametric sum of ranks procedure for relative spread in unpaired samples, *Journal of the American Statistical Association*, **55**, 429-445. (Cited on page 3.)

Wallenstein, S., Naus, J. and Glaz, J. (1993), Power of the scan statistic for detecting of clustering, *Statistics in Medicine*, **12**, 1829-1843. (Cited on pages 1, 14, 18 and 80.)

Wallenstein, S., Naus, J. and Glaz, J. (1994), Power of the scan statistic in detecting

a changed segment in a Bernoulli sequence, *Biometrika*, **81**, 595-601. (Cited on pages 1, 11, 13, 18, 68, 72, 73 and 74.)

Wilcoxon, F.(1945). Individual comparisons by ranking methods, *Biometrics*, **1**, 80-83. (Cited on pages 5, 17 and 26.)

Zhao, X., Cui, L. and Kuo, W.K.(2007). Reliability for sparsely connected consecutive-k systems, *IEEE Transactions on Reliability*, **56**, 516-524. (Cited on page 3.)

Zhao, X. and Cui, L.(2009). On the accelerated scan finite Markov chain imbedding approach, *IEEE Transactions on Reliability*, **58**, 383-388. (Cited on page 3.)