

## Cramér-Von Mises Statistic for Repeated Measures

### El estadístico de Cramér-Von Mises para medidas repetidas

PABLO MARTÍNEZ-CAMBLOR<sup>12,a</sup>, CARLOS CARLEOS<sup>2,b</sup>, NORBERTO CORRAL<sup>2,c</sup>

<sup>1</sup>OFICINA DE INVESTIGACIÓN BIOSANITARIA (OIB), FICYT, OVIEDO, SPAIN

<sup>2</sup>DEPARTAMENTO ESTADÍSTICA E IO Y DM, UNIVERSIDAD DE OVIEDO, ASTURIAS, SPAIN

---

#### Abstract

The Cramér-von Mises criterion is employed to compare whether the marginal distribution functions of a  $k$ -dimensional random variable are equal or not. The well-known Donsker invariance principle and the Karhunen-Loève expansion is used in order to derive its asymptotic distribution. Two different resampling plans (one based on permutations and the other one based on the general bootstrap algorithm, gBA) are also considered to approximate its distribution. The practical behaviour of the proposed test is studied from a Monte Carlo simulation study. The statistical power of the test based on the Cramér-von Mises criterion is competitive when the underlying distributions are different in location and is clearly better than the Friedman one when the sole difference among the involved distributions is the spread or the shape. Both resampling plans lead to similar results although the gBA avoids the usual required interchangeability assumption. Finally, the method is applied on the study of the evolution inequality incomes distribution between some European countries along the years 2000 and 2011.

**Key words:** Asymptotic Distribution, Bootstrap, Cramér-von Mises statistic, Hypothesis testing, Permutation test, Repeated Measures.

#### Resumen

El criterio de Cramér-von Mises es empleado para comparar la igualdad entre las distribuciones marginales de una variable aleatoria  $k$ -dimensional. El conocido principio de invarianza de Donsker y la expansión de Karhunen-Loève se usan para derivar su distribución asintótica. Dos planes de muestreo diferentes (uno basado en permutaciones y el otro basado en el algoritmo bootstrap general, gBA) son usados para aproximar su distribución. El comportamiento práctico del test propuesto es estudiado mediante simulaciones de Monte Carlo. La potencia estadística del test basado en el criterio

---

<sup>a</sup>Biostatistician. E-mail: pmcambor@hotmail.com

<sup>b</sup>Professor. E-mail: carleos@uniovi.es

<sup>c</sup>Lecturer. E-mail: norbert@uniovi.es

de Cramér-von Mises es competitiva cuando la distribuciones subyacentes difieren en el parámetro de localización. Este test es claramente superior al de Friedman cuando las únicas diferencias son en la dispersión o la forma. Ambos planes de remuestreo obtienen resultados similares aunque el gBA evita la hipótesis de intercambiabilidad. Finalmente, el método propuesto es aplicado al estudio de la evolución de las desigualdades en los ingresos entre algunos países Europeos entre los años 2000 y 2011.

**Palabras clave:** Bootstrap, distribución asintótica, estadístico de Cramér-von Mises, medidas repetidas, test de hipótesis, test de permutaciones.

## 1. Introduction

The comparison of the equality among the marginal distribution functions of a  $k$ -dimensional random variable is a common problem in statistical inference (for example, in biomedicine, in problems of comparing diagnostic procedures or bioequivalence (Freitag, Czado & Munk 2007)). In practice, most frequent cases are the study of one feature measured on the same subjects at different time moments (analysis of repeated measures) and matched studies. Despite of, there exists a number of methods of comparing the equality among  $k$ -distributions from independent samples, the  $k$ -sample problem for dependent data has not been as widely studied and, the traditional parametric (ANOVA) and nonparametric (Friedman test) repeated measures procedures are the usual used techniques to solve these problems.

In this context, several rank tests have been proposed. In a non exhaustive revision: Ciba-Geigy & Olsson (1982) developed a specific one for comparing dispersion in paired samples design; Lam & Longnecker (1983) introduced modifications which improve the power of the classical Wilcoxon rank sum test for this topic; Munzel (1999a) used the normalized version of distribution functions to derive an asymptotic theory for rank statistics including ties and considered a mixed model which permits almost arbitrary dependences; Munzel (1999b) studied different nonparametric permutation methods for repeated measures problems in a two sample framework; most recently, Freitag et al. (2007) proposed a test based on the Mallows distance with this goal. Other authors as Govindarajulu (1995) Govindarajulu (1997) or Podgor & Gastwirth (1996) also dealt with this topic from different approaches.

Although the use of bootstrap on multivariate problems is straightforward in order to build confidence intervals and related estimates, the way to resampling under the null (in particular, the way to involve this assumption on the resampling) for preserving the original data structure is not direct and the use of bootstrap on hypothesis testing (which involve paired design) is not so clear. It is not trivial how to involve the (null) hypothesis of equality of the  $k$  marginal distributions of a multivariate random variable. The most common procedure, the *permutation* test (see, for example, Good 2000, Munzel 1999b), implies that the different components of the  $k$ -dimensional random vector must be *interchangeable* (see Venkatraman & Begg 1996 or, most recently, Nelsen 2007). Under the null, this is not a very strong

assumption to compare two samples (most of the previous cited works engage, exclusively, on this particular case) but, for three or more samples it means that the relationship between each pair must be the same (it is also known as *sphericity hypothesis*) and, in spite of for most of the usual statistics, in practice, the permutation test has demonstrated its robustness with respect to this assumption, it is usually violated.

In this paper, the authors deal with the problem of comparing the equality among the  $k$  marginal distribution functions from a typical multivariate problem. With this goal, the traditional Cramér-von Mises criterion is considered. The Donsker invariance principle and the classical Gaussian processes theory, in particular, the Karhunen-Loève expansion, are used in order to obtain (a not explicit version of) the asymptotic distribution for the Cramér-von Mises statistic when the samples are from the same subjects. The properties of this statistic allow to develop a resampling procedure which does not need the (usual) *interchangeability* (or sphericity) assumption. This method is described and its consistency is proved. We think it is worth mentioning that, the considered procedures (the asymptotic, permutation and the bootstrap ones), are simple, useful and easily to implement. A simulation study is carried out (Section 3); its results suggest that the Cramér-von Mises criterion obtains good results in all considered situations and it is clearly better than the Friedman test when distributions differ mainly in their spread or shape. These results are the usual ones when the Cramér-von Mises criterion is used in other context (see, for example, Martínez-Cambor & Uña-Álvarez (2009) or Martínez-Cambor (2011)). Finally, the proposed method is applied on the study of the inequality incomes between thirty European countries during the years 2000 and 2011 (Section 4).

During the revision process of this paper, it has been published the work of Quessy & Éthier (2012) (QE) which, from a slight different approach, deals with the same problem. The main results of the present manuscript had been developed around 2008-2009 and, of course, independently of the previously cited work. In order to keep this independence and, in spite of several reported results are overlapping with the obtained by QE, we have maintained them in the appendix.

## 2. Cramér-von Mises Statistic for Repeated Measures

The well-known Cramér-von Mises criterion introduced, separately by Harald Cramér and Richard Edler von Mises (Cramér 1928, Von Mises 1991), was originally to compare the goodness of fit of a probability distribution  $F^*$  and a fixed distribution function  $F_0$  and is given by

$$W^2 = \int (F^*(t) - F_0(t))^2 dF_0(t)$$

In the immediately one-sample applications,  $F_0$  is the theoretical cumulative distribution function (CDF) and  $F^*$  is the empirical cumulative distribution function (ECDF),  $\hat{F}_n$ . Csörgő & Faraway (1996) derived the exact distribution for

this statistic and proposed a correction for its asymptotic distribution. Anderson (1962) derived the asymptotic distribution for the two-sample case. A standard  $k$ -dimensional generalization was proposed by Kiefer (1959) which considered the expression

$$W_k^2 = \sum_{i=1}^k n_i \int (\hat{F}_{n_i}(X_i, t) - \hat{F}_N(X, t))^2 d\hat{F}_N(X, t)$$

where  $N = \sum_{i=1}^k n_i$  and  $\hat{F}_{n_i}(X_i, t)$  and  $\hat{F}_N(X, t)$  are the ECDF referred to the  $i$ th sample ( $1 \leq i \leq k$ ) and to the pooled sample, respectively. Brown (1982) also dealt with the  $k$ -sample problem, he studied the asymptotic distribution and introduced a permutation test based on the same criterion. In Martínez-Cambor & Uña-Álvarez (2009), the statistical power of this statistic was considered in a simulation study (joint with other six statistics based on the ECDF and four more based on the kernel density estimator). The  $W_k^2$  test obtained very competitive results in the eight considered models (four symmetrical and four asymmetrical).

