Data driven smooth test for paired populations

B. Ghattas\textsuperscript{d,*}, D. Pommeret\textsuperscript{d}, L. Reboul\textsuperscript{e}, A.F. Yao\textsuperscript{f}

\textsuperscript{a}Institut de Mathématiques de Luminy, Campus de Luminy, 13288 Marseille Cedex 9, France
\textsuperscript{b}Université de Poitiers, IUT de Niort, Département STID, 79000 Niort Cedex, France
\textsuperscript{c}Centre d’Océanologie de Marseille, Campus de Luminy, 13288 Marseille Cedex 9, France

*Corresponding author

Email addresses: ghattas@univmed.fr (B. Ghattas), pommeret@iml.univ-mrs.fr (D. Pommeret), reboul@iml.univ-mrs.fr (L. Reboul), anne-francoise.yao@univmed.fr (A.F. Yao)

Preprint submitted to Elsevier March 14, 2010
Data driven smooth test for paired populations

B. Ghattas\textsuperscript{d,}\textsuperscript{*}, D. Pommeret\textsuperscript{d}, L. Reboul\textsuperscript{e}, A.F. Yao\textsuperscript{f}

\textsuperscript{d}Institut de Mathématiques de Luminy, Campus de Luminy, 13288 Marseille Cedex 9, France
\textsuperscript{e}Université de Poitiers, IUT de Niort, Département STID, 79000 Niort Cedex, France
\textsuperscript{f}Centre d'Océanologie de Marseille, Campus de Luminy, 13288 Marseille Cedex 9, France

Abstract
In this paper we propose a smooth test of comparison of two distribution functions. This test adapts to the classical two-sample problem as well as that of paired populations, including discrete distributions. A simulation study and an application to real data show its good performances.

Keywords: Smooth test, Paired sample, Copulas, Score statistic, Schwarz’s rule

1. Introduction
In various contexts, it is often of interest to compare the distributions of two populations. More specifically, let $f_X$ and $f_Y$ be the unknown densities of two populations $X$ and $Y$. A popular problem is to test the nonparametric hypothesis

$$H_0 : f_X = f_Y$$

versus the omnibus alternative that the two distributions are different, on the basis of an i.i.d. sample of each population. When the two populations are unrelated, testing (1) is generally referred to as the two-sample testing problem. In many situations however, independence of the two populations under study cannot be assessed. This is the case whenever there exists a natural coupling between an observation in one sample and a unique observation in the other sample, irrespective of their actual values (repeated measures on the same unit, observation of two different personal characteristics on individuals, etc...). To give a practical example, it is a common practice in ChIP-On-chip\textsuperscript{1} experiments to compare densities coming from two channels corresponding to paired measurement of genes expressions. In such cases, we expect to achieve more precise results when testing (1) with methods that are specifically designed to reflect this pairing.

\textsuperscript{*}Corresponding author

Email addresses: ghattas@univmed.fr (B. Ghattas), pommeret@iml.univ-mrs.fr (D. Pommeret), reboul@iml.univ-mrs.fr (L. Reboul), anne-francoise.yao@univmed.fr (A.F. Yao)

\textsuperscript{1}Chromatin ImmunoPrecipitation on chip
The two-sample nonparametric testing problem (1) has been widely studied. Classical methods are distribution-based omnibus tests (see e.g. Smirnov [27], Rosenblatt [24]) and linear rank tests (see e.g. Wilcoxon [28], Mood [21], Lepage [19]). Insightful comments on the weakness of such tests are given in Eubank and LaRiccia [6], Fan [7], Janic-Wróblewska and Ledwina [15]. Expanding the unknown distributions under study in a basis of functions, the authors show that the classical procedures only use information contained in the first few dimensions of the expansion, corresponding essentially to location-scale alterations of the null hypothesis. This explains the low power of these tests in detecting more subtle differences, corresponding to higher frequency alterations such as local bumps, changes in skewness or kurtosis. In contrast, the last two authors propose testing procedures that give equal weights to all components in the expansion so that they should be able to properly detect other directions of departure than constant location/scale shifts. Such tests are particular cases of smooth tests. The idea underlying smooth tests is to model the alternative distribution as a series expansion along a \( k \)-dimensional family of functions. Testing \( H_0 \) thus reduces to test the nullity of the coefficients in the expansion. For that task, one generally uses a score statistic. The number \( k \) of selected functions in the expansion turns out to be a nuisance parameter to be optimized. A pioneer paper about smooth tests is Neyman [22]. Many references to this work can be found in Hart [11]. The performances of such tests partially rely on the right choice of \( k \) (for discussions, see D’Agostino and Stephens [4] and Rayner and Best [23]). For a long time, many authors have restricted attention to study \( k \in \{1, 2, 3, 4\} \). However, such choice can be misleading in some situations (see Inglot et al. [13]). More generally, since the right choice depends on the type of alternative, which is unknown, a deterministic choice of \( k \) is not satisfactory. A solution to this problem has been first proposed by Ledwina [18]. The author studies a data-based method for choosing \( k \) between 1 and some arbitrary fixed integer \( K \), inspired from Schwarz [26]’s criterion selection rule. The method adapts to the data without prior knowledge about the considered alternative. Afterward, some refinements have been proposed by several authors, among whose Kallenberg and Ledwina [16], Inglot et al. [14]. In these papers, \( K \) is allowed to tend to infinity at a given rate as the sample size tends to infinity. Once the optimal \( k \) is chosen, the corresponding score test statistic is performed and used for testing \( H_0 \).

Smooth tests theory as well as data-driven methods for the selection of \( k \) have been adapted to the two-sample testing problem in different ways. Fan [7] uses a wavelet decomposition to build a test statistic. In Janic-Wróblewska and Ledwina [15], both samples are transformed via their empirical distribution functions and the test statistic is based on the ranks of the modified samples. This standard approach is followed in Albers et al. [1] and Ghosh [8]. However, in the paired situation, the dependence of the ranks and empirical distribution functions turns out to be an important difficulty. To get around this problem, Chervoneva and Iglewicz [3] use another approach based on orthogonal bases for detecting differences between continuous distributions. But unlike in the previous studies, the number \( k \) of coefficients in the expansion is chosen in a somewhat arbitrary way.

