Homogeneity and change-point detection tests for multivariate data using rank statistics

Titre: Tests d'homogénéité et de détection de ruptures pour des données multivariées en utilisant des statistiques de rang

Alexandre Lung-Yut-Fong[1], Céline Lévy-Leduc[2] and Olivier Cappé[1]

Abstract: Detecting and locating changes in highly multivariate data is a major concern in several current statistical applications. In this context, the first contribution of the paper is a novel non-parametric two-sample homogeneity test for multivariate data based on the well-known Wilcoxon rank statistic. The proposed two-sample homogeneity test statistic can be extended to deal with ordinal or censored data as well as to test for the homogeneity of more than two samples. We also provide a detailed analysis of the power of the proposed test statistic (in the two sample case) against asymptotic local shift alternatives. The second contribution of the paper concerns the use of the proposed test statistic to perform retrospective change-point detection. It is first shown that the approach is computationally feasible even when looking for a large number of change-points thanks to the use of dynamic programming. Computable asymptotic *p*-values for the test are available in the case where a single potential change-point is to be detected. The proposed approach is particularly recommendable in situations where the correlations between the coordinates of the data are moderate, the marginal distributions are not well modelled by usual parametric assumptions (e.g., in the presence of outliers) and when faced with highly variable change patterns, for instance, if the potential changes only affect subsets of the coordinates of the data.

Résumé : La détection et la localisation de changements dans des données de très grande dimension est un problème majeur dans plusieurs domaines d'applications. Dans ce contexte, la première contribution de notre papier est un nouveau test d'homogénéité non-paramétrique à deux échantillons pour des données multivariées fondé sur la statistique de rang de Wilcoxon. Le test d'homogénéité à deux échantillons que nous proposons peut être étendu au cas de données censurées et pour proposer un test d'homogénéité pour plus de deux échantillons. Nous proposons également une analyse détaillée du calcul de la puissance de notre statistique de test vis à vis de certaines alternatives locales. La seconde contribution de notre papier concerne l'utilisation de notre statistique de test pour faire de la détection rétrospective de ruptures. Nous montrons que notre méthode peut-être implémentée de façon efficace d'un point de vue algorithmique grâce à la programmation dynamique et nous proposons une méthode pour calculer les *p*-valeurs. Nous recommandons particulièrement notre approche dans les situations suivantes : lorsque les corrélations entre les coordonnées des observations sont modérées, lorsque les lois marginales ne peuvent pas être modélisées par les lois paramétriques usuelles ou lorsque les changements n'affectent qu'une partie des coordonnées des observations.

---

[1] LTCI, CNRS & Telecom ParisTech.
   E-mail: alexandre@lungyutfong.fr and E-mail: olivier.cappe@telecom-paristech.fr
[2] AgroParisTech / UMR INRA MIA 518.
   E-mail: celine.levy-leduc@agroparistech.fr

1. Introduction

Detection and location of distributional changes in data is a major statistical challenge that arises in many different contexts. This very general concern can be particularised to more specific tasks such as segmentation, novelty detection or significance tests. In this contribution, we focus on two types of problems: *homogeneity testing*, where the statistician is presented with pre-specified groupings of the data that are believed to be comparable, and *change-point detection*, in which a series –most often, a time series– is to be segmented into homogeneous contiguous regions. These two tasks are obviously related but the latter is more challenging as the appropriate groupings of the data are unknown, although one does have the strong prior assumption that homogeneous regions of the data are contiguous. Homogeneity testing and/or change-point detection are instrumental in applications that range from the surveillance of industrial processes Basseville and Nikiforov (1993), to computer security Tartakovsky et al. (2006); Lévy-Leduc and Roueff (2009), processing of audiovisual data Désobry et al. (2005), financial and econometric modelling Bai and Perron (2003); Talih and Hengartner (2005), health monitoring Brodsky and Darkhovsky (2000), or bioinformatics Picard et al. (2005); Vert and Bleakley (2010).

In light of the important available literature on change-point detection it is important to make two additional distinctions. First, in many cases the data to be analysed can be assumed to present some form of global reproducibility and to include several instances of actual changes. In this case, it seems reasonable to fit a model to the data to profit from the available statistical information regarding various relevant aspects of the problem such as the distribution of the data in the absence of change, the typical change-point patterns, etc. In such situations, very convincing results have been demonstrated using Bayesian approaches due to the existence of efficient computational methods to explore the posterior distribution, even when using very flexible models Barry and Hartigan (1992); Fearnhead (2006). In contrast, in this contribution, we consider scenarios in which the data are either scarce or very variable or where potential changes occur somewhat infrequently. In this alternative context, the goal is to develop approaches that make as few assumptions as possible regarding the underlying distribution of the data or the nature of the changes and that do not rely on the observation of actual change patterns. The second important distinction is that many works in the time series literature consider the *online* or *sequential* change-point detection framework Siegmund (1985); Poor and Hadjiliadis (2009); Tartakovsky et al. (2014). in which the data have to be processed on-the-fly or with minimal delay using for instance the CUSUM algorithm initially proposed by Page (1954). In the following, we consider the opposite situation, sometimes referred to as *retrospective* analysis, in which all the data to be tested have been recorded and are available for analysis.

In this context, the first contribution of this work consists in novel homogeneity tests for dealing with possibly high-dimensional multivariate observations. We focus on situations where the potential changes are believed to have a strong impact on the mean but where the distribution of the data is otherwise mostly unknown. For scalar observations, there are well-known robust solutions for testing homogeneity in this context such as the Wilcoxon/Mann–Whitney or Kruskal-Wallis procedures Lehmann (1975) to be further discussed below. For multivariate observations, the situation is far more challenging as one would like to achieve robustness with respect both to the form of the marginal distributions and to the existence of correlations (or other form of dependence) between coordinates. The latter aspect has been addressed in a series of works by

Möttönen et al. (1997); Hettmansperger et al. (1998); Oja (1999); Topchii et al. (2003) who studied multivariate extensions of sign and rank tests. These tests are affine invariant in the sense that they behave similarly to Hotelling's T^2 test (see Section 2.3 below) —which is optimal for Gaussian distributions– for general classes of multivariate distributions having ellipsoidal contours (e.g., for multivariate t distributions). For a very recent review of nonparametric multivariate rank tests, we refer the reader to Jurečková and Kalina (2012). Another promising approach investigated by several recent works consists in using kernel-based methods Désobry et al. (2005); Gretton et al. (2006); Harchaoui et al. (2008). In our experience however, these methods that can achieve impressive results for moderately multidimensional data or in specific situations (*e.g*., if the data lie on a low-dimensional manifold) lack robustness when moving to larger dimensions. In particular, as illustrated in Section 4.1 below, kernel-based methods are not robust with respect to the presence of contaminating noise and to the fact that the changes to be detected may only affect a subset of the components of the high-dimensional data. The latter scenario is of very important practical significance in applications where the data to be analysed consist in exhaustive recordings of complex situations that are only partly affected by possible changes (see Lévy-Leduc and Roueff (2009) for an example regarding the detection of computer attacks). The method proposed in this work is based on a combination of marginal rank statistics, following the pioneering idea of Wei and Lachin (1984). Compared to the latter, our contribution is twofold: first, we show how to correct the bias that appears in the test statistic proposed by Wei and Lachin (1984) whenever the two samples are not balanced in size; we then show how to extend this idea to the case of more than two groups. The numerical simulation presented in Section 4 confirms that the proposed test statistic is significantly more robust than kernel-based methods or approaches based on least-squares or Hotelling's T^2 statistics. These empirical observations are also supported by the results detailed in Section 2.3 regarding the power of the proposed test statistic (in the two sample case) with respect to asymptotic local shift alternatives. In particular, although the proposed test is not strictly affine invariant [1] it compares favourably to Hotelling's T^2 statistic for important classes of distributions under the assumption that the condition number of the correlation matrix of the data is not too large.

We then consider the use of the proposed approach for change-point detection by optimising the test statistic over the –now considered unknown– positions of the segment boundaries. Although simple this idea raises two type of difficulties. The first one is computational as the resulting optimisation task is combinatorial and cannot be solved by brute force enumeration when there is more than one change-point (that is, two segments). In the literature this issue has been previously tackled either using dynamic programming Bai and Perron (2003); Harchaoui and Cappé (2007) or more recently using Lasso-type penalties Harchaoui and Lévy-Leduc (2010); Vert and Bleakley (2010). We show that the generic dynamic programming strategy is applicable to the proposed test statistic making it practically suitable for retrospective detection of multiple change-points. The second difficulty is statistical as the optimisation with respect to the change-point locations modifies the distribution of the test statistics. Thus, the design of quantitative criterions for assessing the significance of the test is a challenging problem in this context. This issue has been considered before, mostly in the case of a single change-point, for various test statistics Csörgő and Horváth (1997); Chen and Gupta (2000). In many cases the asymptotic distribution of the test

---

[1] It is highly unlikely that there exist statistics which are both invariant under monotonic transformations of the marginals –as the proposed method is– and affine invariant.

remains hard to characterise and must be calibrated using Monte Carlo simulations. We show that for a simple modification of the proposed rank-based statistic, one can indeed obtain computable asymptotic p-values that can be used to assess the significance of the test when looking for a single change-point.

Contributions of the paper. Section 2 is devoted to homogeneity testing, starting first with the two-sample case and then considering the more general situation where several predefined groups are available. The end of Section 2 is dedicated to the study of the asymptotic behaviour of the two-sample homogeneity test under local shift alternatives. In Section 3, the proposed test statistic is modified to provide a method for detecting and locating change-points, with the computation of p-values (in the single change-point case) being discussed in Section 3.2. The results of numerical experiments carried out both on simulated and on real data are then reported in Section 4.

2. Testing for Homogeneity

We first tackle in this Section the so-called two-sample problem, that is testing the homogeneity between two partitions of data. The proposed test statistic is then extended in Section 2.2 to deal with more than two groups of data.

2.1. Two-sample homogeneity test

Consider n K-dimensional multivariate observations $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and denote by $X_{i,k}$ the kth coordinate of X_i, such that $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,K})'$, where the prime is used to denote transposition. We consider the classical statistic test framework with the null (or baseline) hypothesis, (H_0): "$(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ are identically distributed random vectors", and the alternative hypothesis, (H_1): "$(\mathbf{X}_1, \ldots, \mathbf{X}_{n_1})$ are distributed under $\mathbb{P}_1$ and $(\mathbf{X}_{n_1+1}, \ldots, \mathbf{X}_n)$ under $\mathbb{P}_2$, with $\mathbb{P}_1 \neq \mathbb{P}_2$". In this setting, the potential change point n_1 is assumed to be given but the data distributions are fully unspecified both under (H_0) and (H_1). The proposed test statistic extends the well-known Wilcoxon/Mann–Whitney rank-based criterion to multivariate data by considering the asymptotic joint behaviour of the rank statistics that can be computed from each coordinate of the observations. For k in $\{1, \ldots, K\}$, define the vector-valued statistic $\mathbf{U}_n(n_1) = (U_{n,1}(n_1), \ldots, U_{n,K}(n_1))'$ by

$$U_{n,k}(n_1) = \frac{1}{\sqrt{nn_1(n-n_1)}} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} \left\{ \mathbf{1}(X_{i,k} \leq X_{j,k}) - \mathbf{1}(X_{j,k} \leq X_{i,k}) \right\} . \tag{1}$$

Although the form of the statistic given above is more appropriate for mathematical analysis as well as for discussing possible generalisations of the approach (see Section 2.1.2 below), it is important to realise that $U_{n,k}(n_1)$ is related to the classical Wilcoxon/Mann–Whitney statistic computed from the series $X_{1,k}, \ldots, X_{n,k}$. Assuming that there are no ties in the data, let $R_j^{(k)}$ denote the rank of $X_{j,k}$ among $(X_{1,k}, \ldots, X_{n,k})$, that is, $R_j^{(k)} = \sum_{i=1}^{n} \mathbf{1}(X_{i,k} \leq X_{j,k})$. Noticing that $\sum_{j=1}^{n} R_j^{(k)} = n(n+1)/2$, it is then easily verified that $U_{n,k}(n_1)$ can be equivalently defined as

$$U_{n,k}(n_1) = \frac{2}{\sqrt{nn_1(n-n_1)}} \sum_{i=1}^{n_1} \left(\frac{n+1}{2} - R_i^{(k)} \right) = \frac{2}{\sqrt{nn_1(n-n_1)}} \sum_{j=n_1+1}^{n} \left(R_j^{(k)} - \frac{n+1}{2} \right) . \tag{2}$$

This alternative form of $U_{n,k}(n_1)$ is more appropriate for computational purposes as discussed in Section 2.1.1 below. For convenience, we denote by $\hat{F}_{n,k}(t) = n^{-1}\sum_{j=1}^{n} \mathbf{1}(X_{j,k} \leq t)$ the empirical cumulative distribution function (c.d.f. in short) of the kth coordinate, such that $\hat{F}_{n,k}(X_{i,k}) = R_i^{(k)}/n$. Let $\hat{\Sigma}_n$ denote the K-dimensional empirical covariance matrix defined by

$$\hat{\Sigma}_{n,kk'} = \frac{4}{n}\sum_{i=1}^{n}\{\hat{F}_{n,k}(X_{i,k}) - 1/2\}\{\hat{F}_{n,k'}(X_{i,k'}) - 1/2\},\ 1 \leq k, k' \leq K\ . \tag{3}$$

The test statistic that we propose for assessing the presence of a potential change in n_1 is defined as

$$S_n(n_1) = \mathbf{U}_n(n_1)'\hat{\Sigma}_n^{-1}\mathbf{U}_n(n_1)\ . \tag{4}$$

Theorem 1 below (proved in Appendix A) gives the limiting behaviour of the test statistic $S_n(n_1)$ under the null hypothesis.