In this section we study the different approximations for the distribution of  $W_k^2$  when data are from a multivariate variable i.e., in our case, we have a  $k$ -dimensional random sample  $\mathbf{X} = (X_1, \dots, X_k)$  with  $X_i = (x_{i1}, \dots, x_{in})$  ( $n$  subjects have been collected) for  $i \in 1, \dots, k$ , from a  $k$ -dimensional random variable  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k)$ . For each  $\mathbf{u} = (u_1, \dots, u_k)$  with  $u_1, \dots, u_k \in \mathbb{R}$  the  $k$  dimensional functions

$$\begin{aligned} \hat{\mathbf{F}}_n(\mathbf{X}, \mathbf{u}) &= (\hat{F}_{n,1}(X_1, u_1), \dots, \hat{F}_{n,k}(X_k, u_k)) \\ \mathbf{F}(\mathbf{u}) &= (F_1(u_1), \dots, F_k(u_k)) \end{aligned}$$

denote the vectors with the ECDFs and the theoretical cumulative distribution functions (CDFs), respectively. In Theorem 1, it is proved (we must remark; a non explicit version of) the asymptotic distribution for the statistic

$$W_k^2(n) = \sum_{i=1}^k n \int (\hat{F}_{n,i}(X_i, t) - \hat{F}_{n,\bullet}(\mathbf{X}, t))^2 d\hat{F}_{n,\bullet}(\mathbf{X}, t)$$

where  $\hat{F}_{n,i}(X_i, t)$  ( $1 \leq i \leq k$ ) is the ECDF referred to the  $i$ th sample and  $\hat{F}_{n,\bullet}(\mathbf{X}, t) = k^{-1} \sum_{i=1}^k \hat{F}_{n,i}(X_i, t)$ , when the (null) hypothesis

$$H_0 : F_1 = \dots = F_k (= F) \tag{1}$$

is true.

**Theorem 1.** *Let  $\boldsymbol{\xi}$  be a  $k$ -dimensional random vector and let  $\mathbf{X}$  be a random sample from  $\boldsymbol{\xi}$  (with size  $n$ ), by using the above notation, if  $F_1 = \dots = F_k (= F)$  (null hypothesis), it is hold the (weak) convergence*

$$W_k^2(n) \xrightarrow{\mathcal{L}}_n \sum_{i=1}^k \sum_{l \in \mathbb{N}} \lambda_{i,l} M_{i,l}^2$$

where  $\{\mathbf{M}_l = (M_{1,l}, \dots, M_{k,l})\}_{l \in \mathbb{N}}$  is a sequence of  $k$ -dimensional, normal distributed random variables whose marginals follow a  $\mathcal{N}(0,1)$  distribution and  $\{\{\lambda_{i,l}\}_{i=1}^k\}_{l \in \mathbb{N}}$  are non negative constants satisfying  $\sum_{l \in \mathbb{N}} \lambda_{i,l}^2 < \infty$  for  $1 \leq i \leq k$ .

The above Theorem guarantees the consistency and gives the convergence rate for the studied statistic. However, strictly speaking, this result does not provide its distribution in full. In order to build asymptotic critical regions, the explicit values for the  $\{\{\lambda_{i,l}\}_{i=1}^k\}_{l \in \mathbb{N}}$  coefficients must be known (eigenvalues and eigenfunctions must be computed). However, we want to note that this is a non-trivial problem which involves complex (and sometimes, for some readers, cumbersome analysis see, for example, Deheuvels 2005). In addition, these eigenvalues depend on the covariance data structure and they should be computed particularly for each problem. The following remark is devoted to point out some comments about the eigenvalues calculation in the two-sample case.

**Note 1.** In the two sample-case, the asymptotic distribution of  $W_k^2(n)$  under the null is equivalent to the distribution of  $W^2 = \int (\mathcal{W}_1\{t\} - \mathcal{W}_2\{t\})^2 dt$ , where  $\mathcal{W}_i\{t\}$  ( $i \in 1, 2$ ) is a standard Brownian bridge. Eigenvalues and eigenfunctions are the non zero solutions to the Fredholm type integral equation

$$\lambda_j e_j(u) = \int \mathbb{C}(u, v) e_j(v) dv$$

with the above restrictions on the eigenfunctions,  $e_j$  (i.e. orthonormality). In this particular setting,

$$\begin{aligned} \mathbb{C}(u, v) &= \mathbb{E}[(\mathcal{W}_1\{u\} - \mathcal{W}_2\{u\})(\mathcal{W}_1\{v\} - \mathcal{W}_2\{v\})] \\ &= 2(u \wedge v - uv) - (f(u, v) + f(v, u)) \end{aligned}$$

where  $f(s, t) = \mathbb{E}[\mathcal{W}_1\{s\}\mathcal{W}_2\{t\}]$  (note that  $f(s, t) = 0$  for independent samples). Obviously, the particular solutions depend on the function  $f$ . For instance, assuming  $f(u, v) + f(v, u) = u \wedge v - uv$ , functions  $\sin(j\pi u)$  and  $\cos(j\pi u)$  ( $j \in \mathbb{N}$ ) are possible solutions which lead to eigenvalues in the form  $\lambda_j = (j\pi)^{-2}$  (see, for instance, Van der Vaart, 1998).  $\square$

Usually, in order to approximate the asymptotic distribution, the largest eigenvalue is taken and the other ones are ignored i.e., by using the coefficients properties (see the Theorem 1 proof in the Appendix, in particular, equation (8)), it is obtained the approximation

$$\sum_{i=1}^k \sum_{l \in \mathbb{N}} \lambda_{i,l} M_{i,l}^2 = \sum_{i=1}^k \left( \sum_{l \in \mathbb{N}} \lambda_{i,l} (M_{i,l}^2 - 1) + C_i \right) \sim \sum_{i=1}^k (\lambda_{i,1} (M_{i,1}^2 - 1) + C_i)$$

Unfortunately, the first eigenvalue is also unknown. However, for each  $i \in 1, \dots, k$ , we can approximate the first (and, therefore, the biggest) eigenvalue by

$$\lambda_{i,1}^2 \sim \tilde{\lambda}_{i,1}^2 = \iint \mathbb{C}_{i,i}(s, t)^2 dF(s) dF(t)$$

Note that it is known that  $\tilde{\lambda}_{i,1} \geq \lambda_{i,1}$  ( $i \in 1, \dots, k$ ) and the equality is true only when  $\lambda_{i,l} = 0 \quad \forall l > 1$ . Finally, in order to save the relationship among the different involved samples, we build  $\mathbf{M}_1 = (M_{1,1}, \dots, M_{k,1})$  such that, for  $1 \leq i, j \leq k$

$$C_i C_j \mathbb{E}[M_{i,1}^2 M_{j,1}^2] = \mathbb{E} \left[ \int \{Y_{F_i}^i(t) - \bar{Y}_{F, \bullet}(t)\}^2 dF(t) \int \{Y_{F_j}^j(t) - \bar{Y}_{F, \bullet}(t)\}^2 dF(t) \right]$$

We can work, without loss of generality, with  $\mathbb{E}[M_{1,1}M_{2,1}]$ . It is easy to check that  $2\mathbb{E}[M_{1,1}M_{2,1}]^2 = \mathbb{E}[M_{1,1}^2M_{2,1}^2] - 1$  and

$$\begin{aligned}\mathbb{E}[M_{1,1}^2M_{2,1}^2] &= \frac{1}{C_1C_2} \mathbb{E} \left[ \int \{Y_{F_1}^1(t) - \bar{Y}_{F,\bullet}(t)\}^2 dF(t) \int \{Y_{F_2}^2(t) - \bar{Y}_{F,\bullet}^2(t)\}^2 dF(t) \right] \\ &= \frac{1}{C_1C_2} \mathbb{E} \left[ \iint \{Y_{F_1}^1(t) - \bar{Y}_{F,\bullet}(t)\}^2 \{Y_{F_2}^2(s) - \bar{Y}_{F,\bullet}(s)\}^2 dF(t)dF(s) \right] \\ &= \frac{1}{C_1C_2} \iint \mathbb{E}[(\mathbf{Y}_F(t)\mathbf{a}_1^t)^2(\mathbf{Y}_F(s)\mathbf{a}_2^t)^2] dF(t)dF(s)\end{aligned}$$

With some additional computes and taking into account that, for  $1 \leq i, j \leq k$ ,  $F_{i,j}(u, v) = F_{j,i}(v, u)$ , it is obtained

$$\begin{aligned}C_{1,2}(s, t) &= \mathbb{E}[(\mathbf{Y}_F(s)\mathbf{a}_1^t)(\mathbf{Y}_F(t)\mathbf{a}_2^t)] \\ &= F_{1,2}(s, t) - \bar{F}_{1,\cdot}(s, t) - \bar{F}_{2,\cdot}(t, s) + \bar{F}_{\cdot,\cdot}(s, t)\end{aligned}\quad (2)$$

then,

$$\mathbb{E}[M_{1,1}^2M_{2,1}^2] = \frac{1}{C_1C_2} \iint (2C_{1,2}^2(s, t) + C_1(s)C_2(t)) dF(s)dF(t)\quad (3)$$

and the asymptotic distribution can be approximated by

$$C_A = \sum_{i=1}^k \left( \tilde{\lambda}_{i,1} (M_{i,1}^2 - 1) + C_i \right)\quad (4)$$

We compute  $\tilde{\lambda}_{i,1}$  ( $1 \leq i \leq k$ ) by using the estimation of some parameters of the statistic and, unfortunately, from this method we cannot estimate any other eigenvalue. In the independent case, the quality of this approximation has been checked via simulations (see, for instance, Martínez-Camblor, Carelos & Corral (2012), and references therein).

Note that both the expected value and the variance of  $C_A$  are equal to the  $W_k^2(n)$  ones. The (theoretical) unknown parameters which are involved in the equation (4) can be estimated by putting the respective ECDFs instead of the theoretical ones (typical plug-in method) in their explicit expressions (equations (2) and (3)). At this point, it is worth to remember that, under the null hypothesis, all the marginal functions are equal. Once these values are computed, the asymptotic distribution under the null might be approximated by using some bound for the quadratic forms (see, for example, Alkarni & Siddiqui 2001) or by using the Monte Carlo method: Generating  $T$  independent samples (with the original sample size) from the  $k$ -dimensional normal distribution (previously we must compute its correlation matrix by using the corresponding equations) and computing the respective  $T$  asymptotic values of the statistic by using (4). In Section 3, the latter possibility is employed in the simulation study.