In this paper, we propose a data driven smooth test based on orthogonal series, inspired from the above-mentioned references. This testing method generalizes the results of
Chervoneva and Iglewicz [3] from two points of view: we propose an automatic selection of $k$; our test can be used to compare continuous as well as discrete distributions. Moreover, the method adapts to the paired-sample case.

Assume one observes two possibly paired i.i.d. samples with unknown distributions $f_X$ and $f_Y$ with respect to a given measure $\nu$ ( $\nu$ can be chosen such that it is not necessarily dominated by Lebesgue’s measure, which allows to deal, for instance with discrete distributions). In order to test (1), we consider the series expansions of $f_X$ and $f_Y$ along a $k$-dimensional family of orthogonal functions in $L^2(\nu)$, for a given $k > 0$.

Testing (1) thus reduces to the parametric testing problem that the coefficients of the same order in both expansions are equal. We then derive a test statistic, based on the random vector $U_n(k)$ of the differences between the estimated coefficients. This statistic is the score one under some assumptions. Next, we use Kallenberg and Ledwina [16]’s data-driven method to choose the optimal dimension of the orthogonal family, say $S_n$. Replacing $k$ by $S_n$ in the formula of the test statistic, we obtain our main test statistic $\tilde{T}_n$. Our test strategy relies on the critical values of the asymptotic distribution of $\tilde{T}_n$ under the null. This distribution is obtained under some conditions on the orthogonal family and on the eigenvalues of the variance-covariance matrix of $U_n(k)$, as $k$ gets large. We also prove the consistency of our test.

The rest of the paper is organized as follows. In Section 2 we introduce the test statistic and study its asymptotic distribution for a given $k > 0$. In Section 3 we use the data-driven method to optimize $k$ and derive our main test strategy. Sections 4 and 5 contain simulation results and an application to a real dataset from ChIP-on-chip microarrays. The proofs are devoted to Section 6.

2. A sequence of test statistics

Let $(X_s)_{1 \leq s \leq n}$ and $(Y_s)_{1 \leq s \leq n}$ be two possibly paired i.i.d. samples (see Remark 1 for the classical two-sample case). Let $\nu$ be a given probability measure with density $h$ with respect to some reference measure $\lambda$ (Lebesgue’s or counting measure for instance). We denote by $f_X$ and $f_Y$ the respective unknown marginal densities of the $X_s$’s and the $Y_s$’s with respect to $\lambda$. We assume that $f_X$ and $f_Y$ belong to $L^2(\nu)$. We wish to test:

$$H_0 : f_X = f_Y \quad \text{against} \quad H_1 : f_X \neq f_Y.$$  

(2)

For that task, we consider the expansions of $f_X$ and $f_Y$ along a dense family of orthogonal functions in $L^2(\nu)$. Denoting by $(Q_j)_{j \in \mathbb{N}}$ this family, we have

$$f_X = \sum_{j \geq 0} a_j Q_j \quad \text{and} \quad f_Y = \sum_{j \geq 0} b_j Q_j,$$

with

$$a_j = \mathbb{E} (\tilde{Q}_j(X_1)) = \int_{\mathbb{R}} Q_j(t)f_X(t)d\nu(t) \quad \text{and} \quad b_j = \mathbb{E} (\tilde{Q}_j(Y_1)) = \int_{\mathbb{R}} Q_j(t)f_Y(t)d\nu(t).$$

4
and $\tilde{Q}_j = bQ_j$ for all $j \in \mathbb{N}$. Since in general the sequences $(a_j)_{j \in \mathbb{N}}$ and $(b_j)_{j \in \mathbb{N}}$ rapidly tend towards zero as $j$ gets large, it is reasonable to assume that there exists some $k > 0$ such that the approximations

$$f^{[k]}_X = \sum_{j=0}^{k} a_j Q_j \quad \text{and} \quad f^{[k]}_Y = \sum_{j=0}^{k} b_j Q_j$$

are tight enough for a sufficiently wide range of density functions. Therefore, our testing problem reduces to the parametric testing problem

$$H_0 : \forall j = 1 \ldots k, a_j = b_j \quad \text{against} \quad H_1 : \exists j = 1 \ldots k, a_j \neq b_j. \quad (4)$$

A natural test strategy for (4) can be defined as follows: for every $s = 1, \ldots, n$, let us set $V_s(k) = (\tilde{Q}_j(X_s) - \tilde{Q}_j(Y_s))_{1 \leq j \leq k}$ and

$$U_n(k) = \frac{1}{\sqrt{n}} \sum_{s=1}^{n} V_s(k).$$

Moreover, assume

(A_k): There exists some positive constant $M_k$ such that

$$\mathbb{E}_0 \left( \|V_1(k)\|^4 \right) < M_k.$$ 

It is clear that under $H_0$, $U_n(k)$ has mean zero and finite $k \times k$ variance-covariance matrix

$$W(k) = \mathbb{E}_0 \left( V_1(k) V_1(k)' \right),$$

where $\mathbb{E}_0$ denotes the expectation under $H_0$ and $V_1(k)'$ is the transposition of $V_1(k)$. Next, let us define the empirical version of $W(k)$ under $H_0$, that is the $k \times k$ matrix

$$\hat{W}_n(k) = \frac{1}{n} \sum_{s=1}^{n} V_s(k) V_s(k)'.$$

In the following, we assume that the inverse of $\hat{W}_n(k)$ exists (see Remark 2 for a generalization). Denoting by $\|\cdot\|$ the euclidean norm on $\mathbb{R}^k$, we shall compute the test statistic

$$T_n(k) = U_n(k)' \hat{W}_n(k)^{-1} U_n(k) = \|\hat{W}_n(k)^{-1/2} U_n(k)\|^2,$$

then reject the null hypothesis in (4) whenever the value of $T_n(k)$ is too large. Hereafter, we derive the asymptotic distribution of $T_n(k)$ under $H_0$:

**Proposition 1.** Assume that (A_k) holds. Then, under $H_0$, $T_n(k)$ converges in distribution to a chi-squared random variable with $k$ degrees of freedom as $n$ tends toward infinity.
So, a test strategy for (2) should be to reject $H_0$ as soon as $T_n(k)$ exceeds the $\alpha$-upper percentile of its asymptotic distribution. It is worth noticing that such a test is expected to be powerful if (3) holds. Therefore, a key concern is to choose a convenient model (3) among those corresponding to $k > 0$. If $k$ is too small, the approximations may be crude and the test loses power. If $k$ is too large, (3) encompasses unnecessary components so that power dilution may occur. In the following section, we propose a data-driven method for choosing $k$.

