THE TWO-SAMPLE PROBLEM IN $\mathbb{R}^m$ AND MEASURE-VALUED MARTINGALES

J.H.J. EINMAHL[1] AND E.V. KHMALADZE[2]

Eindhoven University of Technology and Eurandom, and University of New South Wales and A. Razmadze Mathematical Institute

The so-called two-sample problem is one of the classical problems in mathematical statistics. It is well-known that in dimension one the two-sample Smirnov test possesses two basic properties: it is distribution free under the null hypothesis and it is sensitive to 'all' alternatives. In the multidimensional case, i.e. when the observations in the two samples are random vectors in $\mathbb{R}^m$, $m \geq 2$, the Smirnov test loses its first basic property. In correspondence with the above, we define a solution of the two-sample problem to be a 'natural' stochastic process, based on the two samples, which is (α) asymptotically distribution free under the null hypothesis, and which is, intuitively speaking, (β) as sensitive as possible to all alternatives. Despite the fact that the two-sample problem has a long and very diverse history, starting with some famous papers in the thirties, the problem is essentially still open for samples in $\mathbb{R}^m, m \geq 2$. In this paper we present an approach based on measure-valued martingales and we will show that the stochastic process obtained with this approach is a solution to the two-sample problem, i.e. it has both the properties (α) and (β), for any $m \in \mathbb{N}$.

AMS subject classifications: 62G10, 62G20, 62G30; secondary 60F05, 60G15, 60G48.
Keywords and phrases: Dirichlet (Voronoi) tessellation, distribution free process, empirical process, measure-valued martingale, non-parametric test, permutation test, two-sample problem, VC class, weak convergence, Wiener process.

1 Introduction

Suppose we are given two samples, that is, two independent sequences $\{X_i'\}_1^{n_1}$ and $\{X_i''\}_1^{n_2}$ of i.i.d. random variables taking values in m-dimensional Euclidean space $\mathbb{R}^m, m \geq 1$. Denote with P_1 and P_2 the probability distributions of each of the X_i' and X_i'' and write $\widehat{P}_{n_1}$ and P_n for the empirical distributions of the first sample and of the pooled sample $\{X_i'\}_1^{n_1} \cup \{X_i''\}_1^{n_2}$ respectively, i.e.

$$(1.1) \qquad \widehat{P}_{n_1}(B) = \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbb{1}_B(X_i'),$$
$$P_n(B) = \frac{1}{n}\left(\sum_{i=1}^{n_1} \mathbb{1}_B(X_i' i) + \sum_{i=1}^{n_2} \mathbb{1}_B(X_i'')\right), \quad n = n_1 + n_2,$$

[1]Research partially supported by European Union HCM grant ERB CHRX-CT 940693.
[2]Research partially supported by the Netherlands Organization for Scientific Research (NWO) while the author was visiting the Eindhoven University of Technology, and partially by the International Science Foundation (ISF), Grant MXI200.

where B is a measurable set in $\mathbb{R}^m$ and $\mathbb{1}_B$ is its indicator function. Consider the difference

$$(1.2)\qquad v_n(B) = \left(\frac{n_1 n}{n_2}\right)^{\frac{1}{2}} (\widehat{P}_{n_1}(B) - P_n(B)), \quad B \in \mathcal{B},$$

and call the random measure $v_n(\cdot)$ the (classical) two-sample empirical process with the indexing class $\mathcal{B}$. Throughout we avoid the double index (n_1, n_2); this can be done without any ambiguity letting $n_1 = n_1(n)$ and $n_2 = n_2(n)$. We will always assume $n_1, n_2 \to \infty$ as $n \to \infty$. The indexing class $\mathcal{B}$ is important for functional weak convergence of v_n and will be specified in Sections 3–5.

The problem of testing the null hypothesis $H_0 : P_1 = P_2$, called 'the two-sample problem', is one of the classical problems of statistics. The literature on the two-sample problem is enormous. In here we are able to mention only very few of the papers on the subject, namely those in direct relation to the aims of the present work. The specific feature of the two-sample problem is that the under H_0 presumed common distribution $P(= P_1 = P_2)$ remains unspecified and can be any within some typically very large class $\mathcal{P}$. Hence, it is important to have some supply of test statistics such that their null distributions, at least asymptotically as $n \to \infty$, are independent of this common distribution $P \in \mathcal{P}$. Such statistics are called asymptotically distribution free.

The classical solution of the two-sample problem when the dimension $m = 1$ is associated with Smirnov (1939) where first the two-sample empirical process

$$(1.3)\qquad v_n(x) = \left(\frac{n_1 n}{n_2}\right)^{1/2} (\widehat{F}_{n_1}(x) - F_n(x)), \quad x \in \mathbb{R}^1,$$

was introduced, where $\widehat{F}_{n_1}$ and F_n stand for the empirical distribution functions of the first and the pooled sample respectively, and the limiting distribution of its supremum was derived. This limiting distribution was shown to be free from P provided $P \in \mathcal{P}_c$, the class of all distributions on $\mathbb{R}^1$ with a continuous distribution function. This classical statement was an early reflection of the now well-known fact that the process

$$(1.4)\qquad v_n \circ F_n^{-1}(t), \quad t \in [0, 1],$$

converges in distribution, for all $P \in \mathcal{P}_c$, to a standard Brownian bridge v (see, e.g., Shorack and Wellner (1986)). Then, for a large collection of functionals φ, the statistics $\varphi(v_n \circ F_n^{-1})$ converge in distribution to $\varphi(v)$ and hence are asymptotically distribution free. Recently Urinov (1992) considered another version of the two-sample empirical process in $\mathbb{R}^1$:

$$(1.5) \qquad M_n(x) = \left(\frac{n_1 n}{n_2}\right)^{1/2} \left(\widehat{F}_{n_1}(x) - \int_{-\infty}^{x} \frac{1 - \widehat{F}_{n_1}(y-)}{1 - F_n(y-)} F_n(dy)\right),$$

($x \in I\!R^1$), and proved that the process $M_n \circ F_n^{-1}$ is also asymptotically distribution free: for all $P \in \mathcal{P}_c$ it converges to a standard Brownian motion W on $[0,1]$.