On the other hand, the Cramér-von Mises statistic properties allow to propose an useful resampling plan in order to approximate its distribution for paired samples in small size problems. The following subsection is devoted to develop a bootstrap approximation in the current context.

## 2.1. Bootstrap Approximation

The *bootstrap*, introduced and explored in detail by Bradley Efron (Efron 1979, Efron 1982), is an (not only but mainly nonparametric) intensive computer-based method of statistical inference which is often used in order to solve many real questions without the need of knowing the underlying mathematical formulas. Besides, under regularity conditions, the distribution bootstrap estimation is asymptotically minimax among all possible estimates (Beran 1982).

Despite of the bootstrap method has received a great deal of attention and popularity, its use on statistical hypothesis testing has received considerable, although minor attention (Martin 2007). Following Hall & Wilson (1991), many authors such as Westfall & Young (1993) have promoted null resampling as critical to the proper construction of bootstrap tests. However, in related sample distribution comparison, it is not straightforward how to resample under the null and the permutation tests (Good 2000) are the usual ones employed with this goal. In order to guarantee the consistency of the last method, *exchangeability* among the different components must be assumed (Venkatraman & Begg 1996) and, although for most of the statistics, in practice, this technique has proved its robustness with respect to this assumption, in  $k$ -dimensional problems ( $k > 2$ ) it is usually violated. Recently, Martínez-Camblor et al. (2012) proposed a general resampling plan which focus its use on hypothesis testing without the need of assuming additional conditions. In particular, for the present problem, under the null, it is easy to prove that:

**Theorem 2.** *Under the Theorem 1 assumptions and by using the same notation. Let  $\mathbf{X}^* = (X_1^*, \dots, X_k^*)$  be an independent random sample generated from  $\hat{F}_n(\mathbf{X}, \cdot)$  (multivariate ECDF referred to the random sample  $\mathbf{X}$ ). If*

$$W_k^{2,*}(n) = \sum_{i=1}^k n \int \{\hat{F}_{n,i}^*(X_i^*, t) - \hat{F}_{n,i}(X_i, t)\}^2 d\hat{F}_{n,\bullet}^*(\mathbf{X}^*, t) \\ - nk \int \{\hat{F}_{n,\bullet}^*(\mathbf{X}^*, t) - \hat{F}_{n,\bullet}(\mathbf{X}, t)\}^2 d\hat{F}_{n,\bullet}^*(\mathbf{X}^*, t)$$

where for each  $i \in 1, \dots, k$ ,  $\hat{F}_{n,i}^*(X_i^*, t)$  is the ECDF referred to  $X_i^*$  and  $\hat{F}_{n,\bullet}^*(\mathbf{X}^*, t) = k^{-1} \sum_{i=1}^k \hat{F}_{n,i}^*(X_i^*, t)$ . Under the null, it is held,

$$\left\{ \mathcal{P}_{\mathbf{X}} \left( \mathcal{W}_{\parallel}^{\epsilon,*}(\setminus) \leq \square \right) - \mathcal{P} \left( \mathcal{W}_{\parallel}^{\epsilon}(\setminus) \leq \square \right) \right\} \rightarrow_n 0 \quad a.s.$$

where  $\mathcal{P}_{\mathbf{X}}$  denotes probability conditionally on sample  $\mathbf{X}$ .

The above result proves the punctual convergence (for each, fixed,  $u \in \mathbb{R}$ ) of the bootstrap method. Uniform convergence can also be derived (under mild and usual conditions) from general theory of  $U$  and  $V$  statistics (see, for example, Arcones & Gine 1992). Theorem 2 guarantees that the distribution of  $W_k^2(n)$  can be approximated by the  $W_k^{2,*}(n)$  one and, as usual, this distribution can be approximated by using the Monte Carlo method following the algorithm:

$B_1$ . From the original sample,  $\mathbf{X}$ , compute  $W_k^2(n)$ .

$B_2$ . From the multivariate cumulative empirical distribution function,  $\hat{\mathbf{F}}_n(\mathbf{X}, t)$ , draw  $B$  independent  $k$ -dimensional random samples with size  $n$ ,

$$\mathbf{X}^{*,b} = (X_1^*, \dots, X_k^*), \quad 1 \leq b \leq B$$

$B_3$ . For  $b \in 1, \dots, B$  compute  $W_k^{2,*,b}(n)$ , from  $\mathbf{X}^{*,b}$ .

$B_4$ . The distribution of  $W_k^2(n)$  is approximated by  $\{W_k^{2,*,1}(n), \dots, W_k^{2,*,B}(n)\}$  i.e., the final  $P$ -value is given by

$$P = \frac{1}{B} \sum_{b=1}^B I\{W_k^{2,*,b}(n) > W_k^2(n)\}$$

The main difference between this algorithm and the *classical* bootstrap is that, in the proposed method, the null hypothesis (and only the null hypothesis) is used in order to compute the statistic (bootstrap) values instead of to be used to draw the bootstrap samples. We do not resampling from the null and this fact, allows to preserve the original data structure.

Permutation method is based on the idea that within the same subject, each value can be located in any position. For this claim, not only the null must be true but the interchangeability it also must be hold. Although, in practice, the permutation method has proved its robustness for a wide variety of statistics, let us to go to an extreme. We consider a three-sample problem (sample size  $n$ ) where the first and second variables are the same and the third one is independent from the other two. In this setting it is derived the equality

$$\begin{aligned} W_k^2(n) = & n \int \left[ \frac{1}{9} \{ \hat{F}_{n,1}(X_1, t) - \hat{F}_{n,3}(X_3, t) \}^2 \right] d\hat{F}_{n,\bullet}(\mathbf{X}, t) \\ & + n \int \left[ \frac{1}{9} \{ \hat{F}_{n,1}(X_1, t) - \hat{F}_{n,3}(X_3, t) \}^2 \right] d\hat{F}_{n,\bullet}(\mathbf{X}, t) \\ & + n \int \left[ \frac{4}{9} \{ \hat{F}_{n,3}(X_3, t) - \hat{F}_{n,1}(X_1, t) \}^2 \right] d\hat{F}_{n,\bullet}(\mathbf{X}, t) = S_{n,1} + S_{n,2} + S_{n,3}. \end{aligned}$$

It is obvious that the value of the difference between the  $\hat{F}_{n,1}$  and  $\hat{F}_{n,3}$  has not the same weight in the three summands. However the permutation algorithm assumes that the summands have the same distribution, in particular the same expected value. Table 1 depicts the means of  $S_{n,i}$  (labelled as  $\bar{S}_{n,i}$ ) for  $i \in 1, 2, 3$  in 2,000 Monte Carlo (MC) simulations and when the permutation (P) and the proposed bootstrap (B) are used. The observed rejection proportion ( $\alpha = 0.05$ ) and the value of  $W_k^2(n)$  are also included. The underlying distributions are uniforms on  $[0, 1]$  and  $n = 50$ .

Although the  $W_k^2(n)$  is, in general, well estimated by the permutation method (the mean is similar to the expected one), the total value is evenly distributed



TABLE 1: Means for the three summands and for  $W_k^2(n)$  in the problem above described for the Monte Carlo approximation (MC) (2,000 iterations) and for the Bootstrap and Permutation methods. Observed rejection percentages ( $\alpha = 0.05$ ) are also included.

	$\bar{S}_{n,1}$	$\bar{S}_{n,2}$	$\bar{S}_{n,3}$	$W_k^2(n)$	% Rejection
MC	0.0378	0.0378	0.1456	0.2213	5.0%
P	0.0753	0.0744	0.0749	0.2247	8.8%
B	0.0368	0.0368	0.1473	0.2210	5.4%

among the three summands. In spite of for the  $W_k^2(n)$  the results are, in general, good (the observed rejection proportion is too big, but this is an extreme and bit realistic problem), permutation method does not reflect the data structure and this fact can drive to mistake when different weighting are considered for the involved summands or, for instance, in presence of different missing data frameworks. In summary, the correct performance of the permutation method can not be guaranteed in absence of the exchangeability hypothesis.

### 3. Simulation Study

In order to investigate the practical behaviour of the proposed methodology, a Monte Carlo simulation study has been carried out. We estimate the statistical power ( $\alpha = 0.05$ ) from 2,000 Monte Carlo replications for different problems. For the Cramér-von Mises test, asymptotic approximation,  $C_A$  (the  $P$ -value is approximated from 499 Monte Carlo replications following the approximation given in (4)), bootstrap approximation,  $C_B$  ( $B = 499$  in algorithm  $B_1$ - $B_4$ ) and the permutation method,  $C_P$  (the  $P$ -value is also approximated from  $T = 499$  replications) are considered. Although the number of random combinations is small to obtain a good estimation for a particular  $P$ -value, it is enough to obtain a good estimation for the statistical power. Note that, here, we are not interested in the result for each particular problem but in the final rejection proportion. The classical non-parametric Friedman ( $F_R$ ) test is also included as the reference one. Let be  $\mathbf{Z} = (Z_1, Z_2, Z_3)$  a three dimensional random vector from a  $\mathcal{N}_3(\mathbf{0}, \Sigma)$  distribution where  $\mathbf{0} = (0, 0, 0)$  and the components of the covariance matrix are  $\sigma_{i,j} = 1$  if  $i = j$  and  $\sigma_{1,2} = \sigma_{1,3} = 1/4$  and  $\sigma_{2,3} = b$  (cases  $b = 1/4$  and  $b = 3/4$  are considered) and let be  $N_i$ ,  $1 \leq i \leq 4$ , independent random variables with standard normal distribution. A three dimensional random sample with size  $n$ ,  $\mathbf{X} = (X_1, X_2, X_3)$ , is drawn from the following symmetrical models (MD):

**0-I.**  $X_1 \equiv Z_1, X_2 \equiv Z_2, X_3 \equiv Z_3$  (Null hypothesis).

**1-I.**  $X_1 \equiv Z_1, X_2 \equiv Z_2, X_3 \equiv (1-a) * Z_3 + a * 3Z_3$ .

**2-I.**  $X_1 \equiv Z_1, X_2 \equiv Z_2, X_3 \equiv (1-a) * Z_3 + a * (Z_3 + 1)$ .

**3-I.**  $X_1 \equiv Z_1, X_2 \equiv Z_2, X_3 \equiv (1-a) * Z_3 + a * (\sqrt{3}Z_3 + 1)$ .

The following asymmetrical models are also considered:

**0-II.**  $X_1 \equiv Z_1^2 + N_1^2 + N_2^2, X_2 \equiv Z_2^2 + N_1^2 + N_3^2, X_3 \equiv Z_3^2 + N_2^2 + N_3^2$  (Null hypothesis).