**Remark 1.** In the case of two independent samples with possibly different sample sizes $n_X$ and $n_Y$, Proposition 1 still holds using in place of $U_n(k)$ the random vector with elements

$$\sqrt{\frac{2n_Xn_Y}{n_X+n_Y}} \left( \frac{1}{n_X} \sum_{s=1}^{n_X} \tilde{Q}_j(X_s) - \frac{1}{n_Y} \sum_{s=1}^{n_Y} \tilde{Q}_j(Y_s) \right), j = 1 \ldots k.$$

**Remark 2.** There may exist two different reasons leading to a problem of specification of $\hat{W}_n(k)^{-1}$:

- $\hat{W}_n(k)$ is not of full rank. In this case, let $K$ be the greatest rank such that $\hat{W}_n(K + 1)$ exists. Since $\hat{W}_n(K + 1)$ is not of full rank, we remove polynomials of degree $K + 1$ from the construction of the test statistic and directly add polynomials of degree $K + 2$. We then get a new test statistic based on the $(K + 1) \times (K + 1)$ matrix obtained from the family $\{P_1, P_2, \ldots, P_K, P_{K+2}\}$. The degree of freedom of the test statistic is reduced accordingly (see Theorems 6.1.1. and 6.1.2 of Rayner and Best [23]).

- $\hat{W}_n(k)$ is of full rank but $k$ is large (as it has been formerly mentioned, this case seldom happens). Thus, denoting by $\lambda_{\min}(k)$ the smallest eigenvalue of $W(k)$, we have that $\lambda_{\min}(k)$ tends to zero when $k$ goes to infinity, so that $\hat{W}_n(k)$ is ill-conditioned. This case can be dealt using similar arguments as in Section 3. Namely, Proposition 1 continues to hold if Condition (B) of Section 3 holds, replacing $d(n)$ with $k$.

3. Calibrating the test statistic

In this section, we adapt the data-driven method introduced by Kallenberg and Ledwina [16] to optimize the nuisance parameter $k$. It is based on a modified version of Schwarz [26]'s Bayesian information rule. For classical parametric testing problems, this rule consists in maximizing a penalized version of the first order approximation of the likelihood ratio test statistic. A partial interpretation of this criterion in our particular setting is given in Remark 3. Heuristically, let $d(n)$ be an increasing sequence such that $\lim_{n \to \infty} d(n) = \infty$. We first select a likely model among the $d(n)$ models given by (3), fitting the data at hand. For that task, we choose the optimal value of $k$, denoted by $S_n$, such that

$$S_n = \min \left\{ \arg \max_{1 \leq k \leq d(n)} \{ T_n(k) - k \log(n) \} \right\}. \quad (6)$$
Once $S_n$ is determined, we apply the test (4) to the fitted model, leading to use (5) as the test statistic, with $k = S_n$. More precisely, we use for our testing problem the statistic

$$\tilde{T}_n = T_n(S_n).$$

Hereafter, we derive the asymptotic distribution of $\tilde{T}_n$ under the null and prove that the test based on the limiting quantile is consistent. We first need to make further assumptions that ensure the existence and the consistency of $\hat{W}_n(d(n))^{-1}$. Namely,

(A): There exists some positive constant $M$ such that

$$\sup_{k>0} \mathbb{E}_0 \bigg( \|V_1(k)\|^4 \bigg) < M.$$

(B): Setting $\lambda_n^*$ the smallest eigenvalue of $W(d(n))$,

$$d(n) \lambda_n^* = O_P(\log n).$$

Thus

**Theorem 1.** Let $d(n)$ be an increasing sequence such that $\lim_{n \to \infty} d(n) = \infty$. Assume that (A) and (B) hold. Thus, under $H_0$, $\tilde{T}_n$ converges in distribution to a chi-squared random variable with 1 degree of freedom.

Moreover, let us consider the alternatives having the form:

$$H^*_1 : \exists q \in \mathbb{N}, \forall i = 1, \ldots, q - 1, a_i = b_i \text{ and } a_q \neq b_q.$$

**Proposition 2.** Assume that (A) holds. Then, under $H^*_1$, $\tilde{T}_n$ converges in probability to infinity. Moreover, if (B) holds, the test based on the limiting distribution of $\tilde{T}_n$ is consistent against every alternative of the form $H^*_1$.

So, an obvious test strategy for (2) should be to reject $H_0$ whenever the value of $\tilde{T}_n$ exceeds the $\alpha$-upper percentile of its asymptotic distribution. However, several studies dealing with smooth tests (see for instance Kallenberg and Ledwina [16], Janic-Wróblewska and Ledwina [15]) emphasize that the asymptotic approximation is not very accurate for small and moderate sample sizes so that we shall propose a sharper one. By (6), it is easily seen that $T_n \geq T_n(1)$, which means that $\mathbb{P}_0(T_n \leq x)$ is overestimated by its asymptotic approximation. Therefore, using the $\alpha$-upper percentile of the limiting distribution leads to underestimate the actual level and power of the test. A sharper approximation for $\mathbb{P}_0(T_n \leq x)$ can be obtained following Kallenberg and Ledwina [17]:

$$P_0(\tilde{T}_n \leq x) \simeq \mathbb{P}_0(\tilde{T}_n \leq x, S_n = 1) + \mathbb{P}_0(\tilde{T}_n \leq x, S_n = 2)$$

$$= \mathbb{P}_0(U_1^2 \leq x) \mathbb{P}_0(U_2^2 \leq \log n) + \mathbb{P}_0(U_1^2 + U_2^2 \leq x, U_2^2 \geq \log n)$$

$$= \mathbb{P}_0(U_1^2 \leq x) - \mathbb{P}_0(U_1^2 \leq x, U_1^2 + U_2^2 \geq x, U_2^2 \geq \log n),$$
where $U_1$ and $U_2$ denote two standard independent Gaussian variables. This can be straightforwardly reformulated as

$$\mathbb{P}(\tilde{T}_n \leq x) \simeq \begin{cases} (2\Phi(\sqrt{x}) - 1)(2\Phi(\log n) - 1) & \text{if } x \leq \log n \\ (2\Phi(\sqrt{x}) - 1)(2\Phi(\log n) - 1) + 2\Phi(-\sqrt{\log n}) & \text{if } x \geq 2 \log n \\ \text{linearize otherwise,} & \end{cases}$$

(7)

where $\Phi$ is the cumulative distribution function of a standard Gaussian variable.