The convergence in distribution of the process (1.3) when $x \in I\!R^m, m \geq 2$, was studied in Bickel (1969). Though asymptotically distribution free processes or statistics were not obtained in that paper, the general approach was well-motivated. Namely, to obtain an asymptotically correct approximation for the distribution of statistics based on v_n, like, for example, the Smirnov statistic $\sup_{x \in I\!R^m} |v_n(x)|$, he studied the conditional distribution of v_n given F_n. This conditioning, also adopted in Urinov (1992), and being also a part of the approach of the present paper (see Sections 3 and 4), is motivated by the fact that, under H_0, one can construct the two-sample situation as follows. Let $\{X_i\}_1^n$ be a sample of size n from a distribution $P \in \mathcal{P}$. Let also $\{\delta_i\}_1^n$ be n Bernoulli random variables independent of $\{X_i\}_1^n$ and sampled without replacement from an urn containing n_1 'ones' and n_2 'zeros'. Now the set of X_i's with $\delta_i = 1$ is called the first sample and those with $\delta_i = 0$ is called the second sample. Any permutation of $\{(X_i, \delta_i)\}_1^n$ independent of $\{\delta_i\}_1^n$ will not alter the distribution of $\{\delta_i\}_1^n$. Hence, for statistics $\varphi(\{X_i\}_1^n, \{\delta_i\}_1^n)$ their conditional distribution given F_n is induced by a distribution free from P.

Actually, this is the basic approach of all permutation tests and dates back at least as far as Fisher (1936) and Wald and Wolfowitz (1944). Well-known permutation tests for the multivariate two-sample (and multi-sample) problem were developed in the mid-60's (see, e.g., Chatterjee and Sen (1964) and Puri and Sen (1966, 1969)). It should be noted, however, that most of the permutation tests are based on asymptotically linear in $\{\delta_i\}_1^n$, and hence asymptotically normal, statistics. To essentially nonlinear statistics, like the Smirnov statistic, this approach was first applied in Bickel (1969), to the best of our knowledge.

There are several other methods for obtaining statistically important versions or substitutes of the two-sample empirical process, see, e.g., Friedman and Rafsky (1979), Bickel and Breiman (1983), Kim and Foutz (1987), and Henze (1988) for interesting approaches.

Though we just discussed the two-sample problem and its solution, the precise mathematical formulation of the problem has not been given yet. The requirement of asymptotically distribution freeness can not be sufficient to formulate the problem for it can be trivially satisfied. Another condition on 'sensitivity' towards alternatives must be also imposed.

In this paper we propose two related formulations of the problem (Section 2), one of them imposes quite strong requirements. Then in Section 3 we construct a (signed-)measure-valued martingale M_n, which is a generalization of the process (1.5) of Urinov (1992), and its renormalized versions u_n and w_n. We prove limit theorems for the asymptotically distribution free modifications u_n and w_n as well as for M_n, both under the null hypothesis (Section 4) and under alternatives (Section 5) and show that under natural conditions u_n and w_n are solutions of the two-sample problem.

2 General notations; some preliminaries; formulation of the two-sample problem

As we remarked in the Introduction, in the classical two-sample problem in $\mathbb{R}^1$ it is required that under H_0 the common distribution P belongs to the class $\mathcal{P}_c$ of distributions having continuous distribution functions, and for this class of P's, the Smirnov process $v_n \circ F_n^{-1}$ and the Urinov process $M_n \circ F_n^{-1}$ are asymptotically distribution free. In $\mathbb{R}^m$, we also need some requirements under H_0. Let μ denote Lebesgue measure and let from now on $\mathcal{P}$ denote the class of all distributions P with the properties

(C1) $P([0,1]^m) = 1$;

(C2) $f = dP/d\mu > 0$ a.e. on $[0,1]^m$.

Condition (C1) is not an essential restriction since it can be satisfied in several ways. For example, if $Y_1, ..., Y_m$ denote the coordinates of some absolutely continuous m-dimensional random vector Y and if $G_1, ..., G_m$ are some absolutely continuous distribution functions on $\mathbb{R}$ such that the range of Y_k is contained in the support of $G_k, k = 1, ..., m$, then the random vector X with coordinates $X_k = G_k(Y_k)$, $k = 1, ..., m$, has an absolutely continuous distribution on $[0,1]^m$. Another, perhaps better, possibility is to reduce the pooled sample to the sequence $\{R_i\}_1^n$, where the coordinates of each R_i are the normalized coordinatewise ranks of the corresponding coordinates of the i-th observation. (Note that the thus obtained two-sample empirical process is equal to $v_n \circ (F_{1n}^{-1}, ..., F_{mn}^{-1})$, where $F_{jn}, j = 1, \ldots, m$, are the pooled marginal empirical distribution functions.) Though there is definitely no absolute continuity of the distribution of R_i, $i = 1, ..., n$, we will indicate below how the subsequent program can go through for these ranks (see e.g. Lemma 3.5). Condition (C2) represents a certain restriction. Observe, however, that the processes u_n and w_n, defined below, have limiting distributions which depend on P only through its support.

Besides the classical two-sample empirical process v_n there can be many other random measures which are also functionals of $\widehat{P}_{n_1}$ and P_n and could also be called two-sample empirical processes. We will obtain versions of such processes which will be asymptotically distribution free from $P \in \mathcal{P}$. It is also needed that such a process is sufficiently sensitive to possible alternatives. To formulate both requirements precisely we need to describe the class of alternatives. In fact, it will be the class of all compact contiguous alternatives to the two-sample null hypothesis. Here is the precise condition:

(C3) The distributions P_1 and P_2 of each of the X_i' and X_i'', respectively, depend on n and are, for each n, absolutely continuous w.r.t. some $P \in \mathcal{P}$, and the densities dP_j/dP, $j = 1, 2$, admit the following asymptotic representation

$$\left(\frac{dP_1}{dP}\right)^{\frac{1}{2}} = 1 + \frac{1}{2\sqrt{n_1}}\sqrt{1-p_0}h_{1n}, \quad \left(\frac{dP_2}{dP}\right)^{\frac{1}{2}} = 1 - \frac{1}{2\sqrt{n_2}}\sqrt{p_0}h_{2n}, \tag{2.1}$$

and $\int (h_{jn} - h)^2 dP \to 0$, $j = 1, 2$, for some h with $0 < ||h||^2 := \int h^2 dP < \infty$, while $n_1/n \to p_0 \in (0, 1)$.