Theorem 1. *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}, \mathbf{X}_{n_1+1}, \ldots, \mathbf{X}_n$ *be* $\mathbb{R}^K$-*valued i.i.d. random vectors, such that for all k in $\{1, \ldots, K\}$, the cumulative distribution function F_k of $X_{1,k}$ is a continuous function. Assume that $n_1/n \to t_1$ in (0,1) as n tends to infinity, and that the $K \times K$ covariance matrix Σ defined by*

$$\Sigma_{kk'} = 4\,\mathrm{Cov}\left(F_k(X_{1,k}); F_{k'}(X_{1,k'})\right),\ 1 \leq k, k' \leq K\ , \tag{5}$$

is positive definite. Then, the test statistic $S_n(n_1)$ defined in (4) *converges in distribution to a χ^2 distribution with K degrees of freedom.*

Theorem 1 shows that the proposed test is well normalised with respect to the dimension K, the length n of the data and the postulated change-point location n_1. It is asymptotically distribution-free in the sense that its limiting behaviour under (H_0) does not depend on the distribution of the data. By construction, it is also invariant under any monotonic transformation of the coordinates of $\mathbf{X}_i$.

The matrix Σ, which corresponds to the asymptotic covariance matrix of the vector $\mathbf{U}_n(n_1)$ is equal, up to a multiplicative constant, to the Spearman correlation matrix of X_i Lehmann (1975); van der Vaart (1998). This is a well-known robust measure of dependence that appears in particular when using copula models. A sufficient condition for ensuring the invertibility of Σ is thus that no linear combination of the $F_k(X_{1,k})$'s should be almost surely equal to a constant, which is arguably a very weak condition. It is easily checked that the diagonal elements of Σ are all equal to $1/3$ and that $\Sigma_{k\ell} = \Sigma_{\ell k} = 0$ whenever the k-th and ℓ-th coordinates of $\mathbf{X}_i$ are independent. It appears, in practice, that the diagonal elements of $\hat{\Sigma}_n$ converge very rapidly to $1/3$ value and we did not observe any significant improvement when trying to take into account this fact when estimating Σ.

Theorem 1 defines the asymptotic false alarm rate associated with the test statistic $S_n(n_1)$. The test is consistent (*i.e.*, its power tends to 1) for all alternatives such that the condition ensuring the consistency of the standard Wilcoxon/Mann–Whitney two-sample test holds true for at least one coordinate. More formally, the proposed test is consistent when there exists k in $\{1, \ldots, K\}$ such that $\mathbb{P}(X_{k,1} \leq X_{k,n}) \neq 1/2$. In the scalar case, this condition is known to hold for general classes of changes such as shift (change-in-the-mean) models or scale (multiplicative) change for positive variables. We defer to Section 2.3 a more detailed investigation of the asymptotic power of the test in the case of multivariate shift alternatives.

As Theorem 1 is an asymptotic result, we have carried out Monte Carlo simulations to asses the accuracy of the approximation for finite sample sizes. Using data with independent coordinates [2] we found that the distribution of $S_n(n_1)$ defined in (4) can be considered close enough to the limiting distribution, as measured by the Kolmogorov-Smirnov test at level 1%, when n is at least 8 times larger than K. For instance, for $K = 20$, a value of $n = 210$ was sufficient; $K = 100$ required $n = 840$ samples, etc. The empirical density of the test statistics is illustrated in the upper part of Figure 1 when $K = 10$ and $n = 200$.
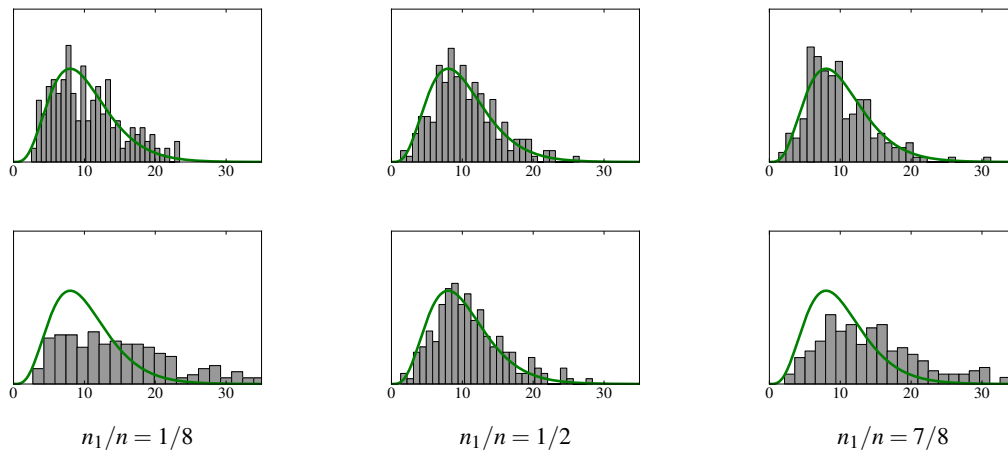

$n_1/n = 1/8$ $n_1/n = 1/2$ $n_1/n = 7/8$

FIGURE 1. *Histograms of the statistics $S_n(n_1)$ (top) and Wei and Lachin's (bottom) compared to the χ_K^2 p.d.f., as a function of the ratio n_1/n. Data corresponds to $n = 200$ samples of a ten-dimensional standard Gaussian distribution.*

The pioneering work of Wei and Lachin (1984) describes a result analogous to that of Theorem 1 in the case of possibly upper-censored data. However, the proof technique used by Wei and Lachin (1984) relies on a different interpretation of Σ which is used to derive a weighting matrix that is not equal to $\hat{\Sigma}_n$ as defined in (3). In contrast, our proof (see Appendix A) is based on a standard argument for U-statistics (the Hoeffding decomposition) that directly returns an expression of Σ in terms of covariances, for which usual estimators, such as $\hat{\Sigma}_n$, may be used. The test statistics of Wei and Lachin (1984) thus differs from $S_n(n_1)$ and turns out to be biased in cases where $n_1 \neq n/2$, *i.e.*, when the change does not occur in the centre of the observation frame, as shown by the bottom part of Figure 1. This bias becomes problematic when the potential change location n_1 is unknown because the values of $S_n(n_1)$ for different values of n_1 cannot be validly compared.

2.1.1. *Implementation Issues*

As noted above, the vector $(U_{n,k}(n_1))_{1 \leq k \leq K}$ should be computed from the marginal rank statistics in the form given in (2). $\hat{\Sigma}_n$ also is a simple function of those marginal ranks. Thus, $(U_{n,k}(n_1))_{1 \leq k \leq K}$ can be computed in $\mathcal{O}(Kn\log(n))$ operations using a sort for computing the ranks, as the average numerical complexity of usual sorting algorithms is of the order of $n\log(n)$ operations. The

---

[2] Note that by construction, the test statistic is then fully invariant with respect to the precise distribution used for the Monte Carlo simulations.

computation of $\hat{\Sigma}_n$ then requires $\mathcal{O}(K^2 n)$ operations and its inversion $\mathcal{O}(K^3)$ operations. Note that if the test statistic needs to be recomputed at a neighbouring index, say $n_1 + 1$, neither the ranks nor $\hat{\Sigma}_n$ and its inverse need to be recomputed. Hence the number of additional operations required to compute $\hat{S}_n(n_1 + 1)$ is indeed very limited.

In some situations, it may happen that the empirical estimate $\hat{\Sigma}_n$ becomes ill-conditioned rendering its inversion numerically unstable. Wei and Lachin (1984) suggested to circumvent the problem by adding some small positive value to the diagonal elements of $\hat{\Sigma}_n$. It is important to realise however that a particular case where Σ itself can be ill-conditioned is when coordinates of $\mathbf{X}_1$ are strongly dependent. In the limiting case where coordinates of $\mathbf{X}_1$, say two of them for illustration, are duplicated, Σ becomes a matrix of rank $K - 1$. In such a case, the correct statistic is obtained by simply discarding one of the coordinates that are duplicated. Hence, to regularise $\hat{\Sigma}_n$ in such cases, we suggest inverting it using its Moore-Penrose pseudo inverse: if $\hat{\Sigma}_n = USU'$ denotes the singular value decomposition of $\hat{\Sigma}_n$, with $S = \mathrm{diag}(s_1, \ldots, s_K)$ being the diagonal matrix of eigenvalues of Σ, then the pseudo inverse $\Sigma_n^\dagger$ is defined as $U' \mathrm{diag}(s_1^\dagger, \ldots, s_K^\dagger) U$ where $s_i^\dagger = s_i^{-1} \mathbf{1}(s_i > \varepsilon)$ and ε is a fixed positive threshold. Instead of relying on the asymptotic result of Theorem 1, it is suggested to compare $S_n(n_1)$ to the quantiles of the $\chi^2_{K'}$ distribution, where K' is the number of non-null values among the $s_i^\dagger$'s. As already mentioned however, some terms of $\hat{\Sigma}_n$ appear to converge very rapidly and the matrix is only very rarely ill-conditioned, even when n is only slightly larger than K. On the other hand, the regularised variant described above was found to be effective for dealing with signals whose coordinates can be extremely dependent, *e.g.*, if there is a quasi-deterministic relationship between two coordinates.

2.1.2. Discrete, missing or censored data

Theorem 1 requires the continuity of the c.d.f. F_k of each coordinate; hence it is not directly applicable, for instance, to discrete variables. In such cases however, Theorem 1 is still valid upon redefining Σ as

$$\Sigma_{kk'} = \mathbb{E}\left[\{F_k(X_{1,k}^-) + F_k(X_{1,k}) - 1\}\{F_{k'}(X_{1,k'}^-) + F_{k'}(X_{1,k'}) - 1\} \right], \qquad (6)$$

where $F_k(x^-)$ denotes the left-limit of the c.d.f. in x. In this case, (2) has to be replaced by

$$U_{n,k}(n_1) = \frac{2}{\sqrt{nn_1(n-n_1)}} \sum_{j=n_1+1}^{n} \left\{ R_j^{(k)} - \frac{n + \sum_{i=1}^{n} \mathbf{1}(X_{i,k} = X_{j,k})}{2} \right\}.$$

Another useful extension concerns the case of censored or missing data that can be dealt with in great generality by introducing lower $\underline{X}_{i,k}$ and upper $\overline{X}_{i,k}$ censoring values such that $\underline{X}_{i,k} \le X_{i,k} \le \overline{X}_{i,k}$, where a strict inequality indicates censoring (for missing values, simply set $\underline{X}_{i,k} = -\infty$ and $\overline{X}_{i,k} = +\infty$). In this case, we define a modified statistics from the censoring bounds $\underline{X}_{i,k}$ and $\overline{X}_{i,k}$ by

$$U_{n,k}(n_1) = \frac{1}{\sqrt{nn_1(n-n_1)}} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} \left\{ \mathbf{1}(\overline{X}_{i,k} \le \underline{X}_{j,k}) - \mathbf{1}(\overline{X}_{j,k} \le \underline{X}_{i,k}) \right\}. \qquad (7)$$

It can be shown, adapting the arguments of Lung-Yut-Fong et al. (2012) who studied the use of the scalar statistic (7) for change-point detection, that Theorem 1 then holds with

$$\Sigma_{kk'} = \mathbb{E}\left[\{\overline{F}_k(\underline{X}_{1,k}) + \underline{F}_k(\overline{X}_{1,k}^-) - 1\}\{\overline{F}_{k'}(\underline{X}_{1,k'}) + \underline{F}_{k'}(\overline{X}_{1,k'}^-) - 1\}\right],\tag{8}$$

where $\overline{F}_k$ and $\underline{F}_k$ denote the c.d.f.'s of $\overline{X}_{1,k}$ and $\underline{X}_{1,k}$, respectively.

2.2. Testing homogeneity within several groups of data

In this section, the procedure presented so far is extended to deal with more than two groups of multivariate data. The resulting test statistic is again based on a proper combination of marginal statistics involved in the Kruskal-Wallis procedure that generalises the classical Wilcoxon-rank test when there are more than two groups of data.

Consider the null hypothesis that L given groups, $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}; \mathbf{X}_{n_1+1}, \ldots, \mathbf{X}_{n_2}; \ldots; \mathbf{X}_{n_{L-1}+1}, \ldots, \mathbf{X}_{n_L}$, share the same distribution, where we shall use the convention that $n_0 = 0$ and $n_L = n$. For j in $\{1, \ldots, n\}$ and k in $\{1, \ldots, K\}$, denote as previously by $R_j^{(k)}$ the rank of $X_{j,k}$ among $(X_{1,k}, \ldots, X_{n,k})$ that is, $R_j^{(k)} = \sum_{i=1}^n \mathbf{1}_{\{X_{i,\ell} \leq X_{j,\ell}\}}$. For ℓ in $\{0, \ldots, L-1\}$, define the average rank in group ℓ for the kth coordinate by $\bar{R}_\ell^{(k)} = (n_{\ell+1} - n_\ell)^{-1} \sum_{j=n_\ell+1}^{n_{\ell+1}} R_j^{(k)}$. Consider the following test statistic:

$$T(n_1, \ldots, n_{L-1}) = \frac{4}{n^2} \sum_{\ell=0}^{L-1} (n_{\ell+1} - n_\ell) \bar{\mathbf{R}}_\ell' \, \hat{\Sigma}_n^{-1} \, \bar{\mathbf{R}}_\ell,\tag{9}$$

where the vector $\bar{\mathbf{R}}_\ell$ is defined as $\bar{\mathbf{R}}_\ell = (\bar{R}_\ell^{(1)} - (n+1)/2, \ldots, \bar{R}_\ell^{(K)} - (n+1)/2)'$, and $\hat{\Sigma}_n$ is again the matrix defined in (3). Theorem 2, proved in Appendix B, describes the limiting behaviour of the test statistic $T(n_1, \ldots, n_{L-1})$ under the null hypothesis.

Theorem 2. *Assume that $(\mathbf{X}_i)_{1 \leq i \leq n}$ are $\mathbb{R}^K$-valued i.i.d. random vectors such that, for all k, the c.d.f. F_k of $X_{1,k}$ is a continuous function. Assume also that for $\ell = 0, \ldots, L-1$, there exists $t_{\ell+1}$ in $(0,1)$ such that $(n_{\ell+1} - n_\ell)/n \to t_{\ell+1}$, as n tends to infinity. Then, $T(n_1, \ldots, n_{L-1})$ defined in (9) satisfies*

$$T(n_1, \ldots, n_{L-1}) \xrightarrow{d} \chi^2((L-1)K), \ as \ n \to \infty,\tag{10}$$

where d denotes convergence in distribution and $\chi^2((L-1)K)$ is the chi-square distribution with $(L-1)K$ degrees of freedom.