**Remark 3.** In the paper, we do not give any formal interpretation of (6) in our general setting. However, such an interpretation is easy in the particular case where the maximum likelihood estimator $\hat{\theta}_n$ of $\theta = \mathbb{E}(V_1(k))$ is the empirical mean of the $V_i(k)$’s, that is $\hat{\theta}_n = \frac{U_n(k)}{\sqrt{n}}$. It is true for instance when the distribution of $V_1(k)$ belongs to an exponential family. So, rewriting (4) as $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$, $T_n(k)$ is the score statistic and (6) is the simplified Schwarz’s rule exposed in Kallenberg and Ledwina [16].

**Remark 4.**

- The matrix $W(k)$ is a nuclear operator when $k$ tends to infinity (see Dauxois et al. [5]) so that we have $\lim \lambda_{\min}(k)_{k \to \infty} = 0$. Therefore, $\lambda_{n}^{k}$ tends to zero as $n$ tends to infinity.

- Condition (B) is a sufficient condition under which Theorem 1 obtains, but it may happens to be too restrictive in some cases. A less demanding condition is

$$\max\left(\log d(n), \frac{d(n)}{\lambda_{n}^{k}}, \frac{d(n)}{\sqrt{n}}\right) \to 0 \text{ in probability.}$$

- It is worth noticing that Condition (B) is satisfied in some well studied situations (see Ash and Gardner [2]). For instance, we shall consider the case where $V_1(k)$ coincides with the $k$ first components of the Karhunen-Loève decomposition of a Wiener or Ornstein-Uhlenbeck process. So, we have $\lambda_{j} \sim a r^{-\gamma}, a > 0, \gamma > 1$ and Condition (B) holds with $d(n) = O\left(|\log n|^{1/(1+\gamma)}\right)$. Another situation is the case where the eigenvalues are given by $\lambda_{j} \sim a r^{j}, a > 0$ and $0 < r < 1$. Condition (B) then holds with $d(n) = O\left(|\log(\log n)|^{1/(1-\log r)}\right)$.

4. Simulation study

In order to evaluate the finite sample performances of our test, Monte Carlo simulations were performed on several sample sizes and models. The models consist of different marginal distributions, within-pair dependence structures and association degrees. For each model and sample size, we compute the empirical level and power of our test as well as that of several competitors and compare them. The nominal level is fixed at $\alpha = 5\%$. We present below the chosen models and tests and the results obtained.
4.1. Models

The simulated examples consist of testing (2) based on samples of \( n \) pairs \((X_s, Y_s)\) with marginal densities \( f_X \) and \( f_Y \) \( (f_X = f_Y \text{ under the null}) \) and a within-pair dependence structure introduced via a copula \( C_\theta \). More precisely, the samples are drawn from the joint cumulative distribution function \( F(x, y) = C_\theta(F_X(x), F_Y(y)) \) \( (F_X = F_Y \text{ under the null}) \), where \( F_X \) and \( F_Y \) are the marginal cumulative distribution functions of the \( X_s \)'s and the \( Y_s \)'s respectively. Three Archimedian copulas \( C_\theta \) (Gumbel, Frank and Clayton) were initially considered, each of them referring to a special kind of dependence structure (see [20] for more details). For given sample size, test procedure and pair of marginal distributions, these different copulas did not yield substantial differences of performances. Thus, we only display the results obtained for Clayton’s copula

\[
C_\theta(u, v) = \left( u^{-\theta} + v^{-\theta} - 1 \right)^{-1/\theta}, \quad (u, v) \in [0, 1]^2, \quad \theta \in \mathbb{R}^+. \tag{8}
\]

The association degree of the pairs is controlled by the value of \( \theta \) and increases with \( \theta \). The limiting case \( \theta = 0 \) corresponds to independence of the pairs. In the following, we use in place of \( \theta \) the Kendall’s tau as a measure of the association degree. It is given by \( \tau = \frac{\theta}{\theta^2} \). The studied pairs of marginal distributions are reported in Table 1. Our choice of the null hypothesis includes light and heavy-tailed symmetric distributions \((A1, A2, A5)\) as well as light and heavy-tailed skewed distributions \((A6, A7)\). The alternative distributions refer to different kind of alterations of the null. The alternative of model \( A1 \) is studied by Fan [7] in the two-sample problem and corresponds to a local alteration of the null. Alternatives in models \( A2, A5, A6, A7 \) correspond to classical location and scale alterations. The Gaussian mean mixture model \( A3 \) is studied in Janic-Wróblewska and Ledwina [15]. The alternative in model \( A4 \) corresponds to a particular case of non constant location shift investigated by Albers et al. [1].

Table 1: Marginal distributions with respect to Lebesgue’s measure used in the simulation. Under \( H_0 \), \( f_X \) is the marginal distribution of both samples. Under \( H_1 \), \( X \) (resp. \( Y \)) is drawn from \( f_X \) (resp. \( f_Y \)). The standard normal density is denoted \( \varphi \).