**1-II.**  $X_1 \equiv Z_1^2 + N_1^2 + N_2^2, X_2 \equiv Z_2^2 + N_1^2 + N_3^2, X_3 \equiv (1-a) * (Z_3^2 + N_2^2 + N_3^2) + a * (Z_3 + 3)$ .

**2-II.**  $X_1 \equiv Z_1^2 + N_1^2 + N_2^2, X_2 \equiv Z_2^2 + N_1^2 + N_3^2, X_3 \equiv (1-a) * (Z_3^2 + N_2^2 + N_3^2) + a * Z_3^2$ .

**3-II.**  $X_1 \equiv Z_1^2 + N_1^2 + N_2^2$ ,  $X_2 \equiv Z_2^2 + N_1^2 + N_3^2$ ,  $X_3 \equiv (1 - a) * (Z_3^2 + N_2^2 + N_3^2) + a * (Z_3^2 + \sum_{i=1}^4 N_i^2)$ .

Where  $a = 3/4$  and  $M = (1 - b) * X + b * Y$  denotes a mixture which takes values on  $X$  with probability  $(1 - b)$  and on  $Y$  otherwise. A graphical representation for the respective density functions is shown in Figure 1.

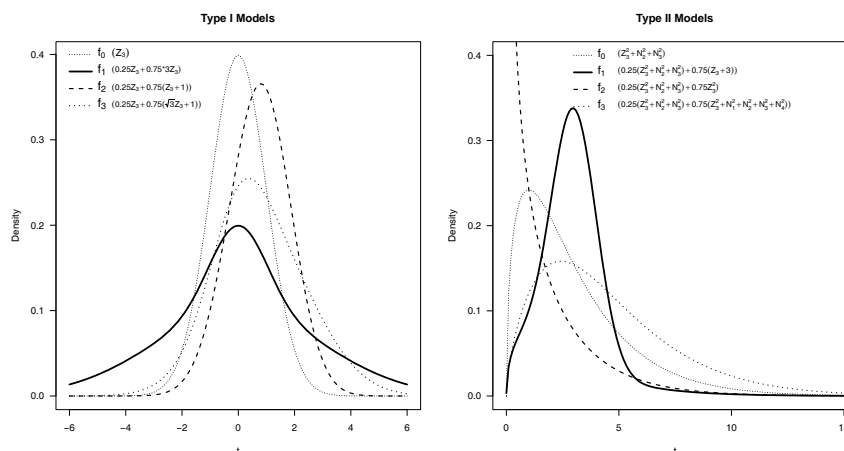


FIGURE 1: Density functions for the different considered models.

Table 2 shows the observed rejection proportions for type I (symmetrical) model for two different sample sizes ( $n = 25, 50$ ). Figure 2 depicts the observed statistical power for the type I models against sample size (sample sizes of 10, 25, 40, 50, 65 and 75 were considered). Despite of the rejected observed percentages are bit larger than the expected ones (in special for the  $C_P$  approximation for  $b = 3/4$ ) the nominal level is, in general, well respected. On the other hand, the Cramér-von Mises test obtains better results than the Friedman one even when the difference among the distributions is only in the position parameter while the variance-covariance structure is the same. Approximations by permutation and bootstrap obtain quite similar results, although the permutation one is a bit better for  $\sigma_{2,3} = 3/4$ . The asymptotic approximation obtains worst results for small sample sizes but quite similar than the other ones for middle sample sizes ( $n > 40$ ).

Table 3 and Figure 3 are analogous to Table 2 and Figure 2, respectively, for type II (asymmetrical) models. The nominal level is well respected in all considered cases. For type II models (Figure 3) the Cramér-von Mises criterion is still the best when the main difference is not the location parameter (model 1-II). When the location parameter is the main difference among the curves, the Friedman test is the best one in model 3-II and the Cramér-von Mises test is the best for model 2-II, although both tests obtain quite similar results. In this scheme the approximation to the asymptotic distribution for the Cramér-von Mises test is slow and, in general, its results are not competitive for  $n \leq 50$ .

TABLE 2: Observed rejection probabilities for type I (symmetrical ones) models. The nominal level is  $\alpha = 0.05$ .

	$n$	$b = 1/4$				$b = 3/4$			
		$C_B$	$C_P$	$C_A$	$F_R$	$C_B$	$C_P$	$C_A$	$F_R$
<b>0-I</b>	<b>25</b>	0.059	0.060	0.037	0.051	0.065	0.074	0.042	0.068
	<b>50</b>	0.051	0.052	0.051	0.048	0.048	0.057	0.045	0.039
<b>1-I</b>	<b>25</b>	0.319	0.335	0.219	0.059	0.365	0.435	0.255	0.058
	<b>50</b>	0.713	0.733	0.699	0.050	0.804	0.856	0.778	0.049
<b>2-I</b>	<b>25</b>	0.793	0.780	0.772	0.711	0.875	0.889	0.815	0.913
	<b>50</b>	0.980	0.980	0.980	0.938	1.000	1.000	1.000	1.000
<b>3-I</b>	<b>25</b>	0.576	0.578	0.500	0.410	0.645	0.682	0.528	0.585
	<b>50</b>	0.874	0.873	0.874	0.688	0.985	0.990	0.980	0.877

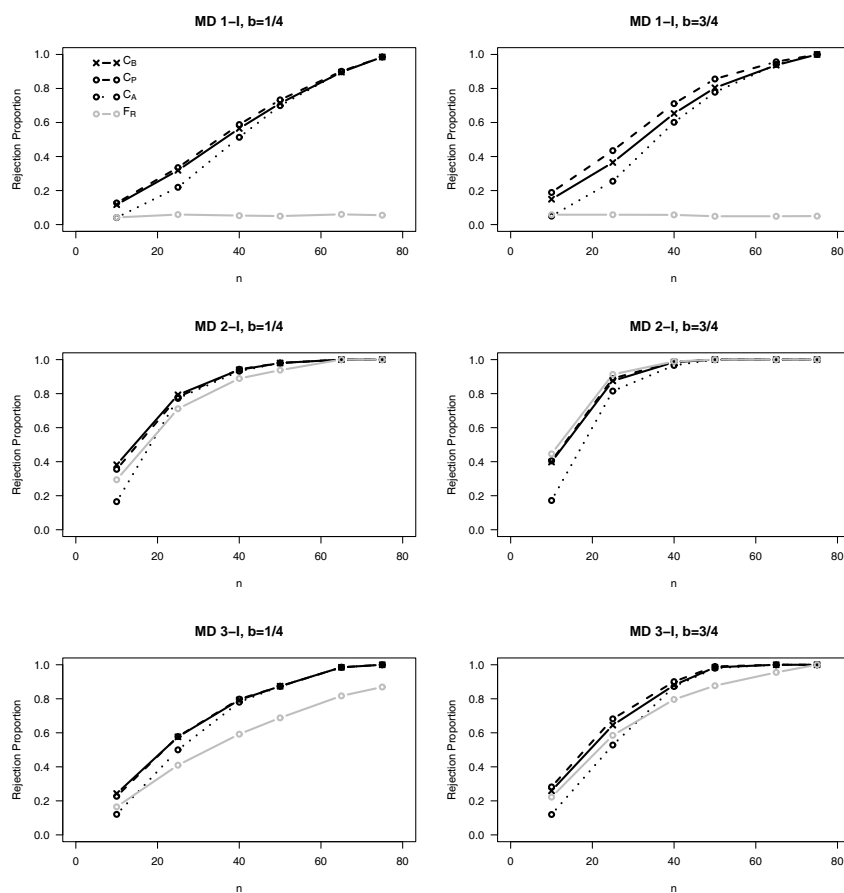


FIGURE 2: Observed rejection probabilities ( $\alpha = 0.05$ ) for the three different considered approximations of the Crámer-von Mises statistic distribution and the Friedman test against sample size for the symmetrical (type I) models.

TABLE 3: Observed rejection probabilities for type II (asymmetrical ones) models. The nominal level  $\alpha = 0.05$ .