Observe that (9) extends the classical Kruskal-Wallis test used for univariate observations to the multivariate setting. Indeed, when $K = 1$, (9) is equivalent to

$$T(n_1, \ldots, n_{L-1}) = \frac{12}{n^2} \sum_{\ell=0}^{L-1} (n_{\ell+1} - n_\ell) \left(\bar{R}_\ell^{(1)} - (n+1)/2\right)^2,\tag{11}$$

where we have replaced $\hat{\Sigma}_{n,11}$ by $\Sigma_{11} = 4\operatorname{Var}(F_1(X_{1,1})) = 4\operatorname{Var}(U) = 1/3$ (U denoting a uniform random variable on $[0,1]$). In the case where there is only one change-point, *i.e.*, when $L = 2$, (9)

reduces to the test statistic proposed in Section 2.1. Indeed, using (2), $T(n_1)$ can be rewritten as follows

$$T(n_1) = \frac{nn_1(n-n_1)}{n^2 n_1}\mathbf{U}_n(n_1)'\hat{\Sigma}_n\mathbf{U}_n(n_1) + \frac{nn_1(n-n_1)}{n^2(n-n_1)}\mathbf{U}_n(n_1)'\hat{\Sigma}_n\mathbf{U}_n(n_1)$$
$$= \mathbf{U}_n(n_1)'\hat{\Sigma}_n\mathbf{U}_n(n_1) = S_n(n_1) \,,$$

where $S_n(n_1)$ is defined in (4).

2.3. Power of the homogeneity test in the two sample case

In this section, we focus on the two sample homogeneity test statistic $S_n(n_1)$ defined in (4) using local alternatives consisting of multivariate shifts to investigate the statistical power of the proposed approach. We consider the following alternative hypotheses $(H_{1,n})$: "$(\mathbf{X}_1,\ldots,\mathbf{X}_{n_1})$ are i.i.d with common multivariate p.d.f. $f(\mathbf{x})$ and $(\mathbf{X}_{n_1+1},\ldots,\mathbf{X}_n)$ are i.i.d with density $f(\mathbf{x}-\delta/\sqrt{n})$", where f is symmetric, positive, and, continuously differentiable and δ denotes an arbitrary shift vector in $\mathbb{R}^k$. Our results can thus be compared directly with those obtained by Hettmansperger et al. (1998), Oja (1999) and Topchii et al. (2003) for affine invariant multivariate generalisations of rank tests.

We first recall the classical result pertaining to the Hotelling-T^2 test

$$H_n(n_1) = (n_1(n-n_1)/n)(\bar{\mathbf{x}}_{\mathbf{n_1}} - \bar{\mathbf{x}}_{\mathbf{n-n_1}})'\hat{C}_n^{-1}(\bar{\mathbf{x}}_{\mathbf{n_1}} - \bar{\mathbf{x}}_{\mathbf{n-n_1}}) \,,$$

where $\bar{\mathbf{x}}_{\mathbf{n_1}} = (\sum_{i=1}^{n_1}\mathbf{X}_i)/n_1$, $\bar{\mathbf{x}}_{\mathbf{n-n_1}} = (\sum_{i=n_1+1}^{n}\mathbf{X}_i)/(n-n_1)$, and, $\hat{C}_n$ is the empirical covariance matrix of the $\mathbf{X}_i$'s. Under $(H_{1,n})$, assuming that n tends to infinity with $n_1/n \to t_1$ and under appropriate moment conditions, it is proved in Bickel (1965) that $H_n(n_1) \xrightarrow{d} \chi_K^2(d_H(\delta))$, where

$$d_H(\delta) = t_1(1-t_1)\delta'C^{-1}\delta \,, \tag{12}$$

$\chi_K^2(d)$ denoting the non-central chi-squared distribution with K degrees of freedom and non-centrality parameter d and C denoting the covariance matrix of the $\mathbf{X}_i$'s.

Let $F(\mathbf{X_i}) = (F_1(X_{i,1}),\ldots,F_K(X_{i,K}))'$ denote the K-dimensional vector of marginal distribution functions and $\nabla \log f(\mathbf{X}_i)$ the score function. The following theorem (proved in Appendix C) establishes the asymptotic behaviour of $S_n(n_1)$ under $(H_{1,n})$.

Theorem 3. *Assume that $\mathbf{X}_1,\ldots,\mathbf{X}_{n_1},\mathbf{X}_{n_1+1},\ldots,\mathbf{X}_n$ are $\mathbb{R}^K$-valued random vectors distributed under $(H_{1,n})$ and that the $K \times K$ covariance matrix Σ defined by (5) is positive definite. Assume also that the Fisher information matrix $I_f = \mathbb{E}_f[\nabla \log f(\mathbf{X}_1)\nabla \log f(\mathbf{X}_1)']$ is finite and that the densities f_k of $X_{1,k}$ are upper bounded for all k. Then, as n tends to infinity with $n_1/n \to t_1 \in (0,1)$,*

$$S_n(n_1) \xrightarrow{d} \chi_K^2\left(4t_1(1-t_1)\delta'A'\Sigma^{-1}A\delta\right) =: \chi_K^2(d_S(\delta)) \,, \tag{13}$$

where Σ is defined in (5) and $A = \mathbb{E}_f[(F(\mathbf{X}_1) - 1/2)\nabla \log f(\mathbf{X}_1)']$.

The following corollary, which is proved in Appendix D, particularises this results to the case where the coordinates of the observations are independent.

Corollary 1. *Assume that the density f may be written as the product of its marginals, $f(\mathbf{x}) = \prod_{k=1}^{K} f_k(x_k)$. Then,*

$$d_H(\delta) = t_1(1-t_1) \sum_{k=1}^{K} \frac{\delta_k^2}{\sigma_k^2} \ \text{and} \ d_S(\delta) = 12t_1(1-t_1) \sum_{k=1}^{K} \frac{\delta_k^2}{\sigma_k^2} \lambda_k^2 \,,$$

where $\sigma_k^2 = \int x^2 f_k(x) dx$ and $\lambda_k = \int (\sigma_k f_k(\sigma_k x))^2 dx$.

Moreover, $d_S(\delta) \geq (108/125) d_H(\delta)$, with $d_S(\delta)$ being equal to $(3/\pi) d_H(\delta)$ when f_k are Gaussian densities.

Corollary 1 states that the so-called *a*symptotic relative efficiency (ARE) with respect to the Hotelling-T^2 test $d_S(\delta)/d_H(\delta)$ is lower bounded by $108/125 \approx 0.864$. Granted that this result holds irrespectively of the choice of the marginals f_k this is a strong guarantee that extends the well-known scalar result. As will be illustrated in Section 4.2.2, the ARE is much larger whenever some of the marginal have strong tails. In the particular case of Gaussian densities, the ARE of the proposed test is equal to $3/\pi \approx 0.955$, which is comparable to the values obtained for affine invariant multivariate rank tests (Oja, 1999, p. 331).

For the particular case of multivariate Gaussian distributions, an explicit expression of the ARE can be obtained without assuming independence, as shown by the following corollary (proved in Appendix E).

Corollary 2. *Assume that f is a multivariate Gaussian p.d.f. with mean 0 and covariance matrix C, then the matrices A and Σ featured in Theorem 3 are given by $A = (2\sqrt{\pi})^{-1} \text{diag}(\sigma_1^{-1}, \ldots, \sigma_K^{-1})$ and $\Sigma_{k\ell} = 2/\pi \arcsin(C_{k\ell}/(2\sigma_k \sigma_\ell))$, for $1 \leq k, \ell \leq K$, where $(\sigma_k^2)_{1 \leq k \leq K}$ denote the diagonal elements of C.*

Furthermore, the asymptotic relative efficiency of the test statistic $S_n(n_1)$ can be lower bounded as follows

$$d_S(\delta)/d_H(\delta) \geq (3/\pi)(\sigma_{\min}^2/\sigma_{\max}^2)(\lambda_{\min}(C)/\lambda_{\max}(|C|)) \,,$$

where $\sigma_{\min}^2$ and $\sigma_{\max}^2$ denote, respectively, the minimal and maximal diagonal terms of C, $\lambda_{min}(C)$ is the minimal eigenvalue of C and $\lambda_{max}(|C|)$ the maximal eigenvalue of $|C| = (|C_{k,\ell}|)_{1 \leq k, \ell \leq K}$.

We observe here an important difference with the test statistic of Möttönen et al. (1997) which is designed so as to guarantee that its ARE in the multivariate Gaussian case does not depend on the value of C. For the proposed statistic, the ARE is lower bounded by the minimal eigenvalue of $A'\Sigma^{-1}AC$, where A and Σ are defined in Corollary 2. Recall that Σ is proportional to the Spearman correlation matrix whereas Corollary 2 implies that ACA' is proportional to the standard correlation matrix. Empirically, it can be checked using numerical simulation that the minimal eigenvalue of $A'\Sigma^{-1}AC$ can only be small when C itself is poorly conditioned. The second statement of Corollary 2 substantiates this claim by providing a lower bound which, for positively correlated C at least, is inversely proportional to the condition number of C. Note that this bound appears to be pessimistic in practice as for values of K in the range 2 to 6, we observed the ARE to be larger than 0.8 for all matrices with condition number smaller than 10 (recall that the ARE is equal to 0.95 in the scalar case); for matrices with condition number up to 100, the ARE was still larger than 0.3. Hence, in situations where the coordinates of the data are not too correlated, despite the fact that the proposed test is not affine invariant, its loss with respect to the (optimal) likelihood

ratio test in the Gaussian case is usually negligible, a fact that will be illustrated by the numerical simulations of Section 4.

3. Change-point estimation and detection

We now consider the setting in which the position of the potential change-points are unknown (still assuming that their number is known). Our proposal is to consider the statistic $T(n_1, \ldots, n_{L-1})$ described in the previous section and to optimise it over all the possible change point locations. This proposal is however faced with two serious difficulties. The first one, which is of computational nature is related to the feasibility of the maximisation when there is more than a single change-point. We start by showing that the maximisation of $T(n_1, \ldots, n_{L-1})$ is amenable to dynamic programming and stays feasible even when L is large. The second difficulty, to which a partial answer is provided in Section 3.2, is statistical and concerns the interpretation of the value of $T(n_1, \ldots, n_{L-1})$ as optimising with respect to the change-point location obviously modifies the distribution of the values of the test statistic. This is a difficult issue in general, but we show how to obtain meaningful and simple-to-compute p-values for a variant of the test in the case of a single change-point.

3.1. Multiple change-point estimation

Assuming a known number of change-points L, we propose to use the test statistic described in Section 2.2 to determine the positions of the segment boundaries $n_1, \ldots, n_{L-1}$. These unknown change-point locations are estimated by maximising the statistic $T(n_1, \ldots, n_{L-1})$ defined in (9) with respect to $n_1, \ldots, n_{L-1}$:

$$(\hat{n}_1, \ldots, \hat{n}_{L-1}) = \underset{1 \leq n_1 < \cdots < n_{L-1} \leq n}{\operatorname{argmax}} T(n_1, \ldots, n_{L-1}) . \tag{14}$$

In practice, direct maximisation by enumeration in (14) is computationally prohibitive as it corresponds to a combinatorial task whose complexity grows exponentially with L. However, due to the fact that the matrix $\hat{\Sigma}_n$ is common to all segments, the statistic $T(n_1, \ldots, n_{L-1})$ defined in (9) has an additive structure which makes it possible to adopt a dynamic programming strategy. We refer here to the classical dynamic programming approach to the segmentation task which is described in Kay (1993) used by, among others, Bai and Perron (2003) and can be traced back to the note by Bellman (1961). More precisely, using the notations

$$\Delta(n_\ell + 1 : n_{\ell+1}) = (n_{\ell+1} - n_\ell)\bar{\mathbf{R}}'_\ell \, \hat{\Sigma}_n^{-1} \, \bar{\mathbf{R}}_\ell \, ,$$

and

$$I_L(p) = \max_{1 < n_1 < \cdots < n_{L-1} < n_L = p} \sum_{\ell=0}^{L-1} \Delta(n_\ell + 1 : n_{\ell+1}) , \tag{15}$$

we prove in Appendix F that

$$I_L(p) = \max_{n_{L-1}} \left\{ I_{L-1}(n_{L-1}) + \Delta(n_{L-1} + 1 : p) \right\} . \tag{16}$$

Thus, for solving the optimisation problem (14), we proceed as follows. We start by computing the $\Delta(i:j)$ for all (i,j) such that $1 \leq i < j \leq n$. All the $I_1(E)$ are thus available for $E = 2,\ldots,n$. Then $I_2(E)$ is computed by using the recursion (16) and so on. The overall numerical complexity of the procedure is thus proportional to $L \times n^2$ only.

3.2. Assessing the significance of the test in the single change-point case

In addition to practical algorithms for estimating change-point locations, one needs tools to assess the plausibility of the obtained change-point configuration. An important step in that direction is to characterise the behaviour of $T(\hat{n}_1,\ldots,\hat{n}_{L-1})$ under the null hypothesis that the data are indeed fully homogeneous. This is a difficult issue in general due to the optimisation over all possible change-point configurations. A possible calibration approach consists in running Monte Carlo experiments, possibly using bootstrap techniques if a representative sample of the baseline data of interest is available. We show below that in the case where $L = 2$, *i.e.*, when looking for a single potential change-point, it is possible to obtain a simple computable approximation to the asymptotic p-value of the test.

To do so, we consider in the rest of this section a modification of the test statistic used in (14). The practical consequences of using this variant rather than the statistic $T(\hat{n}_1)$ when $L = 2$ will be discussed after Theorem 4 which states the main result of this section.