The distribution of the pooled sample $\{X_i'\}_1^{n_1} \cup \{X_i''\}_1^{n_2}$ under P is certainly the n-fold direct product $P^n = P \times \cdots \times P$. It is well-known (Oosterhoff and van Zwet (1979)) that its distribution under the alternative (2.1), which is the direct product $P_1^{n_1} \times P_2^{n_2}$, is contiguous with respect to P^n, and that under P^n

$$L_n = \ln \frac{d(P_1^{n_1} \times P_2^{n_2})}{dP^n} \to_d N(-\tfrac{1}{2}||h||^2, ||h||^2), \tag{2.2}$$

with $N(\mu, \sigma^2)$ denoting a normal random variable with mean μ and variance σ^2. Hence, under $P_1^{n_1} \times P_2^{n_2}$

$$L_n \to_d N(\tfrac{1}{2}||h||^2, ||h||^2). \tag{2.3}$$

Note that, in (2.1), it could seem more natural to start with some functions h_{1n} and h_{2n} converging to h_1 and h_2, instead of converging both to h. However it can be shown that this general situation reduces to (2.1) as it stands, when we replace P by a strategically chosen new P, namely the one such that (P, P) is 'closest' to (P_1, P_2), where this closeness is measured in terms of the distance in variation between P^n and $P_1^{n_1} \times P_2^{n_2}$:

$$d(P_1^{n_1} \times P_2^{n_2}, P^n) = P_1^{n_1} \times P_2^{n_2}(L_n > 0) - P^n(L_n > 0), \tag{2.4}$$

a very proper distance in a statistical context. Indeed, it is clear that

$$d(P_1^{n_1} \times P_2^{n_2}, P^n) = \max_{0 \le \alpha \le 1} (\beta_n(\alpha) - \alpha),$$

where $\beta_n(\alpha)$ is the power of the α-level Neyman-Pearson test for P^n against $P_1^{n_1} \times P_2^{n_2}$. According to (2.2) and (2.3)

$$(2.5) \qquad d(P_1^{n_1} \times P_2^{n_2}, P^n) \to 2\Phi(\tfrac{1}{2}||h||) - 1 =: \lambda,$$

where Φ is the standard normal distribution function.

Now we are prepared to formulate what we mean with a solution of the two-sample problem. In words, we want a 'natural' process, based on the data, that converges in distribution to a limit process not depending on P, and that hence is asymptotically distribution free. Moreover, the distributions of the limiting process under null and contiguous alternative hypothesis should be as far apart in distance in variation as the limiting distance in variation under null and contiguous alternative hypothesis of the data themselves. So basically we want a multivariate process which has the same beautiful properties as the transformed univariate empirical process in (1.4). Here follows the precise mathematical formulation.

Let $\mathcal{B} \subset \mathcal{B}_0$, where $\mathcal{B}_0$ denotes the class of all Borel-measurable subsets of $[0,1]^m$, and consider a sequence of random measures $\{\xi_n\}_{n\geq 1}$ restricted to $\mathcal{B}$. The sequence $\{\xi_n\}_{n\geq 1}$ will be called *a strong $\mathcal{P}$-solution* of the two-sample problem, if there exists a measurable space $\mathcal{X}$ such that

(α) under P^n, for each $P \in \mathcal{P}$, $\xi_n \to_d \xi$ in $\mathcal{X}$ and the distribution $\mathbf{Q}_\xi$ of ξ is the same for all $P \in \mathcal{P}$;

(β) under $P_1^{n_1} \times P_2^{n_2}$, for each sequence of alternatives (2.1), $\xi_n \to_d \widetilde{\xi}$ in $\mathcal{X}$ and the distribution $\mathbf{Q}_{\widetilde{\xi}}$ of $\widetilde{\xi}$ is such that $d(\mathbf{Q}_{\widetilde{\xi}}, \mathbf{Q}_\xi) = \lambda$.

In order to obtain practically relevant solutions we add as in Khmaladze (1993) the heuristic requirement that the process ξ_n (and the subsequent test statistics) are simple enough to make computations feasible. In other words, we want to exclude formally correct solutions, that involve very 'irregular' transformations of the two-sample empirical process v_n, like, e.g., solutions obtained from bimeasurable bijections from $\mathbb{R}^m$ to $\mathbb{R}^1$. Any ξ_n satisfying (α), (β) and this informal requirement provides a proper background for producing two-sample tests. Indeed, not only for any *particular* sequences $\{h_{1n}\}_{n\geq 1}, \{h_{2n}\}_{n\geq 1}$ in (2.1), we can find a (linear) functional based on ξ_n such that it will lead to an asymptotically optimal test against this sequence of alternatives, but also a great variety of good omnibus tests can be constructed in the usual way, e.g., by taking (weighted) Cramér-von Mises-type statistics or (weighted) Kolmogorov-Smirnov-type statistics.

It might be convenient from computational or other points of view to sacrifice a bit of power in favour of, say, computational simplicity: the sequence $\{\xi_n\}_{n\geq 1}$ is called *a weak $\mathcal{P}$-solution* of the two-sample problem if it possesses property (α) and if

(γ) under $P_1^{n_1} \times P_2^{n_2}$, for each sequence of alternatives (2.1), $\xi_n \to_d \widetilde{\xi}$ in $\mathcal{X}$ and $d(\mathbf{Q}_{\widetilde{\xi}}, \mathbf{Q}_\xi) > 0$.

In the subsequent sections our choice of the space $\mathcal{X}$ will be the space $l_\infty(\mathcal{B})$. We will prove that the sequence of random measures $\{u_n\}_{n\geq 1}$ (see (3.15)) is a strong $\mathcal{P}$-solution, and we consider also a sequence of differently normalized random measures $\{w_n\}_{n\geq 1}$ (see (3.16)) and show that under natural assumptions it is a weak $\mathcal{P}$-solution.

In conclusion of this section we give a few remarks which may illuminate the possible nature of strong and weak solutions.

The first remark is that for an appropriate indexing class $\mathcal{B}$ (see Theorem 5.3) the classical two-sample empirical process v_n possesses property (β), though not property (α). When $m = 1$, however, the processes $v_n \circ F_n^{-1}$ in (1.4) and $M_n \circ F_n^{-1}$, with M_n as in (1.5), do satisfy (α) and (β), and hence are strong $\mathcal{P}_c$-solutions to the two-sample problem. For any $m \geq 1$, the process w_n below remains in one-to-one correspondence with v_n for each n (Lemma 3.1 and definition (3.16)) and therefore contains the same amount of 'information' as v_n for each finite n. However, as $n \to \infty$, some 'information' (though not much) is asymptotically 'slipping away' (Theorem 5.3 and following comments).