		$b = 1/4$				$b = 3/4$			
	$n$	$C_B$	$C_P$	$C_A$	$F_R$	$C_B$	$C_P$	$C_A$	$F_R$
<b>0-II</b>	25	0.052	0.058	0.030	0.058	0.056	0.062	0.032	0.054
	50	0.061	0.063	0.057	0.058	0.054	0.060	0.051	0.054
<b>1-II</b>	25	0.293	0.295	0.197	0.106	0.312	0.304	0.207	0.108
	50	0.632	0.628	0.603	0.163	0.619	0.625	0.592	0.159
<b>2-II</b>	25	0.902	0.904	0.872	0.772	1.000	1.000	1.000	0.999
	50	0.932	0.931	0.906	0.845	1.000	1.000	1.000	1.000
<b>3-II</b>	25	0.726	0.738	0.624	0.847	0.737	0.752	0.632	0.875
	50	0.985	0.985	0.979	1.000	1.000	1.000	0.995	1.000

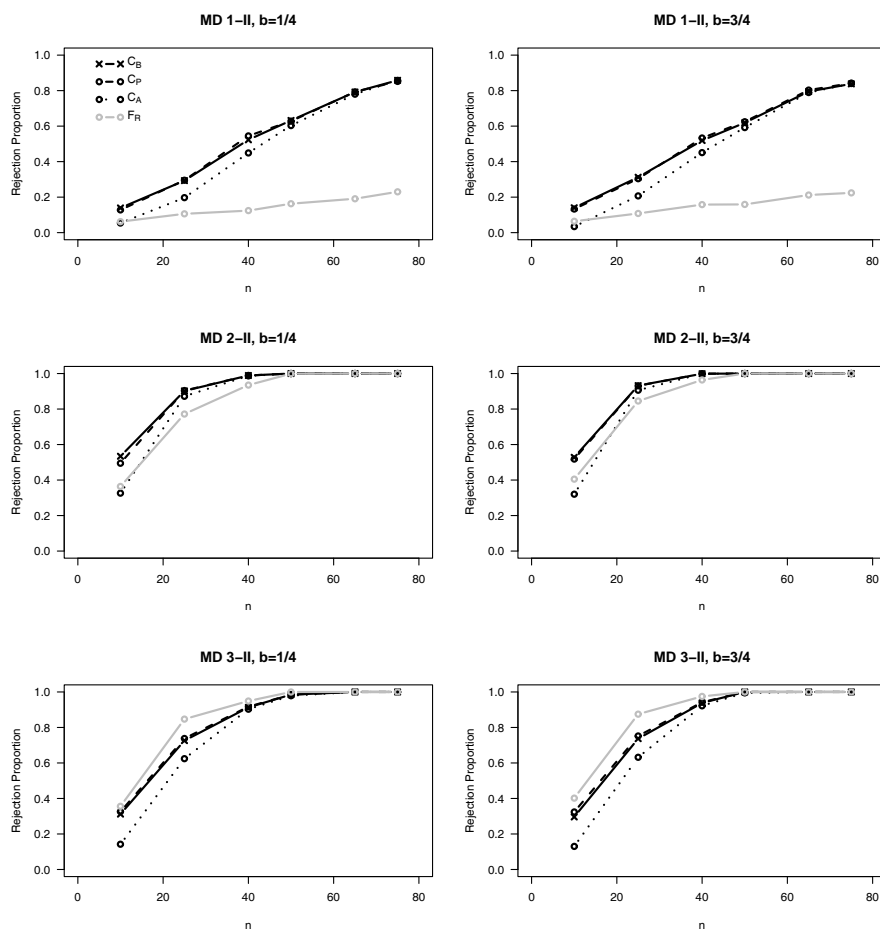


FIGURE 3: Observed rejection probabilities ( $\alpha = 0.05$ ) for the three different considered approximations of the Crámer-von Mises statistic distribution and the Friedman test against sample size for the asymmetrical (type II) models.

## 4. Inequality Incomes Analysis

In order to illustrate the practical performance of the proposed method we considered the study of the inequality incomes between thirty European countries. The inequality measure is a complex problem which has been addressed from different approaches (see Cowel (2009) and references therein). Although the Gini index is, probably, the most popular measure of inequality, other approaches have also been considered (see, for instance, Martínez-Cambor 2007). Our objective is to study the (possible) changes in the income distribution inequalities in Europe. With this goal the GDP per capita in PPS (quote from the site <http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&plugin=1&language=en&pcode=tec00114>: *Gross Domestic Product (GDP) is a measure for the economic activity. It is defined as the value of all goods and services produced less the value of any goods or services used in their creation. The volume index of GDP per capita in Purchasing Power Standards (PPS) is expressed in relation to the European Union (EU-27) average set to equal 100. If the index of a country is higher than 100, this country's level of GDP per head is higher than the EU average and vice versa. Basic figures are expressed in PPS, i.e. a common currency that eliminates the differences in price levels between countries allowing meaningful volume comparisons of GDP between countries. Please note that the index, calculated from PPS figures and expressed with respect to EU27 = 100, is intended for cross-country comparisons rather than for temporal comparisons.*) in thirty European countries in the years 2000 and 2011 have been collected (downloaded from the above website). Due to our objective is not to study the incomes distribution but the inequalities of these incomes, we have considered the relative GDP per capita in PPS distribution i.e., the considered variable are 100 times the original values divided by the European Union one (considering the currently twenty-seven countries members) and the particular mean has been subtracted. Figure 4 depicts the empirical cumulative distribution function (ECDF) and the density estimation function for the considered GDP transformations.

The value of the Cramér-von Mises statistics between these two distributions was 0.171. The approximate  $P$ -values were 0.012, 0.005 and 0.001 from the asymptotic, bootstrap and permutation algorithms (based on 10,000 replications), respectively. All of them reject the null and it can be concluded the inequality of the incomes does not be equal in 2000 and 2011. The Gini indices were 0.251 and 0.220, respectively.

## 5. Main Conclusions

The Cramér-von Mises criterion is widely used in order to compare cumulative distribution functions. Despite of different situations have been considered, independent  $k$ -sample comparison is the most studied problem. We propose the use of this criterion in a typical  $k$ -related sample design. By using the Donsker invariance principle and the Karhunen-Loève decomposition for stochastic Gaussian processes its asymptotic distribution is developed. Although its explicit

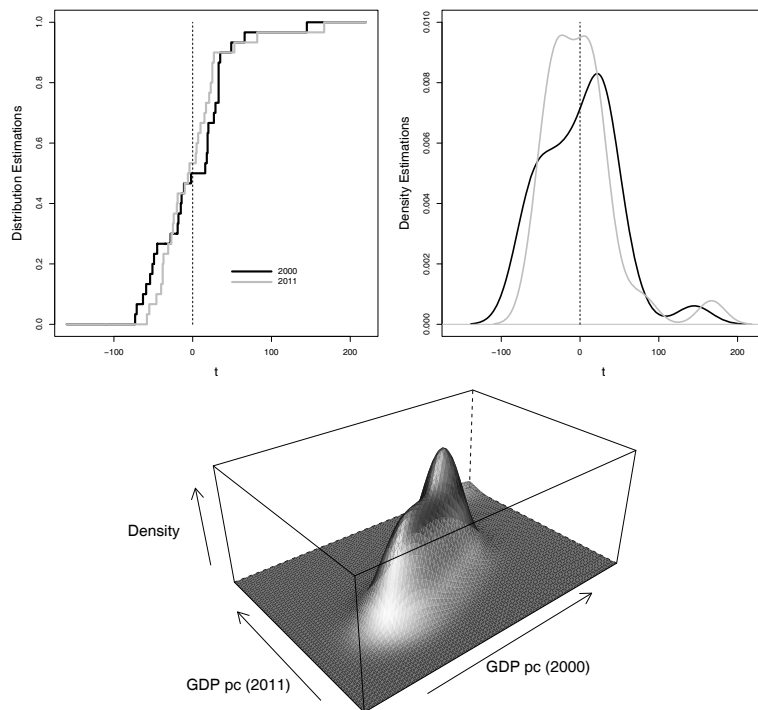


FIGURE 4: Upper, distribution (left) and density (right) estimation functions for the relative GDP per capita in PPS in thirty European countries in the years 2000 (black) and 2011 (gray). Lower, bivariate density estimation for the GDP per capita in PPS in years 2000 and 2011.

asymptotic distribution is still unknown, the obtained results allow to develop an useful approximation. As usual, we also explore two different resampling approximations: the classical and well-known permutation test and the most recent general bootstrap algorithm (gBA).

For independent samples, the Cramér-von Mises statistic is underlying distribution-free, its distribution does not depend on the distribution function where the samples were drawn, and it can be tabulated in order to obtain the  $P$ -value for a particular problem. In a paired sample design, the statistic distribution depends on the relationships among the involved variables; this relationship always must be estimated from the sample (therefore, universal eigenvalues do not exist for this topic), increasing the necessary time to compute the given asymptotic approximation. This is the main handicap for using the asymptotic approximation which, in general, obtains good results for moderate sample sizes.

A general bootstrap algorithm (gBA) and the usual permutation method are also studied. The considered bootstrap procedure exploits a particular pivotal function and introduce the null hypothesis at the moment of computing the value of the (bootstrap) statistic instead of in the random bootstrap samples generation process. The main advantage is that the data structure is preserved and

no additional assumptions (only the null) are required. Some details of its consistency are also reported, however reader is referred to Martínez-Cambor et al. (2012) for more details. This technique has already been used with success in a paired sample extension of the  $\mathcal{AC}$ -statistic (Martínez-Cambor 2010) and in inference on a particular ecological diversity index (Martínez-Cambor, Corral & Vicente 2011). In an extreme example is showed how the permutation method can lead to mistakes when the interchangeability assumption is violated, which is the usual situation when  $k > 2$ . However, the observed statistical power in our simulation study is similar for the three different considered methods. We must remark that the asymptotic and the bootstrap method avoid the exchangeability assumption (Von Mises 1991, Nelsen 2007) and do not increase the methodology complexity.