Let $\mathbf{V}_n(n_1) = (V_{n,1}(n_1),\ldots,V_{n,K}(n_1))'$ denote the vector such that

$$V_{n,k}(n_1) = \frac{1}{n^{3/2}} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} \left\{ \mathbf{1}(X_{i,k} \leq X_{j,k}) - \mathbf{1}(X_{j,k} \leq X_{i,k}) \right\}, \; k = 1,\ldots,K, \qquad (17)$$

and define

$$\tilde{S}_n(n_1) = \mathbf{V}_n(n_1)' \hat{\Sigma}_n^{-1} \mathbf{V}_n(n_1). \qquad (18)$$

Note that $\mathbf{V}_n$ only differs from $\mathbf{U}_n$ by the normalisation, which is now independent of n_1. We now consider the statistic

$$W_n = \max_{1 \leq n_1 \leq n-1} \tilde{S}_n(n_1). \qquad (19)$$

The following theorem, proved in Appendix G, gives the asymptotic p-values of W_n under the null hypothesis that no change in distribution occurs within the observation data.

Theorem 4. *Assume that* $(\mathbf{X}_i)_{1 \leq i \leq n}$ *are* $\mathbb{R}^K$*-valued i.i.d. random vectors such that, for all k, the c.d.f.* F_k *of* $X_{1,k}$ *is a continuous function. Further assume that the* $K \times K$ *matrix* Σ *defined in (5) is invertible. Then,*

$$W_n \xrightarrow{d} \sup_{0 < t < 1} \left(\sum_{k=1}^{K} B_k^2(t) \right), \; as \; n \to \infty, \qquad (20)$$

where d denotes convergence in distribution and $\{B_k(t), t \in (0,1)\}_{1 \leq k \leq K}$ *are independent Brownian bridges.*

To determine the p-value $P_{\text{val}}(W_n)$ associated to (20), one can use the following result due to Kiefer (1959):

$$P_{\text{val}}(b) = \mathbb{P}\left(\sup_{0<t<1}\left(\sum_{k=1}^{K} B_k^2(t)\right) > b\right)$$

$$= 1 - \frac{4}{\Gamma(\frac{K}{2})2^{\frac{K}{2}}b^{\frac{K}{2}}} \sum_{m=1}^{\infty} \frac{(\gamma_{(K-2)/2,m})^{K-2}\exp[-(\gamma_{(K-2)/2,m})^2]/2b}{[J_{K/2}(\gamma_{(K-2)/2,m})]^2}, \quad (21)$$

where J_ν is the Bessel function of the first kind, $\gamma_{\nu,m}$ is the m-th nonnegative zero of J_ν and Γ is the Gamma function. In practice, only a few terms of the series have to be computed. For values of K of forty or less computing the p-values from the thirty first terms of the series was sufficient.

As noted at the beginning of Section 3.2, the normalisation of $\mathbf{V}_n$ differs from that of $\mathbf{U}_n$, resulting in a statistic W_n that does not coincide with $T(\hat{n}_1)$. From our practical experience, replacing $\mathbf{V}_n$ by $\mathbf{U}_n$ in the definition of W_n, that is using $T(\hat{n}_1)$ instead of W_n, produces a statistic that has the same detection and localisation capacities when the potential change occurs in the central region of the observation window, say between $n/4$ and $3n/4$ observations. For potential changes occurring closer to the beginning or to the end of the observation window, $T(\hat{n}_1)$ has an enhanced detection power at the expense of a slight increase in the rate of false alarms, with corresponding spurious detections occurring mostly near the borders of the observation window. Proceeding as in Appendix G, one can prove a result related to Theorem 4 for $T(\hat{n}_1)$ (used when $L = 2$) by imposing some additional conditions on the admissible values of n_1 (namely, that the maximum is searched only for value of n_1 such that n_1/n is bounded from above and below). The resulting limit, expressed in terms of Bessel processes, does not yield easily computable asymptotic p-values, although approximations such as those studied by Estrella (2003) could be used for approximating extreme quantiles (that is, very low p-values).

4. Numerical Experiments

In this section, we report numerical experiments that illustrate different aspects of the methods proposed in Sections 2 and 3. A software implementing these methods in Python is available as a supplementary material of the paper. For easy reference, the two- or multi-sample homogeneity test defined by (9) is referred to as *MultiRank-H* in the following; the change point estimation criterion defined in (14) is referred to as *MultiRank*.

4.1. Illustration of the two-sample homogeneity test

We start by considering the basic two-sample homogeneity test first introduced in Section 2.1. For this, we generate baseline observations distributed as a mixture of two two-dimensional Gaussian densities with common mean $(0,0)$ and diagonal covariance matrices with diagonal terms equal to $(4, 0.2)$ and $(0.2, 4)$, respectively. For the alternative distribution, we generate observations having the same characteristics except that the mean is now equal to $s = (0.5, 0.5)$. In this case, $n = 100$ and the two groups (baseline and alternative data) are of length $n/2 = 50$. Figure 2 (a) shows a typical example of the data, represented as a two-dimensional scatter plot.

MultiRank-H is compared with three other approaches. The first is the the Maximum Mean Discrepancy (MMD) statistics proposed by Gretton et al. (2006), which is a kernel-based test here used with a Gaussian kernel having a bandwidth given by the median distance between the samples as suggested by Gretton et al. (2006), Désobry et al. (2005) and Harchaoui et al. (2008). The second approach is the classical Hotelling's T^2-test (Chen and Gupta, 2000, p. 67) which is optimal in the multivariate Gaussian case. The third method is to use the likelihood ratio (LR) test assuming a known structure for the model whose parameters (mean vectors and diagonal terms of the covariance matrices) are estimated using the Expectation-Maximisation algorithm. This latter approach is optimal in this context but is the only one that uses some knowledge about the distribution of the data. These methods are compared through their ROC (Receiver Operating Characteristic) curves, averaged over 1000 Monte Carlo replications of the data.

From the results of Section 2.3, we know that, as the covariance matrix of the data is proportional to the identity matrix, the asymptotic performance of Hotelling's T^2-test does not depend on the particular choice of the shift s, but only on its L^2 norm. For the Multirank-H approach the answer is less straightforward as the coordinates of the data are not independent in this example. However, Monte Carlo estimation of the asymptotic performance index $d_s(\delta)$ in (13) shows that it does not varies by more than 1% with the direction of δ. We indeed verified that in the setting of Figure 2 the variation in performance with the direction of the shift s is negligible.

The results displayed in Figure 2(b) show that MultiRank-H is on a par with the LR and outperforms the other two approaches. MMD performs somewhat better than Hotelling's T^2 in this context due to the non-Gaussian nature of the data. Figure 2(c) corresponds to the more difficult setup in which eight-dimensional i.i.d. Gaussian random vectors of variance 2.5^2 are appended to the data described previously. In this case, the data are thus ten-dimensional but the change only affects two coordinates. MultiRank-H is again comparable to the LR, which is still optimal in this case. Note the lack of robustness of MMD which is distinctly dominated by Hotelling's T^2 in this second scenario.

4.2. Properties of the change-point detection test

In this section, we investigate the properties of the change-point detection test based on the statistic $T(\hat{n}_1)$ resulting from the optimisation of (19). The simulations reported in this section are based on the following common benchmark scenario: under (H_0), that is, in the absence of change, we generate 100 samples from a five-dimensional standard Gaussian distribution. Under (H_1), the observations are similar, except that the common mean of the Gaussian vector changes to 0.3 at a location which is either equal to 1/4 or 1/2 of the observation window, that is, at indices 25 or 50. This scenario corresponds to a simple situation where all coordinates of the data possibly undergo similar upward shifts. The ROC curves that are plotted in the following are based on 2000 replications of the simulated data.

4.2.1. Comparison with marginal decisions

The MultiRank test statistic is obviously based on a combination of marginal rank statistics. Nevertheless, it incorporates two important aspects of the multivariate change-point detection problem: first, detection of simultaneous changes in multiple coordinates should make the presence of an
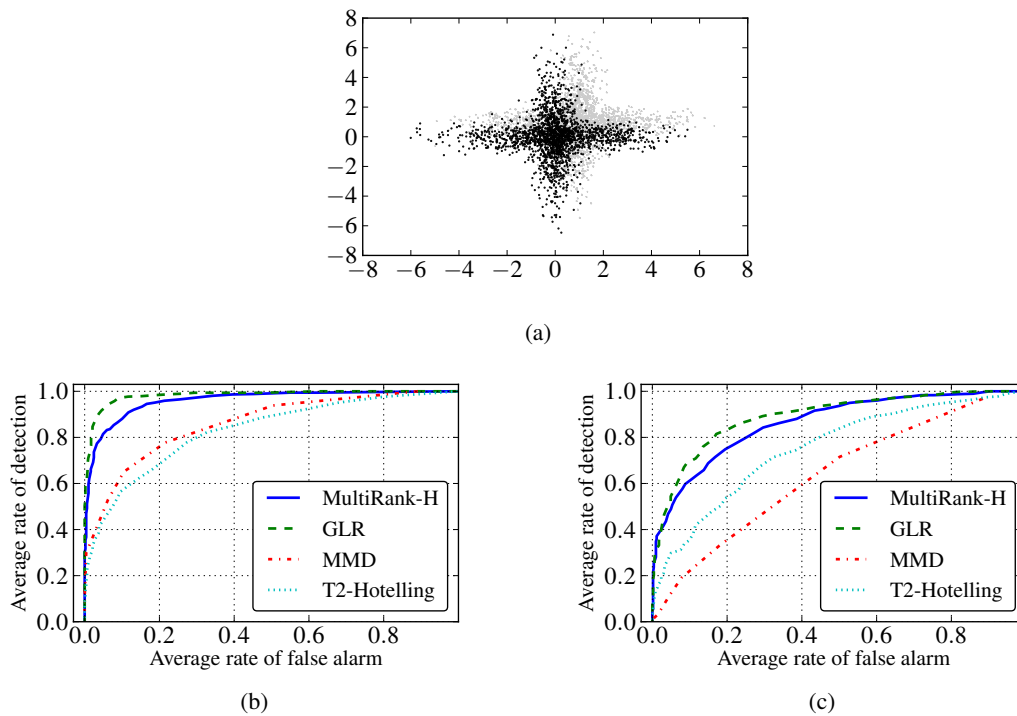
(a)

(b)

(c)

FIGURE 2. *(a) Example of observations under baseline and alternative distributions, (b) ROC curves for MultiRank-H, MMD, Hotelling's T^2 and LR, (c) Same as (b) with eight-dimensional Gaussian noise padding (that is, with $K = 10$).*

actual change-point more likely, and, second, the existence of dependence between the coordinates should influence the decision. To illustrate these observations, we compare MultiRank with a simpler heuristic approach that combines marginal decisions based on Bonferroni bound, using as test statistic $\max_{1 \leq k \leq K} \max_{1 \leq n_1 \leq n-1} V_{n,k}(n_1)$. The results obtained with the data-generating mechanism described at the beginning of Section 4.2 are displayed in the leftmost plot of Figure 3. We also compare both approaches in a setting where the covariance matrix of the Gaussian vector is not the identity matrix anymore but a tridiagonal matrix with a common value of 0.45 (positive correlation) or −0.45 (negative correlation) on the sub- and super-diagonal. The resulting ROC curves are displayed in the middle and right plots of Figure 3, respectively.

The leftmost plot of Figure 3 shows that the approach that combines the marginal statistics by taking into account their correlation, that is MultiRank, outperforms the Bonferroni-type approach. Furthermore, when the coordinates are positively correlated, the rate of detection of the MultiRank method decreases for a given false alarm rate and when negatively correlated, the rate of detection increases. The performance of the Bonferroni-type approach on the other hand does not improve for the negatively correlated data. The MultiRank method captures an important feature of the problem that fails to be exploited by the mere marginal decisions: negative correlations in the data make the detection of simultaneous upward jumps easier while positive correlations render this task more difficult.

Although Figure 3 deals with change-point estimation (which includes the maximisation with

respect to the change point position), we note that Corollary 2 of Section 2.3 implies that the performance of the Multirank-H homogeneity test is here comparable to that of the Hotelling's T^2 test as the condition number of the covariance matrices considered in Figures 3 is relatively small (less than 8.1). The Multirank change detection test obviously inherits this property as its performance is nearly indiscernible from that of the LR test for all three choices of the covariance matrix in Figure 3.
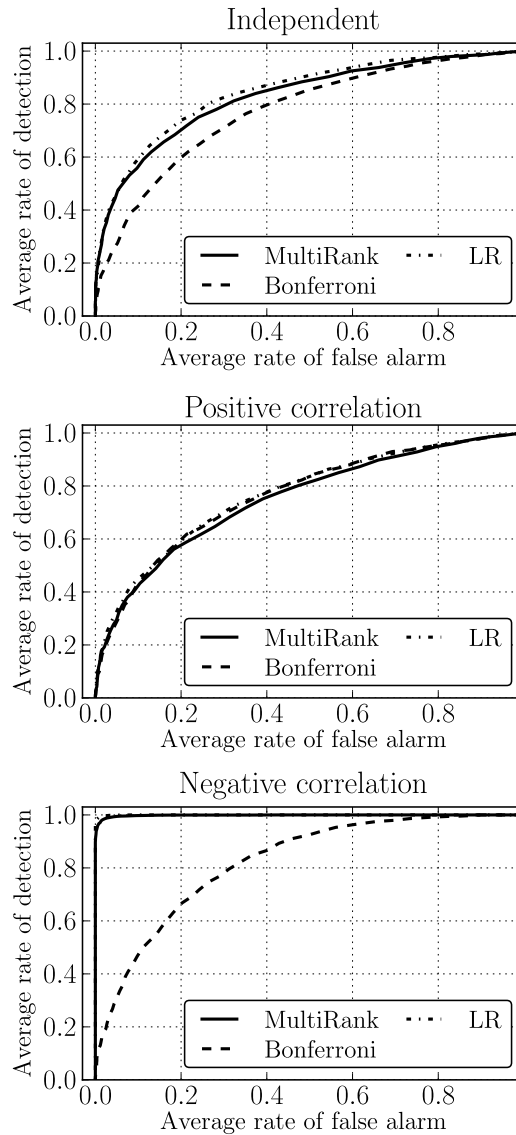