<table>
<thead>
<tr>
<th>Model</th>
<th>( f_X(t) )</th>
<th>( f_Y(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A1 )</td>
<td>( \frac{t}{2} ), ( t \in [-1, 1] )</td>
<td>( \frac{1}{2} + \frac{</td>
</tr>
<tr>
<td>( A2 )</td>
<td>( \varphi(t), t \in \mathbb{R} )</td>
<td>( \frac{1}{2} f_X\left(\frac{</td>
</tr>
<tr>
<td>( A3 )</td>
<td>( 0.7\varphi(t - \frac{</td>
<td>t</td>
</tr>
<tr>
<td>( A4 )</td>
<td>( \exp(-</td>
<td>t</td>
</tr>
<tr>
<td>( A5 )</td>
<td>( \exp(-</td>
<td>t</td>
</tr>
<tr>
<td>( A6 )</td>
<td>( \exp(-</td>
<td>t</td>
</tr>
<tr>
<td>( A7 )</td>
<td>( \varphi(\log(t)), t &gt; 0 )</td>
<td>( \frac{1}{\sigma} f_X\left(\frac{\log(t)}{\sigma}\right), \sigma \in {1.2, 1.6} )</td>
</tr>
</tbody>
</table>

From a practical point of view, to produce a realization \((x_s, y_s)_{1 \leq s \leq n}\) of the bivariate sample \((X_s, Y_s)_{1 \leq s \leq n}\), we generate bivariate observations \((u_s, v_s)_{1 \leq s \leq n}\) drawn from the joint distribution function \( C_\theta \). For that task, we use Marshall and Olkin [20]'s method of simulation of Archimedian copulas. We then apply the inverse probability transform so that for all \( s = 1, \ldots, n \), \( x_s = F_X^{-1}(u_s) \) and \( y_s = F_Y^{-1}(v_s) \), where \( F_X^{-1} \) and \( F_Y^{-1} \) are the respective inverse (or generalized inverse) cumulative distribution functions of \( X \) and \( Y \).
4.2. Tests

We first consider the test procedure described in Section 3. For ease of reference, we denote it by TS. Computing the test statistic $\tilde{T}_n$ first requires the choice of the sequence $d(n)$. A previous study has shown that the empirical levels and powers obtained do not depend on $d(n)$ for sufficiently large values of this parameter. In practice $d(n)$ can be set at 10. Secondly, we need to choose an obvious family $(Q_j)_{j\in\mathbb{N}}$ and a reference measure $\nu$. This choice depends on the support of the distribution under the null of the studied model and we refer to Table 2 for details. Finally, for $k = 1, \ldots, d(n)$, the test statistics (5) are computed and $S_n$ is determined by (6). The critical value of the test is the 5%-upper percentile of the approximated distribution (7). It only depends on the sample size.

Table 2: Orthogonal families and reference measures involved in the construction of TS

<table>
<thead>
<tr>
<th>Models</th>
<th>Measure $\nu$</th>
<th>Family</th>
<th>Recurrence formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>$U[-1,1]$</td>
<td>Legendre</td>
<td>$Q_0 = 1$, $Q_1 = x$, $(j+1)Q_{j+1}(x) = (2j+1)xQ_j - jQ_{j-1}(x)$.</td>
</tr>
<tr>
<td>A2-A6</td>
<td>$N(0,1)$</td>
<td>Hermite</td>
<td>$Q_0 = 1$, $Q_1(x) = 2x$, $2xQ_j(x) = Q_{j+1}(x) + 2jQ_{j-1}(x)$</td>
</tr>
<tr>
<td>A7</td>
<td>Exp(1)</td>
<td>Laguerre</td>
<td>$Q_0 = 1$, $Q_1(x) = 1 - x$, $(j+1)Q_{j+1}(x) = (2j+1 - x)Q_j(x) - jQ_{j-1}(x)$</td>
</tr>
</tbody>
</table>

Next, we consider several competitors. The two main ones are Wilcoxon signed-rank test (WX) and Wilcoxon rank-sum test (WX2), respectively designed for paired (see Hájek and Šidák [10] for details) and unpaired samples. We hope to demonstrate the better ability of our test method to detect alterations that do not correspond to pure location-shift problems. Moreover, the comparison with WX2 should allow us to see what is going on if we omit to take into account the within-pair dependence. In the cases of unpaired samples, we make further power comparisons with three omnibus tests studied in the literature, namely the modified Crâmer-Von-Mises test due to Schmid and Trede [25] (CVM), the data-driven smooth test of Janic-Wróblewska and Ledwina [15] (WL) and the wavelet thresholding test with hard threshold parameter of Fan [7] (F). Crâmer-Von-Mises test is known to have low power for other than location/scale models, WL has good power for smooth alternatives such as A3 and F is specifically adapted to local or high frequency alterations of the null such as A1.

4.3. Empirical levels and powers

For each $n \in \{50, 100, 200\}$, $\tau \in \{0, 0.25, 0.5, 0.75\}$ and marginal distribution $f_X$ (resp. pair of marginal distributions $(f_X, f_Y)$ of Table 1), we investigate the empirical levels (resp. powers) of the tests previously described. They are obtained as follows: using the method described in Subsection 4.1, a sample of $n$ pairs is drawn from the joint distribution $F(x,y) = C_\theta(F_X(x), F_X(y))$ (resp. $F(x,y) = C_\theta(F_X(x), F_Y(y))$), with $C_\theta$ given by (8). Then, for each test procedure, TS, WX
and WX2, we compute the test statistic based on the sample and compare it to the 5%-critical value of its approximated distribution under $H_0$ (formula (7) for TS and gaussian approximation for the others). The empirical level (resp. power) of the test is defined as the percentage of rejection of the null hypothesis over 50000 (resp. 10000) replications of the test statistic. The empirical levels are displayed in Table 3 while the comparative powers of TS, WX and WX2 are described by Figures 1 to 5 (in order to save space, some alternatives are not displayed here, but their study have been taken into account in our conclusions).

Table 3: Empirical levels (in %) for TS, WX and WX2. The within-pair dependency structure is described by Clayton's copula with association degree measured by $\tau$.