As in the independent case (see, for example Martínez-Cambor & Uña-Álvarez 2009), the simulation study results suggest that the Cramér-von Mises criterion is clearly better than the (classical) Friedman test when the main difference among the curves is not the location parameter and it obtains very good results otherwise. On the other hand, the proposed asymptotic approximation obtains similar statistical power than the ones based on resampling for moderate sample sizes. Relevant differences between two considered variance-covariance matrix structures ( $b = 1/4$  and  $b = 3/4$ ) have not been observed.

We think that the considered practical case is specially good in order to illustrate the use of the proposed methodology. When the focus is not the location parameter but the shape, which is the case of the inequality, the Cramér-von Mises statistics conventionally obtains good powers in order to check the equality of the involved distribution functions. In this context, traditional text like Friedman or the Student T-test do not work but the Cramér-von Mises criterion has proved that is a powerful test and a valuable tool for this kind of goals.

Part of the results provided in this manuscript are overlapped with the ones obtained in Quessy & Éthier (2012). However, the present work has been developed independent and previously to the publication of the Quessy and Éthier one. The main differences between the works are:

- (a) Our approach is more practical and, from our point of view more easy to understand for non probabilistic readers.
- (b) The permutation method is considered and discussed. A pathological case where this method fails has been provided.
- (c) A practical use of the gBA and a simulation study where the quality of the provided approximations can be checked.
- (d) The considered practical problem illustrates a situation where the equality among the location parameters is not the hypothesis to be tested.

## Acknowledgements

As usual, we are grateful with Susana Díaz Coto for the final manuscript revision. First author was supported by the research Grant MTM2011-23204 of the Spanish Ministerio de Ciencia e Innovación (FEDER support included).

[Recibido: septiembre de 2013 — Aceptado: diciembre de 2013]

## References

- Alkarni, S. H. & Siddiqui, M. M. (2001), ‘An upper bound for the distribution function of a positive definite quadratic form’, *Journal of Statistical Computation and Simulation* **69**(1), 51–56.
- Arcones, M. A. & Gine, E. (1992), ‘On the bootstrap of  $u$  and  $v$  statistics’, *Annals of Statistics* **20**(2), 655–674.
- Beran, R. (1982), ‘Estimated sampling distributions: The bootstrap and competitors’, *Annals of Statistics* **10**, 212–225.
- Ciba-Geigy, L. S. & Olsson, B. (1982), ‘A nearly distribution-free test for comparing dispersion in paired samples’, *Biometrika* **69**(2), 484–485.
- Cowel, F. A. (2009), Measuring inequality. Accessed 16/04/2013.  
\*[http://darp.lse.ac.uk/papersdb/cowell\\_measuringinequality3.pdf](http://darp.lse.ac.uk/papersdb/cowell_measuringinequality3.pdf)
- Cramér, H. (1928), ‘On the composition of elementary errors: II statistical applications’, *Skandinavisk Aktuarietidskrift* **11**, 141–180.
- Csőrgő, S. & Faraway, J. J. (1996), ‘The exact and asymptotic distributions of Cramér-von Mises statistics’, *Journals of the Royal Statistical Society B* **58**(1), 221–234.
- Deheuvels, P. (2005), ‘Weighted multivariate Cramér-von Mises-type statistics’, *Afrika Statistika* **1**(1), 1–14.
- Efron, B. (1979), ‘Bootstrap methods: Another look at the jackknife’, *Annals of Statistics* **7**, 1–26.
- Efron, B. (1982), The jackknife, the bootstrap and other resampling plans, Regional Conference Series in Applied Mathematics, CBMS-NSF.
- Freitag, G., Czado, D. & Munk, A. (2007), ‘A nonparametric test for similarity of marginals-with applications to the assessment of population bioequivalency’, *Journal of Statistical Planning & Inference* **137**, 697–711.
- Good, P. (2000), *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer Verlag, New York.



- Govindarajulu, Z. (1995), 'A class of asymptotically distribution free test procedures for equality of marginals under multivariate dependence', *American Journal of Mathematical and Management Sciences* **15**, 375–394.
- Govindarajulu, Z. (1997), 'A class of asymptotically distribution free tests for equality of marginals in multivariate populations', *Mathematical Methods of Statistics* **6**, 92–111.
- Hall, P. & Wilson, S. R. (1991), 'Two guidelines for bootstrap hypothesis testing', *Biometrics* **47**, 757–762.
- Lam, F. C. & Longnecker, M. T. (1983), 'Modified Wilcoxon rank sum test for paired data', *Biometrika* (70), 518–513.
- Martin, M. A. (2007), 'Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties', *Computational Statistics & Data Analysis* **51**, 6321–6342.
- Martínez-Camblor, P. (2007), 'Central limit theorems for S-Gini and Theil inequality coefficients', *Revista Colombiana de Estadística* **30**(2), 287–300.
- Martínez-Camblor, P. (2010), 'Nonparametric  $k$ -sample test based on kernel density estimator for paired design', *Computational Statistics & Data Analysis* **54**, 2035–2045.
- Martínez-Camblor, P. (2011), 'Testing the equality among distribution functions from independent and right censored samples via Cramér-von Mises criterion', *Journal of Applied Statistics* **38**(6), 1117–1131.
- Martínez-Camblor, P., Carelos, C. & Corral, N. (2012), 'Sobre el estadístico de Cramér-von Mises', *Revista de Matemáticas: Teoría y Aplicaciones* **19**, 89–101.
- Martínez-Camblor, P., Corral, N. & Vicente, D. (2011), 'Statistical comparison of the genetic sequence type diversity of invasive *Neisseria meningitidis* isolates in northern Spain (1997-2008)', *Ecological Informatics* **6**(6), 391–398.
- Martínez-Camblor, P. & Uña-Álvarez, J. (2009), 'Non-parametric  $k$ -sample tests: density functions vs distribution functions', *Computational Statistics & Data Analysis* **53**(9), 3344–3357.
- Munzel, U. (1999a), 'Linear rank score statistics when ties are present', *Statistics & Probability Letters* **41**, 389–395.
- Munzel, U. (1999b), 'Nonparametric methods for paired samples', *Statistica Neerlandica* **53**(3), 277–286.
- Nelsen, R. B. (2007), 'Extremes of non-exchangeability', *Statistical Papers* **48**, 329–336.

Podgor, M. J. & Gastwirth, J. L. (1996), Efficiency robust rank tests for stratified data, in E. Brunner & M. Denker, eds, 'Research Developments in Probability and Statistics', Festschrift in honor of Madan L. Puri. VSP International Science Publishers, Leiden, Netherlands.

Quessy, J. F. & Éthier, F. (2012), 'Cramér-von Mises and characteristic function tests for the two and  $k$ -sample problems with dependent data', *Computational Statistics and Data Analysis* **56**, 2097–2111.

Venkatraman, E. S. & Begg, C. B. (1996), 'A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment', *Biometrika* **83**(4), 835–848.

Von Mises, R. E. (1991), *Wahrscheinlichkeitsrechnung*, Deuticke, Vienna.

Westfall, P. H. & Young, S. S. (1993), *Resampling-Based Multiple Testing: Examples and Methods for  $p$ -value Adjustment*, Wiley, New York.

## Appendix

In this appendix we provide proofs for the two previously enunciated theorems. In particular, Theorem 1 is based on the well-known Donsker invariance principle and on classical Gaussian processes theory. In particular, we used the Karhunen-Loève decomposition in order to guarantee the existence of the necessary variables and coefficients. These coefficients values (eigenvalues) are not explicitly computed (the respective eigenfunctions are neither computed). These calculus are, in general, cumbersome and complex depending, in the present case, on the data covariance structure (therefore, they are different for each particular problem, universal coefficients do not exist). The following auxiliar result is, directly, derived from the Donsker invariance principle:

**Lemma 1.** *Let  $\boldsymbol{\xi}$  be a  $k$ -dimensional random vector and let  $\mathbf{X}$  be a random sample from  $\boldsymbol{\xi}$  (with size  $n$ ), by using the above notation, it is had the following weak convergence*

$$\sqrt{n}\{\hat{\mathbf{F}}_n(\mathbf{X}, \mathbf{u}) - \mathbf{F}(\mathbf{u})\} \xrightarrow{\mathcal{L}}_n \mathbf{Y}_{\mathbf{F}}(\mathbf{u})$$

where  $\mathbf{Y}_{\mathbf{F}}(\mathbf{u}) = (Y_{F_1^1}(u_1), \dots, Y_{F_k^k}(u_k))$  is a  $k$ -dimensional Gaussian process which follows a distribution  $\mathcal{N}_k(0, \boldsymbol{\Sigma}(\mathbf{u}))$  where

$$\boldsymbol{\Sigma}(\mathbf{u}) = \begin{pmatrix} \sigma_{1,1}(u_1, u_1) & \sigma_{1,2}(u_1, u_2) & \dots & \sigma_{1,k}(u_1, u_k) \\ \dots & \dots & \dots & \dots \\ \sigma_{k,1}(u_k, u_1) & \sigma_{k,2}(u_k, u_2) & \dots & \sigma_{k,k}(u_k, u_k) \end{pmatrix} \quad (5)$$

and  $\sigma_{i,j}(u_i, u_j) = F_{i,j}(u_i, u_j) - F_i(u_i)F_j(u_j)$  and  $F_{i,j}(u_i, u_j) = \mathcal{P}\{\xi_i \leq \square_i \cap \xi_j \leq \square_j\}$  for  $1 \leq i, j \leq k$ .