FIGURE 3. *ROC curves for the MultiRank and the Bonferroni-type approaches when the coordinates are independent (left), positively correlated (middle) and negatively correlated (right). The change-point instant is located at 1/4 and 1/2 of the observations window of length 500.*
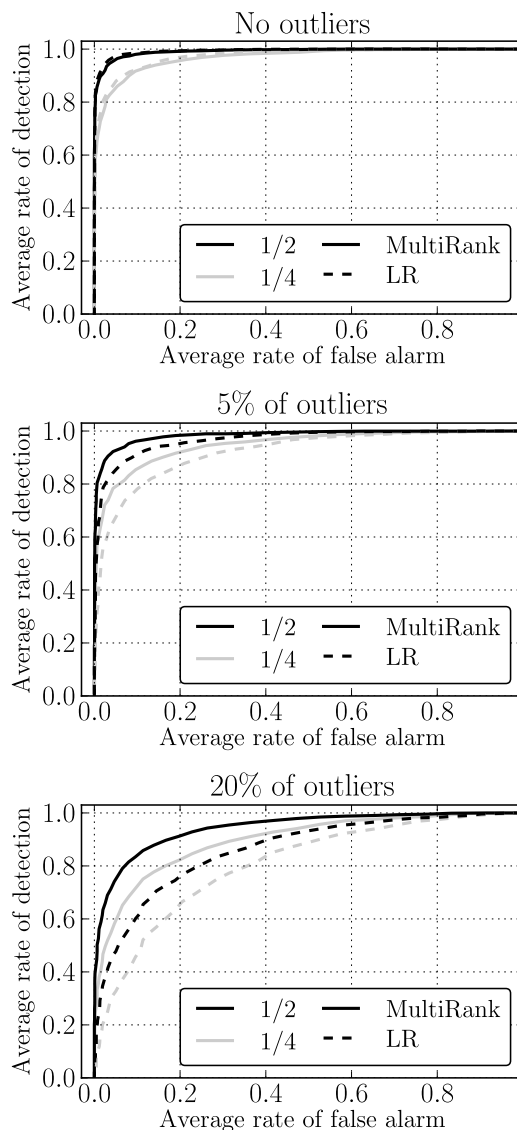
FIGURE 4. *ROC curves for the MultiRank approach and the likelihood-ratio procedure (LR) for three different proportions of outliers (from left to right: 0, 5 and 20%) when the change-point instant is located at 1/4 and 1/2 of the window of length 500.*

4.2.2. Robustness with respect to outliers

Here we illustrate the robustness of the MultiRank approach with respect to outliers in the data by considering the same simulation scenario as in the previous section (with a larger shift of amplitude 0.5) progressively contaminated by large additive outliers. The outliers distribution is the multivariate Gaussian distribution with covariance matrix $10\,\mathrm{Id}_5$, instead of Id_5 for the baseline distribution (where Id_5 refers to the five by five identity matrix). The fraction of outliers is varied

between 0, 5 and 20%. The MultiRank approach is compared to the parametric likelihood-ratio based change-point detection test described by Srivastava and Worsley (1986). This latter method, which is itself based on Hotelling's T^2-test statistic, is optimal in the absence of outliers as the baseline and alternative distributions are both Gaussian. As shown in the leftmost plot of Figure 4, MultiRank has comparable performance with the parametric approach in the case where there are no outliers in the data. However, as shown in the middle and rightmost plots of Figure 4, MultiRank demonstrates its robustness with respect to the presence of outliers as it barely suffers from the presence of additive outliers contrary to the parametric approach.

4.3. Application to genomic hybridisation data

To illustrate the potential of the approach, we consider its application to the segmentation of multiple individual genomic data. We consider the bladder cancer micro-array aCGH dataset studied by Vert and Bleakley (2010) which consists of records of copy-number variations, *i.e.* abnormal alteration of the quantity of DNA sections.

The objective here is to jointly segment data recorded from different subjects so as to robustly detect regions of frequent deletions or additions of DNA which could be characteristic of cancer. Each of the 57 profiles provides the relative quantity of DNA for 2143 probes measured on 22 chromosomes. We ran the change-point estimation algorithm on each of 22 chromosomes separately, thus processing 22 different 9- to 57-dimensional signals (depending on the selected groups of patients at different stages of cancer) of length 50 to 200 (the number of probes varies for each chromosome).

In this paper, we have not considered principled methods for inferring the number of change-points from the data. We describe below an heuristic approach to determine the number of change-points which, despite its simplicity, performs in our experience much better than the use of generic penalties such as AIC or BIC. Values of the statistics $I_L(n)$, for $L = 0, \ldots, L_{\max}$, are first computed using the procedure described in Section 3.1. The algorithm is based on the principle that in the presence of $L^\star \geq 1$ change-points, if $I_L(n)$ is plotted against L, the resulting graph can be decomposed into two distinct regions: the first one, for $L = 0, \ldots, L^\star$ where the criterion is growing rapidly; and the second one, for $L = L^\star, \ldots, L_{\max}$, where the criterion is barely increasing Lavielle (2005). Hence, for each possible value of L in $L = 1, \ldots, L_{\max}$, we compute least square linear regressions for both parts of the graph (before and after L); the estimated number of change-points is the value of L that yields the best fit, that is, the value for which the sum of the residual sums of squares computed on both parts of the graph is minimal. For an illustration of this methodology, see Figure 5. The case $L = 0$ is treated separately and the procedure described above is used only when the value of the test statistic W_n for the presence of a single change-point (see Section 3.2) is significant (*p*-value smaller than 0.1%) based on Theorem 4.

Results are shown for a group of 32 profiles corresponding to Stage T2 of a tumour. In Figure 6 the copy-number data and the segmentation of the whole set of chromosomes is displayed for two particular individuals, together with the corresponding stepwise constant approximations of the data for the 32 individuals (in the bottom of the Figure). Figure 7 details the results pertaining to the 7th chromosome. In both cases, the segmentation result is represented by a signal which is constant (and equal to the mean of the data) within the detected segments. The bottom plot
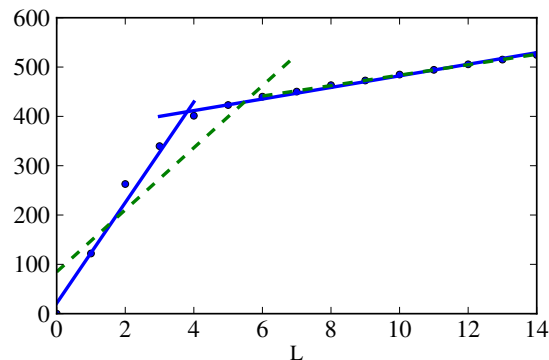
FIGURE 5. *Determining the optimal number of change-points. Here, the actual number of change-points is $L^\star = 4$; the optimal regression is displayed in solid lines, while a non-optimal alternative (for $L = 6$) is displayed in dashed lines.*

in Figure 6 particularly highlights the fact that several coordinates indeed jump at the same time, suggesting that the joint segmentation model is appropriate. On the other hand, it is also obvious that one cannot assume (see, e.g., the third change-point at index 64 in Figure 7) that all coordinates undergo similar changes. Note also that in this application, the fact that the MultiRank test statistic is properly normalised with respect to the length n of the data and their dimension K is particularly important: n corresponds to the number of probes and varies with each chromosome, K represents the number of individuals and varies when considering different groups of subjects.

As a reference, the group fused Lasso algorithm by Bleakley and Vert (2011) outputs similar results. In particular, on the 7th chromosome, change-points are found at positions 21, 44, 65, 102, 107, 112, 124, 132, 156 and 166. On the whole set of chromosomes, 96 change-points are found while the MultiRank estimation procedure outputs 98.

5. Conclusion

We proposed an approach for retrospective detection of multiple changes in multivariate data. The basic idea, used for homogeneity testing when the data groupings are known, is an extension of well-known marginal rank based tests (Wilcoxon/Mann-Whitney and Kruskal-Wallis) based on the idea originally proposed by Wei and Lachin (1984). The use of this approach for change-point detection (when the segments boundaries are unknown) was shown to be computationally feasible. In addition, it incorporates important aspects of the problem, in particular the fact that simultaneous detections in different coordinates make the presence of an actual change more likely. The method was shown to be robust against various alternatives and on a par with optimal methods in benchmark cases. The approach can also be straightforwardly modified to deal with ordinal data, missing or censored values.

To improve the method, it would be desirable to provide significance levels for the change-point detection test when used to detect more than a single potential change-point. We believe that by considering the normalisation used to define the statistic W_n in Section 3.2, it is possible to study the asymptotic behaviour of the change-point detection statistic in the general case. This
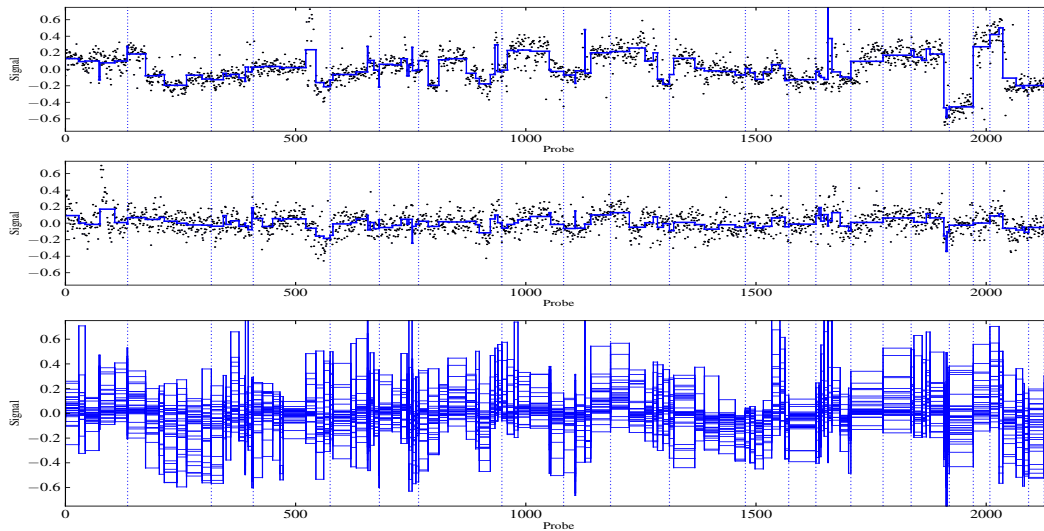
FIGURE 6. *First and second row: copy number data for two different individuals with superimposition of the segmentation. Third row: superimposition of the smoothed bladder tumour aCGH data for 32 individuals in Stage T2 cancer that result from the segmentation. Vertical dashed lines represent the separation between the different chromosomes.*

being said, the difference between the two forms of normalisation could be more significant when applied to more than two segments. On a different level, one should obviously consider more principled approaches for selecting the number of change-points. General-purpose penalisation schemes could be used but we feel that novel ideas need to be developed specifically for the change-point problem given the specific nature of over- and under-estimating the number of change-points. For instance, in many practical applications the significance of over-estimating the number of change-points depends not only on the number of spurious segments but also on their locations. Traditional approaches based on complexity penalties Bai and Perron (2003), Bayesian methods Fearnhead (2006) and sparsity-based criterions Harchaoui and Lévy-Leduc (2010); Vert and Bleakley (2010) are already available but there is certainly room for new developments in these fields.

Appendix A: Proof of Theorem 1

The proof is based on the Hoeffding decomposition of $U_{n,k}(n_1)$ for each k in $\{1,\ldots,K\}$. For further details on the Hoeffding decomposition, we refer the reader to Chapters 11 and 12 of van der Vaart (1998). For each k in $\{1,\ldots,K\}$, let $h_{1,k}(y) = \int h(x,y)dF_k(x)$ and $\tilde{h}_{1,k}(x) = \int h(x,y)dF_k(y)$, where h is defined by $h(x,y) = \mathbf{1}(x \leq y) - \mathbf{1}(y \leq x)$. By the continuity of F_k, $h_{1,k}(y) = 2F_k(y) - 1$ and $\tilde{h}_{1,k}(x) = 1 - 2F_k(x)$. The Hoeffding decomposition of $U_{n,k}(n_1)$ can thus be written as $U_{n,k}(n_1) =$
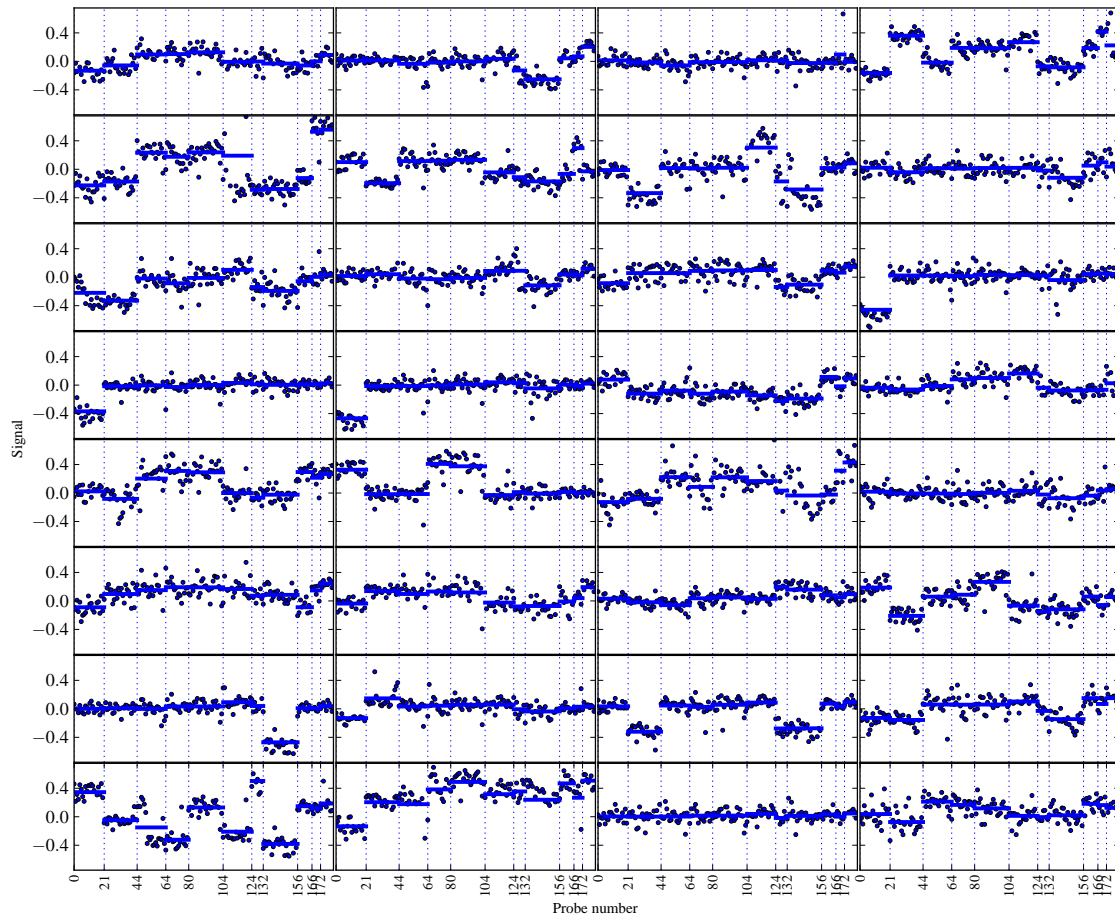
FIGURE 7. *Data for 32 individuals in Stage T2 bladder cancer with superimposed segmentation for chromosome 7. 10 change-points were estimated and the dashed vertical lines correspond to the estimated segment boundaries.*