As the second remark we note one 'obvious' weak solution, which nevertheless is quite interesting for practical purposes: let $\zeta \sim \mathcal{N}(0,1)$ be independent from v_n (say, is generated by a computer programme) and consider $\xi_n(B) = v_n(B) + P_n(B)\zeta$. Since v_n converges to a P-Brownian bridge it is immediate that ξ_n converges to a P-Brownian motion under H_0. Then it can be renormalized exactly in the same way as u_n below (put $t = 1$ in (3.15)) and will become asymptotically distribution free, however, because of the randomization involved, ξ_n will loose property (β) though it will retain property (γ). Curiously enough, in many practical situations the loss is not big (Dzaparidze and Nikulin (1979)).

Finally remark, as shown in Schilling (1983), that the asymptotically distribution free process of Bickel and Breiman (1983) though very interesting from some other points of view can not detect (in a goodness-of-fit context) any of $1/\sqrt{n}$-alternatives. Whether the process of Kim and Foutz (1987) connected with the same initial idea of uniform spacings can detect such alternatives remains formally unclear. However we believe that the phenomena discovered in Chibisov (1961) explain why it may not be likely. For the omnibus statistic of Friedman and Rafsky (1979) the recent result of Henze and Penrose (1999), Theorem 2, leaves little hope that it can detect any of $1/\sqrt{n}$-alternatives. So, to the best of our knowledge, the two-sample problem, as described in this section, is essentially still open when $m \geq 2$.

3 Two-sample scanning martingales

The main object of this section if not of the whole paper is the set-indexed process M_n – see (3.2) below. Though its proper asymptotic analysis requires certain mathematical tools, nothing really is required for the basic idea behind it. Suppose we agreed on some order in which we 'visit' or 'inspect' the elements of pooled sample $\{X_i\}_1^n$, so that we first visit $X_{(1)}$, then $X_{(2)}$ and so on. Suppose this order is independent from the indicators $\{\delta_i\}_1^n$. (This order is formalized by the scanning family $\mathcal{A}$ below.) Then the classical empirical process (1.2) can be written as

$$(3.1) \qquad v_n(B) = \left(\frac{n}{n_1 n_2}\right)^{\frac{1}{2}} \sum_{i=1}^{n} \mathbb{1}_B(X_{(i)})(\delta_i - \frac{n_1}{n}).$$

where n_1/n is, obviously, the unconditional probability of drawing 'one' on the i-th draw (see (3.4)), while the process M_n is defined as

$$(3.2) \qquad M_n(B) = \left(\frac{n}{n_1 n_2}\right)^{\frac{1}{2}} \sum_{i=1}^{n} \mathbb{1}_B(X_{(i)})(\delta_i - \widehat{p}_i).$$

where $\widehat{p}_i$ is the conditional probability of drawing 'one' given that many 'ones' were found before the draw: $\widehat{p}_i$ = number of remaining 'ones' $/(n-i+1)$ – see (3.5). This is the only difference between M_n and v_n. Observe in particular that the processes M_n, and u_n and w_n in the sequel, are indexed by the same *multivariate* B's as v_n, and hence that the, in general, univariate scanning family $\mathcal{A}$ does *not* lead to 'univariate' processes. In several aspects the behaviour of M_n seems simpler and more convenient than that of v_n. At least, we know now how to standardize M_n. At the same time, like v_n, M_n preserves 'all information' that is contained in the samples themselves (Lemma 3.1 and Theorem 5.2. Our final processes u_n and w_n are simply weighted versions of M_n.

Now, let $\mathcal{A} = \{A_t, t \in [0,1]\}$ be a family of closed subsets of $[0,1]^m$ with the following properties:

$$1) \quad A_0 = \emptyset, A_1 = [0,1]^m, \qquad 2) \quad A_t \subset A_{t'} \text{ if } t \le t',$$

$$3) \quad \mu(A_t) \text{ is continuous and strictly increasing in } t.$$

This family will be called a scanning family. Denote with X_i an element of the pooled sample $\{X_i'\}_1^{n_1} \cup \{X_i''\}_1^{n_2}$, with X_i' and X_i'' reordered in some arbitrary and for us unimportant way. Under the two-sample null hypothesis this pooled sample $\{X_i\}_1^n$ is simply a sequence of i.i.d. random variables with distribution $P \in \mathcal{P}$ each. Let

$$(3.3) \qquad t(X_i) = \min\{t : X_i \in A_t\},$$

denote with $\{t_i\}_1^n$ the order statistics based on $\{t(X_i)\}_1^n$ and let $\{X_{(i)}\}_1^n$ be the correspondingly reordered X_i's. Put also $t_0 = 0$ and $t_{n+1} = 1$ when needed. Later it will be useful to have in mind that absolute continuity of P (condition (C2)) implies that all the t_i are different a.s. Under H_0 the sequence of Bernoulli random variables $\{\delta_i\}_1^n$,

$$\delta_i = \mathbb{1}\{X_{(i)} \in \text{first sample}\} ,$$

is independent of the $\{X_{(i)}\}_1^n$ and the distribution of the $\{\delta_i\}_1^n$ is that of sampling without replacement from an urn containing n_1 'ones' and n_2 'zeros' (see Section 1).

Now we define the filtration based on the scanning family $\mathcal{A}$. Let

$$\mathcal{F}_0 = \sigma\{P_n(B), B \in \mathcal{B}_0\},$$

$$\mathcal{F}_t = \sigma\{\widehat{P}_{n_1}(B \cap A_t), B \in \mathcal{B}_0\} \vee \mathcal{F}_0, \quad t \in (0,1],$$

$$\mathcal{F}_{(i)} = \sigma\{\delta_j : j \le i\} \vee \mathcal{F}_0.$$

If P is continuous, then the conditional distribution of $\widehat{P}_{n_1}$ given $\mathcal{F}_0$ is free from P, but conditioning on $\mathcal{F}_{(j)}$ also produces a simple distribution for $\widehat{P}_{n_1}$ free from P:

$$\mathbb{P}\{X_{(i)} \in \text{ first sample } |\mathcal{F}_0\} = \mathbb{P}\{\widehat{P}_{n_1}(X_{(i)}) = \tfrac{1}{n_1}|\mathcal{F}_0\} = \tfrac{n_1}{n} \tag{3.4}$$

and, for $j \le i-1$,

$$\mathbb{P}\{X_{(i)} \in \text{ first sample } |\mathcal{F}_{(j)}\} = \frac{n_1 \widehat{P}_{n_1}(A^c_{t_j})}{n P_n(A^c_{t_j})} , \tag{3.5}$$

where $A^c = [0,1]^m \backslash A$; note that $nP_n(A^c_{t_j}) = n - j$ a.s. We will write

$$\widehat{p}(t) = \frac{n_1 \widehat{P}_{n_1}(A^c_{t-})}{n P_n(A^c_{t-})}, 0 \le t \le t_n;\ \widehat{p}_i = \frac{n_1 \widehat{P}_{n_1}(A^c_{t_{i-1}})}{n P_n(A^c_{t_{i-1}})},\ i = 1, ..., n.$$

Consider now v_n along with the filtration $\{\mathcal{F}_t, t \in [0,1]\}$ in a way similar to the construction introduced in Khmaladze (1993), i.e. for each B consider $v_n(B \cap A_t)$, or, equivalently, consider $\widehat{P}_{n_1}(B \cap A_t)$, as P_n is $\mathcal{F}_0$-measurable.