Furthermore, for  $u, v \in \mathbb{R}$  and  $1 \leq i, j \leq k$ ,  $\mathbb{E}[Y_{F_i}^i(u)Y_{F_j}^j(v)] = \sigma_{i,j}(u, v) = F_{i,j}(u, v) - F_i(u)F_j(v)$  where if  $(u \wedge v) = \min\{u, v\}$ ,

$$F_{i,j}(u, v) = \begin{cases} \mathcal{P}\{\xi \leq \sqcap \cap \xi \leq \sqsupset\} & \text{if } i \neq j \\ \mathcal{P}\{\xi \leq (\sqcap \wedge \sqsupset)\} & \text{if } i = j \end{cases}$$

Under the null given in (1) and if for each  $1 \leq i, j \leq k$  and for  $u, v \in \mathbb{R}$  we define the functions:  $\bar{F}_{\cdot,i}(u) = k^{-1} \sum_{j=1}^k F_{j,i}(u, u)$ ,  $\bar{F}_{\cdot,\cdot}(v) = k^{-1} \sum_{i=1}^k F_{\cdot,i}(v, v)$ ,  $\bar{F}_{\cdot,i}(u, v) = k^{-1} \sum_{j=1}^k F_{j,i}(u, v)$ ,  $\bar{F}_{\cdot,\cdot}(u, v) = k^{-1} \sum_{i=1}^k F_{\cdot,i}(u, v)$ ,  $C_i(u) = F(u) - 2\bar{F}_{\cdot,i}(u) + \bar{F}_{\cdot,\cdot}(u)$  and  $C_{i,j}(u, v) = F_{i,j}(u, v) - \bar{F}_{\cdot,i}(v, u) - \bar{F}_{\cdot,j}(u, v) + \bar{F}_{\cdot,\cdot}(u, v)$  we can obtain the following result,

**Theorem 3.** *By using the Lemma 1 notation, if  $F_1 = \dots = F_k (= F)$  (null hypothesis), it is held the (weak) convergence*

$$W_k^2(n) \xrightarrow{\mathcal{L}}_n \sum_{i=1}^k \sum_{l \in \mathbb{N}} \lambda_{i,l} M_{i,l}^2,$$

where  $\{\mathbf{M}_l = (M_{1,l}, \dots, M_{k,l})\}_{l \in \mathbb{N}}$  is a sequence of  $k$ -dimensional, normal distributed random variables whose marginals follow a  $\mathcal{N}(0, 1)$  distribution and  $\{\{\lambda_{i,l}\}_{l=1}^k\}_{l \in \mathbb{N}}$  are non negative constants satisfying  $\sum_{l \in \mathbb{N}} \lambda_{i,l}^2 < \infty$  for  $1 \leq i \leq k$ .

**Proof.** Keeping the Lemma 1 notation, if  $\bar{Y}_{F_{\cdot,\cdot}}(t) = k^{-1} \sum_{i=1}^k Y_{F_i}^i(t)$  and taking into account the well-known convergence  $\sup_{t \in \mathbb{R}} \{\hat{F}_n(X, t) - F(t)\} \xrightarrow{n} 0$  (a.s.). It is easy to see that

$$W_k^2(n) = \sum_{i=1}^k n \int \{\hat{F}_{n,i}(X_i, t) - \hat{F}_{n,\cdot}(X, t)\}^2 d\hat{F}_{n,\cdot}(X, t) \xrightarrow{\mathcal{L}}_n \sum_{i=1}^k \int \{Y_{F_i}^i(t) - \bar{Y}_{F_{\cdot,\cdot}}(t)\}^2 dF(t) \tag{6}$$

or equivalently,  $\forall u \in \mathbb{R}$ ,

$$\left\{ \mathcal{P} \left( \mathcal{W}_{\parallel}^{\epsilon}(\sqcup) \leq \sqcap \right) - \mathcal{P} \left( \sum_{\parallel}^{\parallel} \int \{\mathcal{Y}_{\mathcal{F}}^{\parallel}(\sqcup) - \bar{\mathcal{Y}}_{\mathcal{F},\cdot}(\sqcup)\}^{\epsilon} [\mathcal{F}(\sqcup) \leq \sqcap] \right) \right\} \xrightarrow{N} 0 \quad a.s.$$

On the other hand, if  $\mathcal{X}_{(\cdot)}(\sqcup) = \{\mathcal{Y}_{\mathcal{F}}^{\parallel}(\sqcup) - \bar{\mathcal{Y}}_{\mathcal{F},\cdot}(\sqcup)\}$  ( $1 \leq i \leq k$ ) then, under the null,

$$\mathcal{X}(t) = (\mathcal{X}_{(\infty)}(\sqcup), \dots, \mathcal{X}_{(\parallel)}(\sqcup))$$

is a centred  $k$ -dimensional Gaussian process. Moreover, if for  $t \in \mathbb{R}$ ,  $\mathbf{t} = (t, \dots, t)$  and  $\mathbf{\Sigma}(\mathbf{t})$  stands for the matrix defined in (2), for symmetry and under the null, for  $i \in 1, \dots, k$  it is obtained,

$$\begin{aligned} \mathbb{E}[\mathcal{X}_{(\cdot)}(\sqcup)^{\epsilon}] &= \mathbf{a}_i \mathbf{\Sigma}(\mathbf{t}) \mathbf{a}_i^t \\ &= \sigma_{i,i}(t, t) - \bar{\sigma}_{\cdot,i}(t) - \bar{\sigma}_{i,\cdot}(t) + \bar{\sigma}_{\cdot,\cdot}(t) = F(t) - 2\bar{F}_{\cdot,i}(t) + \bar{F}_{\cdot,\cdot}(t) = C_i(t) \end{aligned}$$

where for  $1 \leq i \leq k$ ,  $\mathbf{a}_i = (-1/k, \dots, (k-1)/k, \dots, -1/k)$ ,  $\bar{\sigma}_{\cdot, i}(t) = k^{-1} \sum_{j=1}^k \sigma_{j, i}(t, t)$ ,  $\bar{\sigma}_{i, \cdot}(t) = k^{-1} \sum_{j=1}^k \sigma_{i, j}(t, t)$  and  $\bar{\sigma}_{\cdot, \cdot}(t) = k^{-1} \sum_{i=1}^k \bar{\sigma}_{i, \cdot}(t, t)$ . In addition, it is had the covariance

$$\begin{aligned} \mathbb{C}_{i, j}(s, t) &= \mathbb{E}[\mathcal{X}_{(\cdot)}(f) \mathcal{X}_{(\cdot)}(\sqcup)] \\ &= \bar{F}_{i, j}(s, t) - \bar{F}_{\cdot, i}(s, t) - \bar{F}_{i, \cdot}(s, t) + \bar{F}_{\cdot, \cdot}(s, t) = C_{i, j}(s, t) \quad (1 \leq i, j \leq k) \end{aligned}$$

and it is easy to check that, for  $i \in 1, \dots, k$ ,

$$\iint \mathbb{C}_{i, i}(s, t)^2 dF(s) dF(t) < \infty$$

This property allows to apply the Karhunen-Loève decomposition (see, for example, Adler; 1990) in order to obtain the representation

$$\mathcal{X}_{(\cdot)}(\sqcup) = \sum_{\downarrow \in \mathbb{N}} \sqrt{\lambda_{i, \downarrow}} \downarrow_{i, \downarrow}(\sqcup) \mathcal{M}_{i, \downarrow} \quad (\text{for each } i \in 1, \dots, k) \quad (7)$$

where (for each  $i \in 1, \dots, k$ )  $\{e_{i, l}(t)\}_{l \in \mathbb{N}}$  is a *convergent orthonormal sequence* (also known as *eigenfunctions*) i.e.,

$$\int e_{i, p}(t) e_{i, q}(t) dF(t) = \begin{cases} 0 & \text{if } p \neq q \\ 1 & \text{if } p = q \end{cases}$$

$\{\mathbf{M}_l = (M_{1, l}, \dots, M_{k, l})\}_{l \in \mathbb{N}}$  are  $k$ -dimensional random vectors which marginal distributions follow a  $\mathcal{N}(0, 1)$  and, for  $1 \leq i \leq k$   $\{\{\lambda_{i, j}\}_{j=1}^k\}_{j \in \mathbb{N}}$ , are non negative constants (also known as *eigenvalues*) satisfying

$$\lambda_{i, 1} \geq \dots \geq \lambda_{i, l} \geq \dots \geq 0 \quad \forall i \in 1, \dots, k.$$

From (7), it is straightforward that, for  $i \in 1, \dots, k$ ,

$$\begin{aligned} \int \mathcal{X}_{(\cdot)}(\sqcup)^{\epsilon} \uparrow \mathcal{F}(\sqcup) &= \int \left( \sum_{l \in \mathbb{N}} \sqrt{\lambda_{i, l}} e_{i, l}(t) M_{i, l} \right)^2 dF(t) \\ &= \int \sum_{l \in \mathbb{N}} \sum_{j \in \mathbb{N}} \sqrt{\lambda_{i, l}} \sqrt{\lambda_{i, j}} e_{i, l}(t) e_{i, j}(t) M_{i, l} M_{i, j} dF(t) \\ &= \sum_{l \in \mathbb{N}} \sum_{j \in \mathbb{N}} \sqrt{\lambda_{i, l}} \sqrt{\lambda_{i, j}} M_{i, l} M_{i, j} \int e_{i, l}(t) e_{i, j}(t) dF(t) \quad (8) \\ &= \sum_{l \in \mathbb{N}} \lambda_{i, l} M_{i, l}^2 \quad (9) \end{aligned}$$