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>n</th>
<th>$\tau$</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>TS</td>
<td>4.8</td>
<td>4.7</td>
<td>4.6</td>
<td>4.7</td>
<td>4.5</td>
<td>4.5</td>
<td>4.4</td>
<td>5.1</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>WX</td>
<td>4.8</td>
<td>4.9</td>
<td>5.0</td>
<td>4.9</td>
<td>4.8</td>
<td>4.9</td>
<td>4.8</td>
<td>4.9</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>WX2</td>
<td>5.0</td>
<td>1.4</td>
<td>0.1</td>
<td>0.0</td>
<td>4.9</td>
<td>1.4</td>
<td>0.1</td>
<td>0.0</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>WX</td>
<td>4.8</td>
<td>4.7</td>
<td>4.6</td>
<td>4.3</td>
<td>4.5</td>
<td>4.6</td>
<td>4.8</td>
<td>4.5</td>
<td>5.3</td>
</tr>
<tr>
<td>A2-A4</td>
<td>TS</td>
<td>4.8</td>
<td>4.7</td>
<td>4.6</td>
<td>4.3</td>
<td>4.5</td>
<td>4.6</td>
<td>4.8</td>
<td>4.5</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>WX</td>
<td>4.8</td>
<td>5.0</td>
<td>4.8</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>WX2</td>
<td>4.7</td>
<td>1.5</td>
<td>0.1</td>
<td>0.0</td>
<td>5.0</td>
<td>1.4</td>
<td>0.1</td>
<td>0.0</td>
<td>5.0</td>
</tr>
<tr>
<td>A5</td>
<td>TS</td>
<td>4.9</td>
<td>4.9</td>
<td>4.6</td>
<td>4.6</td>
<td>4.6</td>
<td>4.6</td>
<td>4.6</td>
<td>4.5</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>WX</td>
<td>4.9</td>
<td>5.0</td>
<td>4.8</td>
<td>4.9</td>
<td>5.1</td>
<td>4.8</td>
<td>4.8</td>
<td>5.0</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>WX2</td>
<td>5.0</td>
<td>1.4</td>
<td>0.1</td>
<td>0.0</td>
<td>5.0</td>
<td>1.4</td>
<td>0.1</td>
<td>0.0</td>
<td>5.0</td>
</tr>
<tr>
<td>A6</td>
<td>TS</td>
<td>4.8</td>
<td>4.8</td>
<td>4.7</td>
<td>4.5</td>
<td>4.6</td>
<td>4.5</td>
<td>4.7</td>
<td>4.5</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>WX</td>
<td>4.9</td>
<td>4.8</td>
<td>4.9</td>
<td>4.8</td>
<td>4.9</td>
<td>4.9</td>
<td>5.1</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>WX2</td>
<td>4.7</td>
<td>1.4</td>
<td>0.1</td>
<td>0.0</td>
<td>4.9</td>
<td>1.3</td>
<td>0.1</td>
<td>0.0</td>
<td>4.9</td>
</tr>
<tr>
<td>A7</td>
<td>TS</td>
<td>4.7</td>
<td>4.9</td>
<td>4.6</td>
<td>4.3</td>
<td>4.6</td>
<td>4.6</td>
<td>4.3</td>
<td>4.5</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>WX</td>
<td>4.7</td>
<td>5.0</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.1</td>
<td>4.9</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>WX2</td>
<td>4.8</td>
<td>1.4</td>
<td>0.1</td>
<td>0.0</td>
<td>5.0</td>
<td>1.4</td>
<td>0.0</td>
<td>0.0</td>
<td>4.9</td>
</tr>
</tbody>
</table>

We can see in Table 3 that although TS appears to be slightly conservative, TS and WX have stable empirical levels close to 5% for each null hypothesis. Moreover, WX2 breaks down as soon as a within-pair dependence structure exists.

Examining the comparative powers of WX and TS from the whole simulation study, it is seen that except for the pure location shifts versions of $A2$, $A5$, $A6$, TS clearly outperforms WX. Obvious cases are pure scale shift models ($A7$ and the pure scale shift versions of $A2$ and $A5$) for which the power of WX is very low whatever the value of $\tau$ is, leading to a very large gain by TS. The same conclusion holds for the mixture model $A3$. Also, TS has far better ability to detect alterations in both location and scale (the location-scale shift versions of $A2$, $A5$, $A6$) as well as a non constant location shift ($A4$), particularly when the within-pair association is weak. When $\tau$ gets larger, TS and WX get more powerful and the differences of performances between them thin down. The relative behavior of the two tests is similar to a lesser degree for the local alteration hypothesis $A1$, but there, both WX and TS are weak for very localized features ($\mu = 0.33$). Finally, even for pure location shift models, the power of TS increases with the association degree leading to closely related powers for TS and WS as $\tau$ gets larger, at least for moderate sample sizes.
Focusing on the two-sample problem, WX2 and WX have close powers in most cases and are generally outperformed by TS. Nevertheless, we know that both WX2 and WX are low in testing more subtle alterations than pure location shifts. In such cases, it seems more appropriate to compare TS with other tests. We reported from Figure 5.1 of Fan [7], Figure 2 of Janic-Wróblewska and Ledwina [15] the approximate powers of F, WL and CVM for the models \(A_1, A_3\) and \(A_7^2\). They are displayed in Table 4. For hypothesis \(A_1\), it is seen that for a strength feature \(\mu = 0.67\), TS works as well as F, but it breaks down for a small value of \(\mu\). For \(A_3\) and \(A_7\) the power of TS is just slightly lower than that of WL and F and is far better than CVM.

Overall, except for pure location shift problems, our test should be preferred to WX whatever the within-pair association degree is. Moreover, even if it is not specifically designed for the classical two-sample case, it has been shown on models \(A_3\) and \(A_7\) to be far better than classical tests such as CVM and WX2, and it does not break down in front of more sophisticated tests like F and WL.

\footnote{The values of the test WL for \(n = 100\) and model \(A_7\) have been given to us by an anonymous referee.}
Figure 2: Empirical powers (in %) of TS (full line), WX (dashed) and WX2 (square), for alternative $A_2$.