By doing this we obtain a new object in the two-sample theory, which is for each B a semimartingale with respect to $\{\mathcal{F}_t, t \in [0,1]\}$ and for each t a random measure on $\mathcal{B}_0$. Hence we gain the possibility to apply to v_n and $\widehat{P}_{n_1}$ the well-developed theory of martingales and of marked point processes; see e.g., Brémaud (1981)and Jacod and Shiryayev (1987). More specifically, for given B consider the (normalized) martingale part of the submartingale $\{\widehat{P}_{n_1}(B \cap A_t), \mathcal{F}_t, t \in [0,1]\}$. We obtain

$$I\!E(\widehat{P}_{n_1}(B \cap A_{ds})|\mathcal{F}_s) = \frac{n}{n_1}\widehat{p}(s)P_n(B \cap A_{ds}),$$

so that

$$M_n(B,t) = \left(\frac{n_1 n}{n_2}\right)^{\frac{1}{2}} (\widehat{P}_{n_1}(B \cap A_t) - \frac{n}{n_1}\int_0^t \widehat{p}(s)P_n(B \cap A_{ds})) \tag{3.6}$$

is a martingale in t. For a class $\mathcal{B} \subset \mathcal{B}_0$ let

$$a(\mathcal{B}) = \{B \cap A_t : B \in \mathcal{B}, A_t \in \mathcal{A}\}. \tag{3.7}$$

It is clear that $M_n(B,t) = M_n(B \cap A_t, 1)$ for any $B \in a(\mathcal{B})$. Therefore most of the time we will consider the random measure $M_n(\cdot, 1)$ on $a(\mathcal{B})$ and denote it simply $M_n(\cdot)$. However, because the classes $\mathcal{B}$ and $\mathcal{A}$ will play an asymmetric role we will keep also the notation $M_n(B,t)$.

It is easily seen that $M_n(B,t)$ can be rewritten as

$$M_n(B,t) = v_n(B \cap A_t) + \int_0^t \frac{v_n(A_{s-})}{P_n(A^c_{s-})}P_n(B \cap A_{ds}) \tag{3.8}$$

and for $t = 1$ both (3.6) and (3.8) lead to the expression (3.2) which we started with.

Among the first properties of M_n let us mention one: denote $A_{\Delta t} = A_{t+\Delta t} - A_t$, then $M_n(B \cap A_{\Delta t}) = 0$ if $P_n(B \cap A_{\Delta t}) = 0$. The next property is stated in the following

Lemma 3.1 *Let the class $\mathcal{B} \subset \mathcal{B}_0$ be such that $[0,1]^m \in \mathcal{B}$. Given $\mathcal{F}_0$, the restriction of the random measure v_n to $a(\mathcal{B})$ defines the restriction of the random measure M_n to $a(\mathcal{B})$ in a one-to-one way*

The proof of Lemma 3.1 is rather easy, but the lemma itself is important for justification of inference based on M_n, for it says, heuristically, that what can be achieved in testing based on v_n can also be achieved based on M_n and vice versa.

Proof For each $C = B \cap A_\tau$ the value of $M_n(C)$ can be derived using (3.8), since all $A_t \in a(\mathcal{B})$. Now the other way around. Choose $B = [0,1]^m$ and consider the equation

$$M_n(A_t) = v_n(A_t) + \int_0^t \frac{v_n(A_{s-})}{P_n(A^c_{s-})}P_n(A_{ds}). \tag{3.9}$$

It is well-known that it has the unique solution $v_n(A_{t-}) = P_n(A^c_{t-}) \cdot \int_0^{t-} M_n(A_{ds})/P_n(A^c_s)$. Hence, for any B, the unique inverse of (3.8) is

$$(3.10)\qquad v_n(B\cap A_t) = M_n(B\cap A_t) - \int_0^t P_n(B\cap A_{d\tau})\int_0^{\tau-}\frac{M_n(A_{ds})}{P_n(A_s^c)}$$
$$= M_n(B\cap A_t) - \int_0^t \frac{P_n(B\cap A_t) - P_n(B\cap A_s)}{P_n(A_s^c)} M_n(A_{ds}).$$

This concludes the proof. ∎

Not only the properties of $M_n(B,t)$ are convenient in t, but also the properties of it in B are substantially simpler than those of $v_n(B)$ as the next lemma will show. The two martingales $M_n(B,\cdot)$ and $M_n(C,\cdot)$ are called orthogonal if the process

$$\langle M_n(B,\cdot), M_n(C,\cdot)\rangle(t) = \sum_{t_i\le t} I\!E(M_n(B,\Delta t_i)M_n(C,\Delta t_i)|\mathcal{F}_{t_{i-1}})$$

is identically 0. For $C = B$ the process $\langle M_n(B,\cdot), M_n(B,\cdot)\rangle = \langle M_n(B,\cdot)\rangle$ is called the quadratic variation process. In words, $\langle M_n(B,\cdot), M_n(C,\cdot)\rangle$ is a partial sum process of conditional covariances, whereas $\langle M_n(B,\cdot)\rangle$ is a partial sum process of conditional variances. According to (3.5) and (3.2)

$$(3.11)\qquad \langle M_n(B,\cdot), M_n(C,\cdot)\rangle(t) = \frac{n}{n_1 n_2}\sum_{t_i\le t} 1\!\!1_{B\cap C}(X_{(i)})\widehat{p}_i(1-\widehat{p}_i),$$

$$\langle M_n(B,\cdot) - M_n(C,\cdot)\rangle(t) = \frac{n}{n_1 n_2}\sum_{t_i\le t} 1\!\!1_{B\Delta C}(X_{(i)})\widehat{p}_i(1-\widehat{p}_i).$$

This leads to the aforementioned lemma.