Therefore, from the Fubini Theorem, for  $i \in 1, \dots, k$ ,

$$\begin{aligned} \sum_{l \in \mathbb{N}} \lambda_{i, l} &= \mathbb{E} \left( \int \mathcal{X}_{(\cdot)}(\sqcup)^{\epsilon} \uparrow \mathcal{F}(\sqcup) \right) = \int \mathbb{E} (\mathcal{X}_{(\cdot)}(\sqcup)^{\epsilon}) dF(t) \\ &= \int C_i(t) dF(t) = C_i \end{aligned} \quad (10)$$

and

$$\sum_{l \in \mathbb{N}} \lambda_{i,l}^2 = \iint \mathbb{C}_{i,i}(s,t)^2 dF(s)dF(t) < \infty \tag{11}$$

Now, we are interested in studying the joint distribution of  $\mathbf{M}_l$  (for each fixed  $l \in \mathbb{N}$ ). We will prove that  $\sum_{i=1}^k a_i M_{i,l}$  follows a normal distribution for each  $a_1, \dots, a_k \in \mathbb{R}$  and for  $l \in \mathbb{N}$ . Note that, for each (fixed)  $l \in \mathbb{N}$ , we have that

$$\mathbf{X}^*(t) = (a_1 \mathcal{X}_{(\infty)}(\sqcup) \rfloor_{\infty, \updownarrow}(\sqcup), \dots, \mathcal{X}_{(\parallel)}(\sqcup) \rfloor_{\parallel, \updownarrow}(\sqcup))$$

is a  $k$ -dimensional centred *Gaussian* process. From (7), for each  $i \in 1, \dots, k$

$$a_i e_{i,l}(t) \sum_{j \in \mathbb{N}} \sqrt{\lambda_{i,l}} e_{i,j}(t) M_{i,j} = a_i \mathcal{X}_{(\cdot)}(\sqcup) \rfloor_{\cdot, \updownarrow}(\sqcup)$$

hence,

$$a_i \sum_{j \in \mathbb{N}} \sqrt{\lambda_{i,j}} M_{i,j} \int e_{i,l}(t) e_{i,j}(t) dF(t) = a_i \int \mathcal{X}_{(\cdot)}(\sqcup) \rfloor_{\cdot, \updownarrow}(\sqcup) \rfloor \mathcal{F}(\sqcup)$$

We can assume that  $\lambda_{i,l} \neq 0$  for  $1 \leq i \leq k$  (if some  $\lambda_{i,l} = 0$  ( $1 \leq i \leq k$ ),  $M_{i,l}$  does not interfere in any definition and we would have freedom to select it independently with the other ones) hence, from the eigenfunctions properties

$$\sum_{i=1}^k a_i M_{i,l} = \sum_{i=1}^k \frac{a_i}{\sqrt{\lambda_{i,l}}} \int \mathcal{X}_{(\cdot)}(\sqcup) \rfloor_{\cdot, \updownarrow}(\sqcup) \rfloor \mathcal{F}(\sqcup) = \int \sum_{\cdot=\infty}^{\parallel} \frac{\cdot \rfloor}{\sqrt{\lambda_{\cdot, \updownarrow}}} \mathcal{X}_{(\cdot)}(\sqcup) \rfloor_{\cdot, \updownarrow}(\sqcup) \rfloor \mathcal{F}(\sqcup)$$

follows a normal distribution.

From (6), (7) and (8) we know that there exists a sequence of  $k$ -dimensional normal distributed random variables,  $\{\mathbf{M}_j\}_{j \in \mathbb{N}}$  whose marginals follow a  $\mathcal{N}(0, 1)$  distribution and non negative constants  $\{\{\lambda_{i,j}\}_{i=1}^k\}_{j \in \mathbb{N}}$  satisfying (9) and (10), such that

$$W_k^2(n) \xrightarrow{\mathcal{L}}_n \sum_{i=1}^k \sum_{j \in \mathbb{N}} \lambda_{i,j} M_{i,j}^2$$

and the proof is concluded. □

**Theorem 4.** *Under the Lemma 1 assumptions and by using the same notation. Let  $\mathbf{X}^* = (X_1^*, \dots, X_k^*)$  be an independent random sample generated from  $\hat{F}_n(\mathbf{X}, \cdot)$  (multivariate ECDF referred to the random sample  $\mathbf{X}$ ). If*

$$\begin{aligned} W_k^{2,*}(n) &= \sum_{i=1}^k n \int \{\hat{F}_{n,i}^*(X_i^*, t) - \hat{F}_{n,i}(X_i, t)\}^2 d\hat{F}_{n,\bullet}^*(\mathbf{X}^*, t) \\ &\quad - nk \int \{\hat{F}_{n,\bullet}^*(\mathbf{X}^*, t) - \hat{F}_{n,\bullet}(\mathbf{X}, t)\}^2 d\hat{F}_{n,\bullet}^*(\mathbf{X}^*, t) \end{aligned}$$

where for each  $i \in 1, \dots, k$ ,  $\hat{F}_{n,i}^*(X_i^*, t)$  is the ECDF referred to  $X_i^*$  and  $\hat{F}_{n,\bullet}^*(\mathbf{X}^*, t) = k^{-1} \sum_{i=1}^k \hat{F}_{n,i}^*(X_i^*, t)$ . Under the null, it is derived,

$$\left\{ \mathcal{P}_{\mathbf{X}} \left( \mathcal{W}_{\parallel}^{\infty,*}(\setminus) \leq \square \right) - \mathcal{P} \left( \mathcal{W}_{\parallel}^{\infty}(\setminus) \leq \square \right) \right\} \rightarrow_n 0 \quad a.s.$$

where  $\mathcal{P}_{\mathbf{X}}$  denotes probability conditionally on sample  $\mathbf{X}$ .

**Proof.** It is easy to check that,

$$\begin{aligned} W_k^2(n) &= \sum_{i=1}^k n \int \{ \hat{F}_{n,i}(X_i, t) - \hat{F}_{n,\bullet}(\mathbf{X}, t) \}^2 d\hat{F}_{n,\bullet}(\mathbf{X}, t) \\ &= \sum_{i=1}^k n \int \{ \hat{F}_{n,i}(X_i, t) - F_i(t) \}^2 d\hat{F}_{n,\bullet}(\mathbf{X}, t) - nk \int \{ \hat{F}_{n,\bullet}(\mathbf{X}, t) - F(t) \}^2 d\hat{F}_{n,\bullet}(\mathbf{X}, t). \end{aligned}$$

And, directly from the Lemma 1,

$$W_k^2(n) \xrightarrow{\mathcal{L}}_n \sum_{i=1}^k \int Y_{\hat{F}_i}^i(t)^2 dF(t) - k \int \bar{Y}_{F,\bullet}(t)^2 dF(t)$$

Of course, the above equation is equivalent to the one in (6). Also from the Lemma 1 and classical theory of stochastic processes (in particular, Horváth and Steinebach (1999) proved that the expressions  $\sup_{t \in \mathbb{R}} |\hat{F}_n(X, t) - F(t)|$  and  $\sup_{t \in \mathbb{R}} |\hat{F}_n^*(X^*, t) - \hat{F}_n(X, t)|$  where  $X^*$  is an independent random sample with size  $n$  generated from  $\hat{F}_n(X, \cdot)$  (ECDF referred to the random sample  $X$  which sample size is  $n$ ) have the same asymptotic distribution), for each  $u \in \mathbb{R}$ , it is had the convergence,

$$\left\{ \mathcal{P}_{\mathbf{X}} \left( \mathcal{W}_{\parallel}^{\infty,*}(\setminus) \leq \square \right) - \mathcal{P} \left( \sum_{\setminus=\infty}^{\parallel} \mathcal{I}_{\setminus} \left( \mathcal{Y}_{\hat{F}_{\setminus,i}}^i, \mathcal{Y}_{\hat{F}_{\setminus,\bullet}}(\setminus) \right) \leq \square \right) \right\} \rightarrow_n 0 \quad a.s.$$

where  $Y_{\hat{F}_{n,i}}^i$  ( $1 \leq i \leq k$ ) and  $\bar{Y}_{\hat{F}_{n,\bullet}}$  are the processes which appear in the Lemma 2.1 and, for  $i \in 1, \dots, k$ ,

$$I_n \left( Y_{\hat{F}_{n,i}}^i, Y_{\hat{F}_{n,\bullet}}(t) \right) = \int Y_{\hat{F}_{n,i}}^i(t)^2 d\hat{F}_{n,\bullet}(\mathbf{X}, t) - k \int \bar{Y}_{\hat{F}_{n,\bullet}}(t)^2 d\hat{F}_{n,\bullet}(\mathbf{X}, t)$$

Due, under the null hypothesis,  $\forall t \in \mathbb{R}$ , it is had the convergence  $\hat{F}_n(\mathbf{X}, t) \rightarrow_n F(t)$  (a.s.), for each  $u \in \mathbb{R}$ , it is directly derived that

$$\left\{ \mathcal{P}_{\mathbf{X}} \left( \mathcal{W}_{\parallel}^{\infty,*}(\setminus) \leq \square \right) - \mathcal{P} \left( \mathcal{W}_{\parallel}^{\infty}(\setminus) \leq \square \right) \right\} \rightarrow_n 0 \quad a.s.$$

□