5. Application to ChIP-on-chip data

ChIP-on-chip, also known as genome-wide location analysis, is a technique combining Chromatin ImmunoPrecipitation and microarrays for isolation and identification of the DNA sequences occupied by specific DNA binding proteins in cells. The identified binding sites may be used as a basis for annotating functional elements in genomes. The types of functional elements that one can identify using ChIP-on-chip include promoters, enhancers, repressor and silencing elements, insulators, boundary elements, and sequences that control DNA replication. Microarrays design can contain up to 1,000,000 probes, each measuring the binding intensity of a protein (histone, transcription factor,...). Probes are generally randomly distributed over the genome. ChIP-on-chip uses bi-channel microarrays. One channel is the total DNA, and is used as a reference measure (Input). The other channel measures the DNA quantity that binds the immunoprecipitated protein (IP). The relative ratio IP/INPUT is used as a binding score. A probe is enriched in DNA (that means that it is bound by a protein) if the score is higher than the noise. Figure 6 shows the probes densities for the red and green channels using the whole dataset with $N = 243494$ observations (probes). Data come from experiments realized to query the occupancy of ETS1 transcription factor in a human T-cell line (for more details see Hollenhorst et al. [12]).

We can observe that there is a location and scale shift between the paired Input and IP densities. We applied our test to samples of size $n \in \{50, 100, 200\}$, randomly
drawn from the original dataset. We computed the powers of TS, WX and WX2 using 20000 samples. For TS, we fixed \( d(n) = 10 \) and Legendre polynomials were used associated to the uniform distribution as reference measure \((\mu)\). Results are given in Table 5.

Our test outperforms WX and WX2, especially when more than 50 observations are used.

6. Proofs

6.1. Proof of Proposition 1

The application of the central limit theorem yields that \( U_n(k) \) converges in distribution under \( H_0 \) to a Gaussian vector with mean zero and variance-covariance matrix \( W(k) \). By \((A_k)\), it is easily seen that \( \hat{W}_n(k) \) is a consistent estimator of \( W(k) \). So, the random vector \( \hat{W}_n(k)^{-1/2}U_n(k) \) converges to a standard Gaussian vector by Slutsky’s theorem and thus, \( T_n(k) \) converges to a chi-squared random variable with \( k \) degrees of freedom.

6.2. Proof of Theorem 1

Let us denote by \( P_0 \) the probability under \( H_0 \). By Proposition 1, \( T_n(1) \) converges to a chi-squared random variable with one degree of freedom. So, in view to prove
Figure 4: Empirical powers (in %) of TS (full line), WX (dashed) and WX2 (square), for alternative A4.

Table 4: Approximated empirical powers (in %) for TS, WX2, WL, F, CVM in the two-sample problem, for A1 and n = 200, A3 and n = 50, A7 and n = 50 (right) and n = 100 (italic).

<table>
<thead>
<tr>
<th>A1 (n=200)</th>
<th>μ = 0.33</th>
<th>μ = 0.67</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>15</td>
<td>100</td>
</tr>
<tr>
<td>WX2</td>
<td>10</td>
<td>77</td>
</tr>
<tr>
<td>F</td>
<td>60</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A3 (n=50)</th>
<th>μ = 0.33</th>
<th>μ = 0.67</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>15</td>
<td>80</td>
</tr>
<tr>
<td>WX2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>CVM</td>
<td>5</td>
<td>55</td>
</tr>
<tr>
<td>WL</td>
<td>15</td>
<td>90</td>
</tr>
<tr>
<td>F</td>
<td>20</td>
<td>90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A7 (n=50,100)</th>
<th>σ = 1.2</th>
<th>σ = 1.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>WX2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>WL</td>
<td>15</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>96</td>
</tr>
<tr>
<td>F</td>
<td>15</td>
<td>70</td>
</tr>
<tr>
<td>CVM</td>
<td>5</td>
<td>20</td>
</tr>
</tbody>
</table>

Theorem 1, it is enough to show that $P_0(S_n \geq 2)$ tends to zero as $n$ tends to infinity.
Figure 5: Empirical powers (in %) of TS (full line), WX (dashed) and WX2 (square), for alternative $A7$.

Table 5: Empirical powers (in %) for TS, WX and WX2 for the ChIP-on-chip data.

<table>
<thead>
<tr>
<th>n</th>
<th>TS</th>
<th>WX</th>
<th>WX2</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>48.050</td>
<td>54.105</td>
<td>13.785</td>
</tr>
<tr>
<td>100</td>
<td>67.335</td>
<td>62.695</td>
<td>21.655</td>
</tr>
<tr>
<td>200</td>
<td>86.470</td>
<td>68.360</td>
<td>29.065</td>
</tr>
</tbody>
</table>

Let us set $r_n(k) = (k - 1) \log n$. By definition of $S_n$, we have

$$
\mathbb{P}_0(S_n \geq 2) = \sum_{k=2}^{d(n)} \mathbb{P}_0(S_n = k) \leq \sum_{k=2}^{d(n)} \mathbb{P}_0 \left( T_n(k)^{1/2} \geq \sqrt{r_n(k)} \right). \tag{9}
$$

Hereafter, we shall bound the general term of the expansion at the right hand side of (9). As $\hat{W}_n(k)$ is a positive-definite symmetric matrix, we have that on the one hand it is a diagonalizable matrix with strictly positive eigenvalues and we denote by $\lambda_{\text{min}}(k)$ its smallest eigenvalue. On the other hand, $\hat{W}_n(k)^{-1}$ exists and its greatest eigenvalue equals $1/\lambda_{\text{min}}(k)$. Using a basic property of positive-definite symmetric matrices and
Let us bound the two terms at the right hand side of (13). For the first term, Bienaymé-Tchebychev’s inequality yields

$$
\mathbb{P}_0 \left( \|U_n(k)\| \geq \sqrt{\frac{r_n(k)\lambda_{\text{min}}(k)}{2}} \right) \leq \frac{2\mathbb{E}_0(\|U_n(k)\|^2)}{\lambda_{\text{min}}(k)r_n(k)}. \tag{14}
$$
Using the independence of the pairs \((X_s, Y_s)_{1 \leq s \leq n}\), we get

\[
E_0 \left( \|U_n(k)\|^2 \right) = E_0 \left( \|V_1(k)\|^2 \right) \leq \sqrt{E_0 \left( \|V_1(k)\|^4 \right)}.
\]

the last inequality arising from Jensen’s inequality. Finally, Inequality (14) and the above result yield

\[
P_0 \left( \|U_n(k)\| \geq \sqrt{\frac{r_n(k)\lambda_{\text{min}}(k)}{2}} \right) \leq \frac{2}{r_n(k)\lambda_{\text{min}}(k)} E_0 \left( \|V_1(k)\|^4 \right) \quad (15)
\]