$\hat{U}_{n,k}(n_1) + R_{n,k}(n_1)$, where

$$\hat{U}_{n,k}(n_1) = \frac{n_1}{\sqrt{nn_1(n-n_1)}} \sum_{j=n_1+1}^{n} h_{1,k}(X_{j,k}) + \frac{n-n_1}{\sqrt{nn_1(n-n_1)}} \sum_{i=1}^{n_1} \tilde{h}_{1,k}(X_{i,k}) , \qquad (22)$$

$$R_{n,k}(n_1) = \frac{1}{\sqrt{nn_1(n-n_1)}} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} [h(X_{i,k},X_{j,k}) - \tilde{h}_{1,k}(X_{i,k}) - h_{1,k}(X_{j,k})] . \qquad (23)$$

We first prove that $U_{n,k}(n_1) = \hat{U}_{n,k}(n_1) + o_p(1)$ by showing that $\mathrm{Var}[R_{n,k}(n_1)] \to 0$, as n tends to infinity. Using that $\mathbb{E}[U_{n,k}(n_1)] = \mathbb{E}[\hat{U}_{n,k}(n_1)] = 0$, we obtain that $\mathrm{Var}[R_{n,k}(n_1)] = \mathrm{Var}[U_{n,k}(n_1) - \hat{U}_{n,k}(n_1)] = \mathbb{E}[U_{n,k}^2(n_1)] + \mathbb{E}[\hat{U}_{n,k}^2(n_1)] - 2\mathbb{E}[U_{n,k}(n_1)\hat{U}_{n,k}(n_1)]$. By independence of the $(X_{i,k})_{1 \le i \le n}$, we obtain that

$$\mathbb{E}[\hat{U}_{n,k}^2(n_1)] \quad = \quad \frac{n_1^2}{nn_1(n-n_1)} \sum_{j=n_1+1}^{n} \mathbb{E}[h_{1,k}(X_{j,k})^2] \quad + \quad \frac{(n-n_1)^2}{nn_1(n-n_1)} \sum_{i=1}^{n_1} \mathbb{E}[\tilde{h}_{1,k}(X_{i,k})^2]. \quad (24)$$

Using that

$$\mathbb{E}[h_{1,k}(X_{i,k})^2] = 4\mathbb{E}[(F_k(X_{1,k}) - 1/2)^2] = 4\,\mathrm{Var}(\mathscr{U}) = 1/3 \,, \tag{25}$$

where $\mathscr{U}$ has a uniform distribution on $[0,1]$, we get, on the one hand, that

$$\mathbb{E}[\hat{U}_{n,k}^2(n_1)] = \frac{n_1^2(n-n_1)}{3nn_1(n-n_1)} + \frac{(n-n_1)^2 n_1}{3nn_1(n-n_1)} = 1/3 \,. \tag{26}$$

On the other hand

$$\mathbb{E}[U_{n,k}^2(n_1)] = \frac{1}{nn_1(n-n_1)} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} \mathbb{E}[h(X_{i,k}, X_{j,k})^2]$$

$$+ \frac{1}{nn_1(n-n_1)} \sum_{1 \le i \ne i' \le n_1} \sum_{j=n_1+1}^{n} \mathbb{E}[h(X_{i,k}, X_{j,k})h(X_{i',k}, X_{j,k})]$$

$$+ \frac{1}{nn_1(n-n_1)} \sum_{i=1}^{n_1} \sum_{n_1+1 \le j \ne j' \le n} \mathbb{E}[h(X_{i,k}, X_{j,k})h(X_{i,k}, X_{j',k})] \,. \tag{27}$$

We separately study the three terms of the r.h.s of (27). Using that $(X_{i,k})_{1 \le i \le n}$ are i.i.d. , we get

$$\frac{1}{nn_1(n-n_1)} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} \mathbb{E}[h(X_{i,k}, X_{j,k})^2] = \frac{n_1(n-n_1)}{nn_1(n-n_1)} \mathbb{E}[h(X_{1,k}, X_{n_1+1,k})^2] \to 0, \text{ as } n \to \infty \,. \tag{28}$$

Then, by continuity of F_k, we have

$$\frac{1}{nn_1(n-n_1)} \sum_{1 \le i \ne i' \le n_1} \sum_{j=n_1+1}^{n} \mathbb{E}[h(X_{i,k}, X_{j,k})h(X_{i',k}, X_{j,k})]$$

$$= \frac{(n_1^2 - n_1)(n-n_1)}{nn_1(n-n_1)} \int (2F_k(y)-1)(2F_k(y)-1)dF_k(y) = \frac{n_1(n_1-1)(n-n_1)}{3nn_1(n-n_1)} \,. \tag{29}$$

Using similar arguments, the last term of the r.h.s of (27) is equal to $n_1(n-n_1)(n-n_1-1)/(3nn_1(n-n_1))$. With (28) and (29), we obtain

$$\mathbb{E}[U_{n,k}^2(n_1)] \to 1/3, \text{ as } n \to \infty. \tag{30}$$

Since $\mathbb{E}[U_{n,k}(n_1)\hat{U}_{n,k}(n_1)] \to 1/3$, as $n \to \infty$, (26) and (30) lead to $\mathrm{Var}[R_{n,k}(n_1)] \to 0$ and thus $U_{n,k}(n_1) = \hat{U}_{n,k}(n_1) + o_p(1)$, as n tends to infinity. The multivariate central limit theorem then yields $(U_{n,1}(n_1), \dots, U_{n,K}(n_1))' \to \mathcal{N}(0, \Sigma)$, where the (k,k')th entry of Σ is given by $\Sigma_{kk'} = \lim_{n \to \infty} \mathbb{E}[\hat{U}_{n,k}(n_1)\hat{U}_{n,k'}(n_1)]$. Using that the $(X_{i,k})_{1 \le i \le n}$ are i.i.d., we obtain that

$$\mathbb{E}[\hat{U}_{n,k}(n_1)\hat{U}_{n,k'}(n_1)] = \frac{4n_1^2}{nn_1(n-n_1)} \sum_{j=n_1+1}^{n} \mathbb{E}[\{F_k(X_{j,k}) - 1/2\}\{F_{k'}(X_{j,k'}) - 1/2\}]$$

$$+ \frac{4(n-n_1)^2}{nn_1(n-n_1)} \sum_{i=1}^{n_1} \mathbb{E}[\{F_k(X_{i,k}) - 1/2\}\{F_{k'}(X_{i,k'}) - 1/2\}]$$

$$= 4\,\mathrm{Cov}\left(F_k(X_{1,k}), F_{k'}(X_{1,k'})\right) \,.$$

Thus, $\Sigma^{-1/2}(U_{n,1}(n_1), \dots, U_{n,K}(n_1))' \xrightarrow{d} \mathcal{N}(0, \mathrm{Id}_K)$. Since $\hat{\Sigma}_n \xrightarrow{p} \Sigma$, we deduce from Slutsky's Theorem that $\hat{\Sigma}_n^{-1/2}(U_{n,1}(n_1), \dots, U_{n,K}(n_1))' \xrightarrow{d} \mathcal{N}(0, \mathrm{Id}_K)$, which concludes the proof.

Appendix B: Proof of Theorem 2

Using that $R_j^{(k)} = \sum_{i=1}^n \mathbf{1}(X_{i,k} \leq X_{j,k})$, we obtain that

$$\bar{R}_\ell^{(k)} - \frac{n+1}{2} = \frac{1}{n_{\ell+1} - n_\ell} \left(\sum_{j=n_\ell+1}^{n_{\ell+1}} \sum_{i=1}^n \mathbf{1}(X_{i,k} \leq X_{j,k}) \right) - \frac{n+1}{2}$$

$$= \frac{1}{n_{\ell+1} - n_\ell} \sum_{j=n_\ell+1}^{n_{\ell+1}} \sum_{\substack{i=1 \\ i \neq j}}^n \left[\mathbf{1}(X_{i,k} \leq X_{j,k}) - 1/2 \right]. \quad (31)$$

Let $h(x,y) = \mathbf{1}(x \leq y)$, $h_{1,k}(y) = \int \mathbf{1}(x \leq y)\mathrm{d}F_k(x)$ and $h_{2,k}(x) = \int \mathbf{1}(x \leq y)\mathrm{d}F_k(y)$. By continuity of F_k: $h_{1,k}(y) = F_k(y)$ and $h_{2,k}(x) = 1 - F_k(x)$. Using the notation:

$$\mathscr{R}_\ell^{(k)} = (n_{\ell+1} - n_\ell)^{1/2}/n(\bar{R}_\ell^{(k)} - (n+1)/2),$$

the Hoeffding decomposition yields

$$\mathscr{R}_\ell^{(k)} = \frac{(n_{\ell+1} - n_\ell)^{1/2}}{n} \left[\frac{n-1}{n_{\ell+1} - n_\ell} \sum_{j=n_\ell+1}^{n_{\ell+1}} (h_{1,k}(X_{j,k}) - 1/2) + \frac{n_{\ell+1} - n_\ell - 1}{n_{\ell+1} - n_\ell} \sum_{i=1}^n (h_{2,k}(X_{i,k}) - 1/2) \right]$$

$$+ \frac{(n_{\ell+1} - n_\ell)^{1/2}}{n} \left[\frac{1}{n_{\ell+1} - n_\ell} \sum_{j=n_\ell+1}^{n_{\ell+1}} \sum_{\substack{i=1 \\ i \neq j}}^n \left\{ h(X_{i,k}, X_{j,k}) - h_{1,k}(X_{j,k}) - h_{2,k}(X_{i,k}) + 1/2 \right\} \right]$$

$$\overset{\text{def}}{=} \mathscr{R}_{\ell,1}^{(k)} + \mathscr{R}_{\ell,2}^{(k)} + \mathscr{R}_{\ell,3}^{(k)}. \quad (32)$$

Note that $\mathscr{R}_{\ell,3}^{(k)} = o_p(1)$, as n tends to infinity, since it can be proved that $\mathrm{Var}(\mathscr{R}_{\ell,3}^{(k)}) = \mathrm{Var}[\mathscr{R}_\ell^{(k)} - (\mathscr{R}_{\ell,1}^{(k)} + \mathscr{R}_{\ell,2}^{(k)})] \to 0$, as n tends to infinity. Thus, (32) can be rewritten as

$$\mathscr{R}_\ell^{(k)} = \frac{n-1}{n(n_{\ell+1} - n_\ell)^{1/2}} \sum_{j=n_\ell+1}^{n_{\ell+1}} (F_k(X_{j,k}) - 1/2) + \frac{n_{\ell+1} - n_\ell - 1}{n(n_{\ell+1} - n_\ell)^{1/2}} \sum_{i=1}^n (1/2 - F_k(X_{i,k})) + o_p(1).$$

Since $\sum_{i=1}^n (1/2 - F_k(X_{i,k})) = \sum_{p=0}^{L-1} \sum_{j=n_p+1}^{n_{p+1}} (1/2 - F_k(X_{j,k}))$,

$$\mathscr{R}_\ell^{(k)} = \frac{n - (n_{\ell+1} - n_\ell)}{n(n_{\ell+1} - n_\ell)^{1/2}} \sum_{j=n_\ell+1}^{n_{\ell+1}} (F_k(X_{j,k}) - 1/2) - \frac{n_{\ell+1} - n_\ell - 1}{n(n_{\ell+1} - n_\ell)^{1/2}} \sum_{\substack{p=0 \\ p \neq \ell}}^{L-1} \sum_{j=n_p+1}^{n_{p+1}} (F_k(X_{j,k} - 1/2)) + o_p(1)$$

$$\overset{\text{def}}{=} U_k(n_\ell, n_{\ell+1}) + o_p(1).$$

Observe that, for a fixed ℓ in $\{0, \ldots, L-1\}$ and k, k' in $\{1, \ldots, K\}$, we get, as n tends to infinity,

$$4\,\mathrm{Cov}(U_k(n_\ell, n_{\ell+1}), U_{k'}(n_\ell, n_{\ell+1})) =$$

$$\Sigma_{kk'} \left[\left(1 - \frac{(n_{\ell+1} - n_\ell)}{n} \right)^2 + \sum_{\substack{p=0 \\ p \neq \ell}}^{L-1} \frac{(n_{\ell+1} - n_\ell - 1)^2 (n_{p+1} - n_p)}{n^2(n_{\ell+1} - n_\ell)} \right] \to (1 - t_{\ell+1})\Sigma_{kk'}, \quad (33)$$

where we have used that $\sum_{\substack{p=0 \\ p\neq\ell}}^{L-1}(n_{p+1}-n_p)=n-(n_{\ell+1}-n_\ell)$. In the same way, for fixed k,k' in $\{1,\dots,K\}$ and $\ell\neq\ell'$ in $\{0,\dots,L-1\}$, we obtain, as n tends to infinity,

$$4\operatorname{Cov}(U_k(n_\ell,n_{\ell+1}),U_{k'}(n_{\ell'},n_{\ell'+1}))\to-\sqrt{t_{\ell+1}t_{\ell'+1}}\Sigma_{kk'}\ . \tag{34}$$