Lemma 3.2 *If B and C are disjoint, then $M_n(B,\cdot)$ and $M_n(C,\cdot)$ are orthogonal. Therefore, given $\mathcal{F}_0, M_n(\cdot)$ is a random measure with uncorrelated increments: for $B, C\in a(\mathcal{B})$ with $P_n(B\cap C) = 0$*

$$I\!E(M_n(B)M_n(C)|\mathcal{F}_0) = 0.$$

Note that this is certainly not the case for v_n, since

$$I\!E(v_n(B)v_n(C)|\mathcal{F}_0) = \frac{n}{n-1}(P_n(B\cap C) - P_n(B)P_n(C)).$$

The process $M_n(B,t)$ is essentially an analogue in $I\!R^m$ of the process in (1.5), studied in Urinov (1992) in $I\!R^1$; see also Cabaña and Cabaña (1994), Section 3, where 'wave components' of a Wiener process are studied in $I\!R^2$.

The quadratic variation describes the 'intrinsic' time of a martingale. As formula (3.11) suggests $\langle M_n(B,\cdot)\rangle(t) \approx P_n(B\cap A_t)$. Indeed we have, with $a'(\mathcal{B}) = \{C_1\cap C_2 :\ C_1, C_2\in a(\mathcal{B})\}$,

Lemma 3.3 *Assume $P \in \mathcal{P}$ and $n_1/n \to p_0 \in (0,1)$.*
i) *For a.a. sequences $\{P_n\}_{n\geq 1}$, conditionally on $\mathcal{F}_0$,*

$$\sup_{B\in\mathcal{B}_0} |\langle M_n(B,\cdot)\rangle(1) - P_n(B)| \to 0 \quad a.s. \quad (n\to\infty), \tag{3.12}$$

and hence in particular for any $B \in \mathcal{B}_0$, conditionally on $\mathcal{F}_0$,

$$\langle M_n(B,\cdot)\rangle(t) \to P(B\cap A_t) \quad a.s. \quad (n\to\infty).$$

ii) *Suppose the class $\mathcal{B}$ is a Vapnik-Chervonenkis (VC) class. Then*

$$\sup_{C\in a'(\mathcal{B})} |P_n(C) - P(C)| \to 0 \quad a.s. \quad (n\to\infty).$$

The proof of this lemma is essentially contained in the proof of Lemma 3.4 below and will hence be omitted. Observe that (ii) is just a version of the Glivenko-Cantelli theorem; it is stated here under the above explicit condition to stay in line with Lemma 3.4 and Theorem 4.1.

It follows that the provisional limit, under H_0, for $M_n(B)$ is the restriction to $a(\mathcal{B})$ of a mean zero Gaussian random measure $W_P(B)$ with covariance function

$$\mathbb{E}W_P(B)W_P(C) = P(B\cap C) \tag{3.13}$$

(i.e. Wiener random measure with 'time' P) and apparently, $M_n(B)$ is not asymptotically distribution free. Therefore we will renormalize it in two different ways, see u_n and w_n below. The idea of both these normalizations is inspired by the following simple result (see, e.g., Khmaladze (1988)): If W_P is a Wiener random measure on $\mathcal{B}_0$ with covariance function (3.13) and $P \in \mathcal{P}$, then (cf. (C2))

$$W(B) = \int_B \frac{1}{f^{\frac{1}{2}}(x)} W_P(dx), \quad B \in \mathcal{B}_0, \tag{3.14}$$

is a standard Wiener random measure, i.e. it has covariance function

$$\mathbb{E}W(B)W(C) = \mu(B\cap C).$$

An empirical version of this transformation will be applied to M_n. Suppose

(C4) there exists an $\mathcal{F}_0$-measurable density estimator $\widehat{f}_n$ $(0 \leq \widehat{f}_n < \infty)$, such that if each X_i has distribution $P \in \mathcal{P}$, then for all c'

$$\limsup_{n\to\infty} c_n \sup_{x:f(x)\leq c'} |\widehat{f}_n(x) - f(x)| \leq c \text{ a.s.},$$

with $c = c(c')$, c_n *known* and $c_n \to \infty$ $(n\to\infty)$; moreover for all sufficiently large $c' > 1$

$$\liminf_{n\to\infty} \inf_{x:f(x)\geq c'} \widehat{f}_n(x) > 1 \text{ a.s.}$$

It is not difficult to see that such an $\widehat{f}_n$ exists under mild smoothness conditions on f (see, e.g., Silverman (1986) and Scott (1992)). Set $f_n(x) = \widehat{f}_n(x) \vee c_n^{-1}, x \in [0,1]^m$, and introduce the random measure u_n by

$$(3.15) \qquad u_n(B \cap A_t) \equiv u_n(B,t) = \int\limits_{B\cap A_t} \frac{1}{f_n^{\frac{1}{2}}(x)} M_n(dx)$$
$$= \left(\frac{n}{n_1 n_2}\right)^{\frac{1}{2}} \sum_{i=1}^{n} \frac{1}{f_n^{\frac{1}{2}}(X_{(i)})} \mathbb{1}_{B\cap A_t}(X_{(i)})(\delta_i - \widehat{p}_i).$$

Since f_n is $\mathcal{F}_0$-measurable $u_n(B,\cdot)$ is a martingale indeed, and

$$\langle u_n(B,\cdot)\rangle(t) = \frac{n}{n_1 n_2} \sum_{i=1}^{n} \frac{1}{f_n(X_{(i)})} \mathbb{1}_{B\cap A_t}(X_{(i)}) \widehat{p}_i(1-\widehat{p}_i).$$

Eventually we will prove that u_n is a strong $\mathcal{P}$-solution of the two-sample problem. To prove convergence in distribution of u_n we will need the following result. Set

$$\mu_n(C) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{f_n(X_i)} \mathbb{1}_C(X_i)$$

Lemma 3.4 *Assume $P \in \mathcal{P}$,* (C4) *holds and $n_1/n \to p_0 \in (0,1)$.*
(i) *For a.a. sequences $\{P_n\}_{n\geq 1}$, we have conditionally on $\mathcal{F}_0$, for all $B \in \mathcal{B}_0$*

$$\langle u_n(B,\cdot)\rangle(t) \to \mu(B \cap A_t) \ a.s. \ (n \to \infty).$$

(ii) *Suppose the class $\mathcal{B}$ is a VC class. Then*

$$\sup_{C\in a'(\mathcal{B})} |\mu_n(C) - \mu(C)| \to 0 \ a.s. \ (n \to \infty).$$