We now bound the second term at the right hand side of (13). Let us denote by \(\||.||\) the spectral norm on the space of \(k \times k\) matrices. More precisely, for a given matrix \(A\),

\[
\|A\| = \sup_{\|x\|=1, x \in \mathbb{R}^k} \|Ax\|.
\]

Now, \(\hat{W}_n(k)\) and \(W(k)\) are self-adjoint compact (finite-rank here) operators of \(\mathbb{R}^k\) so that applying Corollary 2.3 p31 of Gohberg and Krejn [9], we have

\[
|\lambda_{\text{min}}(k) - \hat{\lambda}_{\text{min}}(k)| \leq \|\hat{W}_n(k) - W(k)\|.
\]

Therefore,

\[
P_0 \left( \hat{\lambda}_{\text{min}}(k) \leq \frac{\lambda_{\text{min}}(k)}{2} \right) = P_0 \left( \lambda_{\text{min}}(k) - \hat{\lambda}_{\text{min}}(k) \geq \frac{\lambda_{\text{min}}(k)}{2} \right) \leq P_0 \left( \|\hat{W}_n(k) - W(k)\| \geq \frac{\lambda_{\text{min}}(k)}{2} \right)
\]

so that by Markov’s and Jensen’s inequalities, we have

\[
P_0 \left( \hat{\lambda}_{\text{min}}(k) \leq \frac{\lambda_{\text{min}}(k)}{2} \right) \leq \frac{2}{\lambda_{\text{min}}(k)} \sqrt{E_0 \left( \|\hat{W}_n(k) - W(k)\|^2 \right)}.
\]

Moreover, calculations using the definition of the spectral norm and the independence of the pairs \((X_s, Y_s)_{1 \leq s \leq n}\) yield

\[
P_0 \left( \hat{\lambda}_{\text{min}}(k) \leq \frac{\lambda_{\text{min}}(k)}{2} \right) \leq \frac{2}{n^{1/2} \lambda_{\text{min}}(k)} \sqrt{E_0 \left( \|V_1(k)\|^4 \right)}.
\]

18
Substituting (19) and (15) in (13) yield
\[
\mathbb{P}_0(T_n(k)^{1/2} \geq \sqrt{r_n(k)}) \leq \frac{2 \left( \mathbb{E}_0(||V_1(k)||^4) \right)^{1/2}}{\lambda_{\min}(k) \left( \frac{1}{r_n(k)} + \frac{1}{n^{1/2}} \right)}. \tag{20}
\]
Finally, using (A) and the fact that \(\sum_{k=1}^{n} \frac{1}{k} = O_{\cdot \cdot \cdot}(\log n)\), we have
\[
\mathbb{P}_0(S_n \geq 2) \leq \frac{2M^{1/2}}{\inf_{1 \leq k \leq d(n)} \lambda_{\min}(k)} \left( \frac{\log d(n)}{\log n} + \frac{d(n)}{n^{1/2}} \right).
\]

Theorem 1 obtains by (B), as soon as we have shown that \((\lambda_{\min}(k))_{k>0}\) is a decreasing sequence. For that, let \(X = (x_i)_{1 \leq i \leq k}\) be the unit eigenvector associated with \(\lambda_{\min}(k)\) and let \(Y = (y_i)_{1 \leq i \leq k+1}\) be the vector such that \(x_i = y_i\) for all \(1 \leq i \leq k\) and \(y_{k+1} = 0\). We have for all \(k\)
\[
\lambda_{\min}(k + 1) \leq Y^t W(k + 1) Y = X^t W(k) X = \lambda_{\min}(k).
\]

6.3. **Proof of Proposition 2**

Let us denote by \(\mathbb{P}_1\) the probability under \(H_1^+\). First, let us notice that \(\lim_{n \to \infty} d(n) > q\) and let us prove that
\[
\lim_{n \to \infty} \mathbb{P}_1(S_n < q) = 0. \tag{21}
\]
For \(k < q\), we have \(\mathbb{P}_1(S_n = k) \leq \mathbb{P}_1(T_n(k) - k \log(n) \geq T_n(q) - q \log(n))\). Then, on the one hand, by the law of large numbers, the variable \(U_n(q) / \sqrt{n}\) converges in \(\mathbb{P}_1\)-probability to a non-null vector and, since \(W_n(q)\) is a positive definite matrix we have \(T_n(q) = U_n(q)^t W_n(q)^{-1} U_n(q) = O_P_1(n)\). So,
\[
T_n(q) - q \log(n) \longrightarrow +\infty \text{ in probability} \tag{22}
\]
under \(H_1^+\). On the other hand, we have
\[
\mathbb{P}_1(T_n(k) - k \log(n) \geq - \log(n)) = \mathbb{P}_1(T_n(k)^{1/2} \geq \sqrt{r_n(k)}).
\]
Since for all \(1 \leq j \leq k, a_j = b_j\), equation (20) holds with \(\mathbb{P}_0\) replaced by \(\mathbb{P}_1\). Therefore, the right-hand side of (23) tends to zero as \(n\) goes to infinity as soon as (A) holds, so that
\[
T_n(k) - k \log(n) \longrightarrow -\infty \text{ in probability} \tag{24}
\]
under \(H_1^+\). Therefore, \(\mathbb{P}_1(S_n = k)\) converges to zero for all \(k < q\), which leads to (21). Now, we have for all \(x \in \mathbb{R}\),
\[
\mathbb{P}_1(\bar{T}_n \leq x) \leq \mathbb{P}_1(\bar{T}_n \leq x, S_n \geq q) + \mathbb{P}_1(S_n < q) \leq \mathbb{P}_1(T_n(q) \leq x) + \mathbb{P}_1(S_n < q). \tag{25}
\]
Using (21) and (22), the two terms at the right-hand side of (25) tend to zero as \(n\) goes to infinity. So, we have
\[
\lim_{n \to \infty} \bar{T}_n = +\infty \text{ in probability}.
\]