Let

$$\bar{R}_n=2\left(\frac{(n_1-n_0)^{1/2}}{n}\bar{R}_0',\dots,\frac{(n_L-n_{L-1})^{1/2}}{n}\bar{R}_{L-1}'\right)'\ .$$

We deduce from (33), (34) and the multivariate central limit theorem that

$$\bar{R}_n\xrightarrow{d}\mathcal{N}(0,\Theta\otimes\Sigma)\,,n\to\infty\,,$$

where Σ is the $K\times K$ matrix defined in (5), $\otimes$ denotes the Kronecker product, $\Theta=\operatorname{Id}_L-\sqrt{t}\sqrt{t}'$ with $\sqrt{t}=(\sqrt{t_1},\dots,\sqrt{t_L})'$. Thus,

$$\bar{R}_n^\Sigma\xrightarrow{d}\mathcal{N}(0,\Theta\otimes\operatorname{Id}_K)\,,n\to\infty\,,$$

where

$$\bar{R}_n^\Sigma=2\left(\frac{(n_1-n_0)^{1/2}}{n}\Sigma^{-1/2}\bar{R}_0',\dots,\frac{(n_L-n_{L-1})^{1/2}}{n}\Sigma^{-1/2}\bar{R}_{L-1}'\right)'\ .$$

Since $\hat{\Sigma}_n\xrightarrow{p}\Sigma$, as n tends to infinity, the same convergence holds when Σ is replaced by $\hat{\Sigma}_n$. Since $\sum_{\ell=0}^{L-1}(n_{\ell+1}-n_\ell)/n=1$, $\sum_{\ell=1}^L t_\ell=1$ and the matrix t has eigenvalue 0 of multiplicity 1 (with eigenspace spanned by $\sqrt{t}$), and eigenvalue 1 of multiplicity $L-1$. Hence, the eigenvalues of $\Theta\otimes\operatorname{Id}_K$ are 0, with multiplicity K, and 1, with multiplicity $(L-1)K$, which concludes the proof using Cochran's theorem.

Appendix C: Proof of Theorem 3

In Appendix A, we proved that under the null hypothesis where the $\mathbf{X}_j$'s are i.i.d. random vectors such that the c.d.f F_k of $X_{1,k}$ is continuous:

$$U_{n,k}(n_1)=\frac{2n_1}{\sqrt{nn_1(n-n_1)}}\sum_{j=n_1+1}^n\{F_k(X_{j,k})-1/2\}$$
$$-\frac{2(n-n_1)}{\sqrt{nn_1(n-n_1)}}\sum_{i=1}^{n_1}\{F_k(X_{i,k})-1/2\}+o_P(1)\,,\text{ as }n\to\infty\,. \tag{35}$$

Since, by assumption, the model $f(\mathbf{x}-\theta)$ is differentiable in quadratic mean at θ, the log-likelihood ratio L_n defined by

$$L_n=\log\left[\frac{\prod_{i=1}^{n_1}f(\mathbf{X}_i)\prod_{j=n_1+1}^n f(\mathbf{X}_j-\delta/\sqrt{n})}{\prod_{i=1}^n f(\mathbf{X}_i)}\right]$$

satisfies the following asymptotic expansion as n tends to infinity

$$L_n = -\frac{\delta}{\sqrt{n}} \sum_{j=n_1+1}^{n} \nabla \log f(\mathbf{X}_j) - \frac{(1-t_1)}{2} \delta' I_f \delta + o_P(1) . \qquad (36)$$

Combining multivariate central limit theorem with (35) and (36) one can thus show that $(\mathbf{U}_n, L_n)$ converges in distribution to a Gaussian random vector. From the expression of the asymptotic covariance matrix and using Le Cam's third lemma, one obtains that under $(H_{1,n})$

$$\mathbf{U}_n(n_1) \xrightarrow{d} \mathcal{N}_K(2\sqrt{t_1(1-t_1)}A\delta, \Sigma) .$$

Using Glivenko-Cantelli Theorem, the weak law of large numbers and the fact that f_k is bounded for all k –which implies that F_k is a Lipschitz function for all k– it can be proved that, under $(H_{1,n})$, $\hat{\Sigma}_n$ converges in probability to Σ. The conclusion follows using Slutsky's Lemma.

Appendix D: Proof of Corollary 1

Assuming independence, $\Sigma = 4\mathbb{E}_0[\{F_1(X_{1,1}) - 1/2\}^2] \mathrm{Id}_K = 1/3 \mathrm{Id}_K$, where Id_K denotes the $K \times K$ identity matrix, and A is the diagonal matrix with elements $A_{k,k} = \int_{\mathbb{R}} F_k(x) f_k'(x) \mathrm{d}x = -\int_{\mathbb{R}} f_k^2(x) \mathrm{d}x = -\sigma_k^{-1} \int_{\mathbb{R}} (\sigma_k f_k(\sigma_k x))^2 \mathrm{d}x$. The lower bound on d_S is obtained by the classical result that $12\lambda_k^2 \geq 108/125$ (van der Vaart, 1998, p. 198). Finally, in the Gaussian case, $\lambda_k = 1/(2\sqrt{\pi})$.

Appendix E: Proof of Corollary 2

Let us first prove that A is a diagonal matrix such that $A_{k,k} = \sigma_k^{-1}/(2\sqrt{\pi})$. Let $D = (d_{k,\ell})_{1 \leq k, \ell \leq K} = C^{-1}$, then

$$A_{k,\ell} = \mathbb{E}_0[(F_k(X_{1,k}) - 1/2)(\sum_{j=1}^{K} d_{\ell,j} X_{1,j})] = \sum_{j=1}^{K} d_{\ell,j} \sigma_j \mathbb{E}[\Phi(\tilde{X}_{1,k})\tilde{X}_{1,j}] ,$$

where Φ is the c.d.f. of a standard Gaussian random variable and $\tilde{X}_{1,j} = X_{1,j}/\sigma_j$ is a standard Gaussian random variable. Since $\Phi(x) - 1/2 = \sum_{p \geq 1} (\alpha_p/p!) H_p(x)$, where $\alpha_p = \mathbb{E}[\Phi(Z)H_p(Z)]$, Z is a standard Gaussian random variable and H_p is the pth Hermite polynomial with leading coefficient equal to one ($H_1(x) = x$, $H_2(x) = x^2 - 1$, ...). By applying Mehler's formula, one obtains $\mathbb{E}[\Phi(\tilde{X}_{1,k})\tilde{X}_{1,j}] = \alpha_1 C_{k,j}/(\sigma_k \sigma_j)$, where $\alpha_1 = \mathbb{E}[Z\Phi(Z)] = \int_{\mathbb{R}} \varphi^2(x) \mathrm{d}x$, using an integration by parts. Thus, $A_{k,\ell} = \sigma_k^{-1}/(2\sqrt{\pi}) \sum_{j=1}^{K} d_{\ell,j} C_{j,k} = \sigma_k^{-1}/(2\sqrt{\pi}) \mathbf{1}_{\{k=\ell\}}$. The expression given for Σ can be obtained similarly and is well-known in the literature as it provides the link between the Spearman correlation and the usual correlation coefficients Kruskal (1958). Regarding the lower bound, first note that

$$\begin{aligned}
\mathrm{ARE} &= (4\pi)^{-1} \delta' \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_K^{-1})(4\Sigma^{-1})\mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_K^{-1})\delta / \delta' C^{-1} \delta \\
&\geq (4\pi)^{-1} \lambda_{\min}\left(\mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_K^{-1})(4\Sigma^{-1})\mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_K^{-1})C\right) \\
&= (4\pi)^{-1} \sigma_{\max}^{-2} \lambda_{\min}(C)/\lambda_{\max}(\Sigma/4) .
\end{aligned}$$

From the expression of Σ, it is easily checked that $\Sigma_{k,\ell} \leq |C_{k,\ell}/3 \sigma_k^{-1} \sigma_\ell^{-1}| \leq |C_{k,\ell}|/3 \sigma_{\min}^{-2}$, which gives the second result.

Appendix F: Proof of (16)

By (15), $I_1(p) = \max_{1<n_1=p}\Delta(1:n_1) = \Delta(1:p)$ and $I_2(p) = \max_{1<n_1<n_2=p}\{\Delta(1:n_1)+\Delta(n_1+1:p)\} = \max_{1<n_1<n_2=p}\{I_1(n_1)+\Delta(n_1+1:p)\}$, which is (16) when $L=2$. By (15), $I_3(p) = \max_{1<n_1<n_2<n_3=p}\{\Delta(1:n_1)+\Delta(n_1+1:n_2)+\Delta(n_2+1:p)\}$. By using the previous expression of $I_2(p)$, we get that $I_3(p) = \max_{1<n_2<p}\{I_2(n_2)+\Delta(n_2+1:p)\}$, which is (16) when $L=3$ and so on, which gives (16).

Appendix G: Proof of Theorem 4

We start by proving (20) when $\hat{\Sigma}_n$ is replaced by Σ in (18). For this, we shall verify the assumptions of (Billingsley, 1968, Theorem 15.6): the convergence of the finite-dimensional distributions:

$$\big(\mathbf{V}_n(\lfloor nt_1\rfloor)'\Sigma^{-1}\mathbf{V}_n(\lfloor nt_1\rfloor),\ldots,\mathbf{V}_n(\lfloor nt_p\rfloor)'\Sigma^{-1}\mathbf{V}_n(\lfloor nt_p\rfloor)\big)$$
$$\xrightarrow{d} \left(\sum_{k=1}^{K}B_k^2(t_1),\ldots,\sum_{k=1}^{K}B_k^2(t_p)\right),\ \text{for } 0<t_1<\ldots<t_p<1\ ,\ n\to\infty\ ,\quad (37)$$

and the tightness criterion for the process:

$$\left\{\mathbf{V}_n(\lfloor nt\rfloor)'\Sigma_n^{-1}\mathbf{V}_n(\lfloor nt\rfloor);\ 0<t<1\right\}\ ,$$

where $\lfloor x\rfloor$ denotes the integer part of x. Let $n_1=\lfloor nt_1\rfloor$, with t_1 in $(0,1)$. In the same way as in Appendix A, as $V_{n,k}(\cdot)$ only differs from $U_{n,k}(\cdot)$ by a normalising factor, we can prove that $V_{n,k}(n_1) = \hat{V}_{n,k}(n_1) + o_p(1)$, with $0<n_1<n$ and

$$\hat{V}_{n,k}(n_1) = \frac{n_1}{n^{3/2}}\sum_{j=n_1+1}^{n}h_{1,k}(X_{j,k}) - \frac{n-n_1}{n^{3/2}}\sum_{i=1}^{n_1}h_{1,k}(X_{i,k})\ ,$$

where $h_{1,k}(x) = 2F_k(x)-1$ and that

$$\mathbb{E}[\hat{V}_{n,k}(n_1)\hat{V}_{n,k'}(n_1)] \to 4t_1(1-t_1)\operatorname{Cov}\big(F_k(X_{1,k}),F_{k'}(X_{1,k'})\big),\ \text{as } n\to\infty\ . \quad (38)$$

Let $n_2=\lfloor nt_2\rfloor$. Since $1<n_1<n_2<n$, $n_1/n\to t_1\in(0,1)$, and $n_2/n\to t_2\in(0,1)$ we get

$$\mathbb{E}[\hat{V}_{n,k}(n_1)\hat{V}_{n,k'}(n_2)] = \mathbb{E}\left[\left\{\frac{n_1}{n^{3/2}}\sum_{j=n_1+1}^{n}h_{1,k}(X_{j,k}) - \frac{n-n_1}{n^{3/2}}\sum_{i=1}^{n_1}h_{1,k}(X_{i,k})\right\}\right.$$
$$\left.\times\left\{\frac{n_2}{n^{3/2}}\sum_{j=n_2+1}^{n}h_{1,k'}(X_{j,k'}) - \frac{n-n_2}{n^{3/2}}\sum_{i=1}^{n_2}h_{1,k'}(X_{i,k'})\right\}\right]\ . \quad (39)$$

By decomposing the interval $[n_1+1,n]$ (resp. $[1,n_2]$) into $[n_1+1,n_2]$ and $[n_2+1,n]$ (resp. $[1,n_1]$ and $[n_1+1,n_2]$) and developing the expression, we obtain

$$\mathbb{E}[\hat{V}_{n,k}(n_1)\hat{V}_{n,k'}(n_2)] =$$

$$\mathbb{E}\left[\frac{(n-n_1)(n-n_2)}{n^3}\sum_{i=1}^{n_1}h_{1,k}(X_{i,k})h_{1,k'}(X_{i,k'}) - \frac{n_1(n-n_2)}{n^3}\sum_{j=n_1+1}^{n_2}h_{1,k}(X_{i,k})h_{1,k'}(X_{i,k'})\right.$$

$$\left.+\frac{n_1 n_2}{n^3}\sum_{j=n_2+1}^{n}h_{1,k}(X_{i,k})h_{1,k'}(X_{i,k'})\right]$$

$$=\frac{n_1(n-n_2)}{n^2}\Sigma_{kk'} \longrightarrow t_1(1-t_2)\Sigma_{kk'}, \text{ as } n\to\infty. \quad (40)$$