It could be noted that the initial observation behind the proof of the lemma is that, according to the Kolmogorov strong law of large numbers (SLLN), for each $B \in \mathcal{B}_0$

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{f(X_i)} \mathbb{1}_B(X_i) \to \mu(B) \ (= \int\limits_B \frac{1}{f} dP) \text{ a.s. } (n \to \infty).$$

Before we prove this lemma let us introduce another normalization of M_n. Consider the Dirichlet (or Voronoi) tessellation of $[0,1]^m$ associated with the sequence $\{X_i\}_1^n$: for each X_i let

$$\Delta(X_i) = \{x \in [0,1]^m : ||x - X_i|| = \min_{1\leq j\leq n} ||x - X_j||\}$$

and let for each C

$$\widetilde{C}_n = \bigcup_{X_i \in C} \Delta(X_i),\ \widetilde{\mu}_n(C) = \mu(\widetilde{C}_n) \overset{\text{a.s.}}{=} \sum_{i=1}^{n} \mu(\Delta(X_i)) 1\!\!1_C(X_i).$$

Now introduce

$$w_n(C) = \frac{n}{(n_1 n_2)^{\frac{1}{2}}} \sum_{i=1}^{n} (\mu(\Delta(X_{(i)})))^{\frac{1}{2}} 1\!\!1_C(X_{(i)})(\delta_i - \widehat{p}_i). \tag{3.16}$$

Then again, since the sequence $\{\mu(\Delta(X_{(i)}))\}_1^n$ is $\mathcal{F}_0$-measurable, $w_n(B \cap A_t)$ is for each t a random measure in B and for each B a martingale in t, and, in the obvious notation,

$$\langle w_n(B, \cdot)\rangle(t) = \frac{n^2}{n_1 n_2} \sum_{i=1}^{n} \mu(\Delta(X_{(i)})) 1\!\!1_{B \cap A_t}(X_{(i)}) \widehat{p}_i (1 - \widehat{p}_i).$$

The expression in (3.16) also can be viewed as another empirical analogue of (3.14):

$$w_n(C) = \int_C \frac{1}{\widetilde{f}_n^{\frac{1}{2}}(x)} M_n(dx),$$

since the step-function $\widetilde{f}_n(x) = (n\mu(\Delta(X_i)))^{-1}$ for all inner points $x \in \Delta(X_i)$ (and let it be 1 on the boundaries $\Delta(X_i) \cap \Delta(X_j)$) can be considered as a density estimator, though an inconsistent one. Its analogue on $I\!R$ is essentially the 1-nearest neighbour estimator. Denote

$$\rho(X_i) = \max_{x \in \Delta(X_i)} ||x - X_i||.$$

We shall consider $\{X_i\}_1^n$ that do not necessarily form a random sample, in order to justify to some extent the possibility of using the normalized ranks $\{R_i\}_1^n$ as mentioned in Section 2. For these more general X_i, the δ_i which determine first and second sample are as in Section 1.

Lemma 3.5 *Suppose that the X_i, $1 \le i \le n$, are random vectors in $[0,1]^m$ with $X_i \ne X_j$ a.s. for $i \ne j$, such that for their empirical distribution P_n we have $P_n \to_w P$ a.s. $(n \to \infty)$, for some $P \in \mathcal{P}$.*
(i) *Then*

$$\rho_n^* = \max_{1 \le i \le n} \rho(X_i) \to 0 \ a.s.\ (n \to \infty).$$

(ii) *For $C \subset \mathcal{B}_0$, set $C^\varepsilon = \{x \in [0,1]^m : ||x - C|| < \varepsilon\}$ and $C_\varepsilon = ((C^c)^\varepsilon)^c$. Suppose $\mathcal{C} \subset \mathcal{B}_0$ is such that $\mathcal{A} \subset \mathcal{C}$ and*

$$\lim_{\varepsilon \downarrow 0} \sup_{C \in \mathcal{C}} \mu(C^\varepsilon \backslash C_\varepsilon) \to 0. \tag{3.17}$$

If $n_1/n \to p_0 \in (0,1)$, then for a.a. sequences $\{P_n\}_{n\geq 1}$, conditionally on $\mathcal{F}_0$,

$$\sup_{C\in\mathcal{C}} |\langle w_n(C,\cdot)\rangle(1) - \mu(C)| \to 0 \ a.s. \ (n\to\infty).$$

(iii) *Also, under (3.17)*

$$\sup_{C\in\mathcal{C}} |\widetilde{\mu}_n(C) - \mu(C)| \to 0 \ a.s. \ (n\to\infty).$$

Proof of Lemma 3.4 Consider

$$|\langle u_n(B,\cdot)\rangle(t) - \mu(B\cap A_t)| \leq |\langle u_n(B,\cdot)\rangle(t) - \frac{1}{n}\sum_{i=1}^n \frac{1}{f(X_{(i)})}1\!\!1_{B\cap A_t}(X_{(i)})| + |\frac{1}{n}\sum_{i=1}^n \frac{1}{f(X_i)}1\!\!1_{B\cap A_t}(X_i) - \mu(B\cap A_t)|.$$

By the SLLN, as $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{f(X_i)}1\!\!1_{B\cap A_t}(X_i) \to \int \frac{1}{f}1\!\!1_{B\cap A_t}dP = \mu(B\cap A_t) \text{ a.s.}$$

So it suffices to consider

$$\left|\frac{1}{n}\sum_{i=1}^n \{\frac{1}{f_n(X_{(i)})}1\!\!1_{B\cap A_t}(X_{(i)})\frac{n^2}{n_1n_2}\widehat{p}_i(1-\widehat{p}_i) - \frac{1}{f(X_{(i)})}1\!\!1_{B\cap A_t}(X_{(i)})\}\right|$$

$$\leq \frac{1}{n}\sum_{i=1}^n \left|\frac{1}{f_n(X_i)} - \frac{1}{f(X_i)}\right| \frac{n^2}{4n_1n_2} + \frac{1}{n}\sum_{i=1}^n \frac{1}{f(X_{(i)})}\left|\frac{n^2}{n_1n_2}\widehat{p}_i(1-\widehat{p}_i) - 1\right|.$$