With (38) and (40), we obtain

$$\begin{pmatrix}\hat{\mathbf{V}}_n(n_1)\\ \hat{\mathbf{V}}_n(n_2)\end{pmatrix} \xrightarrow{d} \mathscr{N}\left(0; \left(\begin{array}{c|c} t_1(1-t_1)\Sigma & t_1(1-t_2)\Sigma \\ \hline t_1(1-t_2)\Sigma & t_2(1-t_2)\Sigma \end{array}\right)\right), \quad (41)$$

which is equivalent to

$$\begin{pmatrix}\hat{\mathbf{V}}_n(n_1)\\ \hat{\mathbf{V}}_n(n_2)\end{pmatrix} \xrightarrow{d} \begin{pmatrix}\Sigma^{\frac{1}{2}} & 0 \\ 0 & \Sigma^{\frac{1}{2}}\end{pmatrix}\begin{pmatrix}\mathbf{B}(t_1)\\ \mathbf{B}(t_2)\end{pmatrix}, \quad (42)$$

where $\mathbf{B}(t) = (B_1(t),\dots,B_K(t))$, $0\le t\le 1$. For the sake of clarity and without loss of generality, (37) is thus proved in the case $p=2$ by applying the continuous function

$$\begin{pmatrix}x_1\\ x_2\end{pmatrix} \longmapsto \begin{pmatrix}x_1' x_1\\ x_2' x_2\end{pmatrix}, \text{ where } x_1, x_2 \in \mathbb{R}^K. \quad (43)$$

In the following, we check the tightness condition, that is, for $0 < t_1 < t < t_2 < 1$, we show that

$$\mathbb{E}\left[|\hat{\mathbf{V}}_n(\lfloor nt\rfloor)\Sigma^{-1}\hat{\mathbf{V}}_n(\lfloor nt\rfloor) - \hat{\mathbf{V}}_n(\lfloor nt_1\rfloor)\Sigma^{-1}\hat{\mathbf{V}}_n(\lfloor nt_1\rfloor)|^2\right.$$
$$\left.\times|\hat{\mathbf{V}}_n(\lfloor nt_2\rfloor)\Sigma^{-1}\hat{\mathbf{V}}_n(\lfloor nt_2\rfloor) - \hat{\mathbf{V}}_n(\lfloor nt\rfloor)\Sigma^{-1}\hat{\mathbf{V}}_n(\lfloor nt\rfloor)|^2\right] \le C|t_2-t_1|^2, \quad (44)$$

where C is a positive constant. Let $\mathbf{x}_n(t) = (x_{n,1}(t),\dots,x_{n,K}(t))' = A\hat{\mathbf{V}}_n(\lfloor nt\rfloor)$, where $A = \Sigma^{-\frac{1}{2}}$, whose (p,q)th entry is denoted by $a_{p,q}$. The l.h.s. of (44) can thus be rewritten as

$$\mathbb{E}\left[|\mathbf{x}_n'(t)\mathbf{x}_n(t) - \mathbf{x}_n'(t_1)\mathbf{x}_n(t_1)|^2|\mathbf{x}_n'(t_2)\mathbf{x}_n(t_2) - \mathbf{x}_n'(t)\mathbf{x}_n(t)|^2\right]$$

$$= \mathbb{E}\left[\left|\sum_{k=1}^{K}\{x_{n,k}^2(t) - x_{n,k}^2(t_1)\}\right|^2 \left|\sum_{k'=1}^{K}\{x_{n,k'}^2(t_2) - x_{n,k'}^2(t)\}\right|^2\right]. \quad (45)$$

Note that

$$\sum_{k=1}^{K}\{x_{n,k}^2(t) - x_{n,k}^2(t_1)\} = \sum_{k=1}^{K}(x_{n,k}(t) - x_{n,k}(t_1))(x_{n,k}(t) + x_{n,k}(t_1))$$

$$= \sum_{k=1}^{K}\left(\sum_{p=1}^{K}a_{k,p}[\hat{V}_{n,p}(\lfloor nt\rfloor) - \hat{V}_{n,p}(\lfloor nt_1\rfloor)]\right)\left(\sum_{p'=1}^{K}a_{k,p'}[\hat{V}_{n,p'}(\lfloor nt\rfloor) + \hat{V}_{n,p'}(\lfloor nt_1\rfloor)]\right)$$

$$= \sum_{p,p'=1}^{K}b_{p,p'}\left[\hat{V}_{n,p}(\lfloor nt\rfloor) - \hat{V}_{n,p}(\lfloor nt_1\rfloor)\right]\left[\hat{V}_{n,p'}(\lfloor nt\rfloor) + \hat{V}_{n,p'}(\lfloor nt_1\rfloor)\right], \quad (46)$$

where $b_{p,p'} = \sum_{k=1}^{K} a_{k,p} a_{k,p'}$ is the (p,p')th element of the matrix $B = A^2 = \Sigma^{-1}$. Similarly

$$\sum_{k'=1}^{K} \{x_{n,k'}^2(t_2) - x_{n,k'}^2(t)\} = \sum_{q,q'=1}^{K} b_{q,q'} \left[\hat{V}_{n,q}(\lfloor nt_2 \rfloor) - \hat{V}_{n,q}(\lfloor nt \rfloor)\right] \left[\hat{V}_{n,q'}(\lfloor nt_2 \rfloor) + \hat{V}_{n,q'}(\lfloor nt \rfloor)\right]. \quad (47)$$

Using the notations $\ell = \lfloor nt \rfloor$, $\ell_1 = \lfloor nt_1 \rfloor$ and $\ell_2 = \lfloor nt_2 \rfloor$, and decomposing the interval $[1,n]$ into $[1, \ell_1]$, $[\ell_1 + 1, \ell]$, $[\ell + 1, \ell_2]$ and $[\ell_2 + 1, n]$, we get

$$\hat{V}_{n,p}(\ell) - \hat{V}_{n,p}(\ell_1) = \frac{\ell - \ell_1}{n^{3/2}} \left(\sum_{i=1}^{\ell_1} h_{1,p}(X_{i,p})\right) - \frac{n - (\ell - \ell_1)}{n^{3/2}} \left(\sum_{i=\ell_1+1}^{\ell} h_{1,p}(X_{i,p})\right)$$
$$+ \frac{\ell - \ell_1}{n^{3/2}} \left(\sum_{i=\ell+1}^{\ell_2} h_{1,p}(X_{i,p})\right) + \frac{\ell - \ell_1}{n^{3/2}} \left(\sum_{i=\ell_2+1}^{n} h_{1,p}(X_{i,p})\right), \quad (48)$$

and

$$\hat{V}_{n,p}(\ell) + \hat{V}_{n,p}(\ell_1) = \frac{\ell + \ell_1}{n^{3/2}} \left(\sum_{i=1}^{\ell_1} h_{1,p}(X_{i,p})\right) - \frac{n - (\ell + \ell_1)}{n^{3/2}} \left(\sum_{i=\ell_1+1}^{\ell} h_{1,p}(X_{i,p})\right)$$
$$+ \frac{\ell + \ell_1}{n^{3/2}} \left(\sum_{i=\ell+1}^{\ell_2} h_{1,p}(X_{i,p})\right) + \frac{\ell + \ell_1}{n^{3/2}} \left(\sum_{i=\ell_2+1}^{n} h_{1,p}(X_{i,p})\right), \quad (49)$$

with similar results for the terms of (47). Equation (45) is the expected value of the product of the squares of (46) and (47). Using Cauchy-Schwarz inequality, (45) is bounded above by the sum of several terms, obtained by inserting (48) and (49) in (46) and (47), respectively. Among these terms, we consider the case of:

$$C_1 \sum_{p,p'=1}^{K} \sum_{q,q'=1}^{K} b_{p,p'}^2 b_{q,q'}^2 \frac{(n-(\ell-\ell_1))^2(\ell+\ell_1)^2(n-(\ell_2-\ell))^2(\ell_2+\ell)^2}{n^{12}}$$
$$\times \mathbb{E}\left[\left|\sum_{i=\ell_1+1}^{\ell} h_{1,p}(X_{i,p})\right|^2 \left|\sum_{i=1}^{\ell_1} h_{1,p}(X_{i,p})\right|^2 \left|\sum_{i=\ell+1}^{\ell_2} h_{1,p}(X_{i,p})\right|^2 \left|\sum_{i=\ell_2+1}^{n} h_{1,p}(X_{i,p})\right|^2\right]. \quad (50)$$

Using the independence of $(X_{i,k})_{1 \le i \le n}$, the expected value in (50) can be separated into the product of four expected values, and thus can be bounded by

$$(\ell - \ell_1)\ell_1(\ell_2 - \ell)(n - \ell_2)/3^4 \le n^2(\ell_2 - \ell_1)^2/3^4. \quad (51)$$

Equation (50) is thus bounded above by a quantity proportional to $(\ell_2 - \ell_1)^2/n^2 = (\lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor)^2/n^2$. All the terms appearing in the expansion of (45) can be treated similarly. This completes the proof of (44) and thus ensures that

$$\sup_{0<t<1} \mathbf{V}_n(\lfloor nt \rfloor)' \Sigma^{-1} \mathbf{V}_n(\lfloor nt \rfloor) \xrightarrow{d} \sup_{0<t<1} \sum_{k=1}^{K} B_k^2(t), \ n \to \infty. \quad (52)$$

In order to prove (52) when Σ is replaced by $\hat{\Sigma}_n$, it enough to prove that $\sup_{0<t<1} |\mathbf{V}_n(\lfloor nt \rfloor)'(\Sigma^{-1} - \hat{\Sigma}_n^{-1})\mathbf{V}_n(\lfloor nt \rfloor)| = o_p(1)$. Note that

$$|\mathbf{V}_n(\lfloor nt \rfloor)'(\Sigma^{-1} - \hat{\Sigma}_n^{-1})\mathbf{V}_n(\lfloor nt \rfloor)| \leq \|\hat{\Sigma}_n^{-1} - \Sigma^{-1}\| \sup_{0<t<1} \|\mathbf{V}_n(\lfloor nt \rfloor)\|^2 ,$$

where $\hat{\Sigma}_n^{-1} \xrightarrow{p} \Sigma^{-1}$ and $\sup_{0<t<1} \|\mathbf{V}_n(\lfloor nt \rfloor)\|^2 = O_p(1)$, by (52), which concludes the proof.

References

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Barry, D. and Hartigan, J. A. (1992). Product partition models for change point models. *Ann. Statist.*, 20:260–279.

Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall.

Bellman, R. (1961). On the approximation of curves by line segments using dynamic programming. *Communications of the ACM*, 4(6):284.

Bickel, P. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2. *Ann. Math. Stat.*, 36:160–173.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley, New York.

Bleakley, K. and Vert, J.-P. (2011). The group fused lasso for multiple change-point detection. Technical report. arXiv:1106.4199.

Brodsky, B. E. and Darkhovsky, B. S. (2000). *Non-parametric statistical diagnosis*. Kluwer Academic Publishers, Dordrecht.

Chen, J. and Gupta, A. K. (2000). *Parametric statistical change point analysis*. Birkhäuser, Boston, MA.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley, New-York.

Désobry, F., Davy, M., and Doncarli, C. (2005). An online kernel change detection algorithm. *IEEE Trans. Signal Process.*, 53(8):2961–2974.

Estrella, A. (2003). Critical values and *p*-values of Bessel process distributions: computation and application to structural break tests. *Econometric Theory*, 19(6):1128–1143.

Fearnhead, P. (2006). Exact and efficient bayesian inference for multiple changepoint problems. *Statist. Comput.*, 16:203–213.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. J. (2006). A kernel method for the two-sample-problem. In *Advances in Neural Information Processing Systems*, pages 513–520.

Harchaoui, Z., Bach, F., and Moulines, E. (2008). Kernel change-point analysis. In *Advances in Neural Information Processing Systems*.

Harchaoui, Z. and Cappé, O. (2007). Retrospective multiple change-point estimation with kernels. In *IEEE Workshop on Statistical Signal Processing*.

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.*, 105(492):1480–1493.

Hettmansperger, T. P., Möttönen, J., and Oja, H. (1998). Affine invariant multivariate rank tests for several samples. *Statist. Sinica*, 8(3):785–800.

Jurečková, J. and Kalina, J. (2012). Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli*, 18(1):229–251.

Kay, S. (1993). *Fundamentals of statistical signal processing: detection theory*. Prentice-Hall, Inc.

Kiefer, J. (1959). *K*-sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests. *Ann. Math. Statist.*, 30:420–447.

Kruskal, W. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.*, 53(284):814–861.

Lavielle, M. (2005). Using penalized contrasts for the change-points problems. *Signal Processing*, 85(8):1501–1510.

Lehmann, E. (1975). *Nonparametrics: statistical methods based on ranks*. Holden-Day Inc.

Lévy-Leduc, C. and Roueff, F. (2009). Detection and localization of change-points in high-dimensional network traffic data. *Ann. Applied Statist.*, 3(2):637–662.

Lung-Yut-Fong, A., Lévy-Leduc, C., and Cappé, O. (2012). Distributed detection/localization of change-points in high-dimensional network traffic data. *Statist. Comput.*, 22(12):485–496.

Möttönen, J., Oja, H., and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.*, 25(2):542–552.

Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.*, 26(3):319–343.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41:100–115.

Picard, F., Robin, S., Lavielle, M., Vaisse, C., and Daudin, J.-J. (2005). A statistical approach for array CGH data analysis. *BMC Bioinformatics*, 6(1):27.

Poor, H. and Hadjiliadis, O. (2009). *Quickest Detection*. Cambridge University Press, Cambridge.

Siegmund, D. (1985). *Sequential analysis*. Springer Series in Statistics. Springer-Verlag, New York.

Srivastava, M. S. and Worsley, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Statist. Assoc.*, 81(393):199–204.

Talih, M. and Hengartner, N. (2005). Structural learning with time-varying components: tracking the cross-section of the financial time series. *J. Royal Statist. Soc. B*, 67(3):321–341.

Tartakovsky, A., Nikiforov, I., and Basseville, M. (2014). *Sequential Analysis : Hypothesis Testing and Changepoint Detection*. CRC Press, Taylor & Francis Group.

Tartakovsky, A., Rozovskii, B., Blazek, R., and Kim, H. (2006). A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Trans. Signal Process.*, 54(9):3372 – 3382.

Topchii, A., Tyurin, Y., and Oja, H. (2003). Inference based on the affine invariant multivariate Mann-Whitney-Wilcoxon statistic. *J. Nonparametr. Stat.*, 15(4-5):403–419.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

Vert, J. and Bleakley, K. (2010). Fast detection of multiple change-points shared by many signals using group LARS. In *Advances in Neural Information Processing Systems 23*.

Wei, L. J. and Lachin, J. M. (1984). Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *J. Amer. Statist. Assoc.*, 79(387):653–661.