First we show that the last term above converges to 0 a.s. We will split this sum in the sum involving the $X_{(i)}$'s for which $X_{(i)} \in A_{1-\varepsilon}$ and the sum involving the $X_{(i)}$'s for which $X_{(i)} \notin A_{1-\varepsilon}$. Since $P\in\mathcal{P}$, we have $P(A_{1-\varepsilon}) < 1$ and hence it follows from a kind of conditional Glivenko-Cantelli theorem that

$$\max_{X_{(i)}\in A_{1-\varepsilon}} |\widehat{p}_i - \frac{n_1}{n}| \leq \sup_{t\leq 1-\varepsilon} |\widehat{p}(t) - \frac{n_1}{n}| \to 0 \text{ a.s. } (n\to\infty).$$

(Actually this conditional Glivenko-Cantelli theorem is well-known and is essentially proved in a version of the proof of the ordinary Glivenko-Cantelli theorem for VC classes, see Gaenssler (1983, pp. 28–34).) This yields in combination with the SLLN that

$$\frac{1}{n}\sum_{X_{(i)}\in A_{1-\varepsilon}} \frac{1}{f(X_{(i)})}\left|\frac{n^2}{n_1n_2}\widehat{p}_i(1-\widehat{p}_i) - 1\right| \to 0 \text{ a.s.}$$

The sum dealing with the $X_{(i)}$'s for which $X_{(i)} \notin A_{1-\varepsilon}$ is not greater than

$$\frac{n^2}{4n_1n_2}\frac{1}{n}\sum_{X_i\notin A_{1-\varepsilon}}\frac{1}{f(X_i)}\to\frac{1}{4p_0(1-p_0)}\mu(A^c_{1-\varepsilon})\text{ a.s.}$$

For arbitrary $\delta > 0$, this last expression is less than δ for ε sufficiently small.

Gathering everything we see that the proof of part (i) is complete if we show that

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{f_n(X_i)}-\frac{1}{f(X_i)}\right|=\int_{[0,1]^m}\left|\frac{1}{f_n}-\frac{1}{f}\right|dP_n\to 0\text{ a.s.}$$

Define for $0<\eta<1<c'$, $D_1=\{x\in[0,1]^m:\ f(x)<\eta\}$, $D_2=\{x\in[0,1]^m,\ \eta\le f(x)\le c'\}$, and $D_3=\{x\in[0,1]^m: f(x)>c'\}$. Then for large enough c', we have by (C4)

$$\limsup_{n\to\infty}\int_{D_3}\left|\frac{1}{f_n}-\frac{1}{f}\right|dP_n\le\limsup_{n\to\infty}\int_{D_3}dP_n=P(D_3)<\delta\text{ a.s.}$$

Also for small enough η we have from the definition of f_n

$$\limsup_{n\to\infty}\int_{D_1}\left|\frac{1}{f_n}-\frac{1}{f}\right|dP_n\le\limsup_{n\to\infty}\int_{D_1}\frac{|f_n-f|}{f_nf}dP_n \tag{3.18}$$

$$\le c\lim_{n\to\infty}\int_{D_1}\frac{1}{f}dP_n=c\mu(D_1)<\delta\text{ a.s.}$$

Finally by (C4)

$$\limsup_{n\to\infty}\int_{D_2}\left|\frac{1}{f_n}-\frac{1}{f}\right|dP_n=\limsup_{n\to\infty}\int_{D_2}\frac{|f_n-f|}{f_nf}dP_n$$

$$\le\frac{1}{\eta^2}\limsup_{n\to\infty}\int_{D_2}|f_n-f|dP_n\le\frac{1}{\eta^2}\lim_{n\to\infty}\sup_{x\in D_2}|f_n(x)-f(x)|=0,$$

almost surely. Since δ is arbitrary this completes the proof of part (i).

Now we will prove part (ii). We have

$$\sup_{C\in a'(\mathcal{B})}|\mu_n(C)-\mu(C)|=\sup_{C\in a'(\mathcal{B})}\left|\int_C\frac{1}{f_n}dP_n-\int_C\frac{1}{f}dP\right|$$

$$\le\sup_{C\in a'(\mathcal{B})}\int_C\left|\frac{1}{f_n}-\frac{1}{f}\right|dP_n+\sup_{C\in a'(\mathcal{B})}\left|\int_C\frac{1}{f}(dP_n-dP)\right|.$$

But

$$\sup_{C\in a'(\mathcal{B})} \int_C |\frac{1}{f_n} - \frac{1}{f}| dP_n = \int_{[0,1]^m} |\frac{1}{f_n} - \frac{1}{f}| dP_n,$$

which converges to 0 a.s. as we just showed. So finally we have to prove

$$\sup_{C\in a'(\mathcal{B})} |\int_C \frac{1}{f}(dP_n - dP)| \to 0 \text{ a.s. } (n \to \infty).$$

This is, however, a routine matter: since $\mathcal{B}$ is a VC class and $\int_{[0,1]^m} f^{-1}\, dP = 1 < \infty$, the class of functions $\{f^{-1}\mathbb{1}_C : C \in a'(\mathcal{B})\}$ is a Glivenko-Cantelli class. ∎

Proof of Lemma 3.5 (i) For $k \in I\!N$, let $\mathcal{H}_k$ be the finite set of hypercubes of the form $\Pi_{j=1}^m [r_j/k, (r_j+1)/k]$, $r_j \in \{0, 1, ..., k-1\}$. Since $P_n \to_w P$ a.s. and $P \in \mathcal{P}$, we see that for all $k \in I\!N$, $\sup_{H\in\mathcal{H}_k} |P_n(H) - P(H)| \to 0$ a.s. But since $\inf_{H\in\mathcal{H}_k} P(H) > 0$, this easily implies that $\rho_n^* \to 0$ a.s.

We now prove part (iii). Let $\varepsilon > 0$. Since $\rho_n^* \to 0$ a.s. and for all $C \in \mathcal{C}$, $\widetilde{C}_n \subset C^\varepsilon$ and $(\widetilde{C}_n)^c \subset (C^c)^\varepsilon$, that is $C_\varepsilon \subset C_n \subset C^\varepsilon$, as soon as $\rho_n^* < \varepsilon$, we have

$$\limsup_{n\to\infty} \sup_{C\in\mathcal{C}} |\widetilde{\mu}_n(C) - \mu(C)| \le \sup_{C\in\mathcal{C}} \mu(C^\varepsilon \backslash C_\varepsilon) \text{ a.s.}$$

Now (3.17) proves part (iii).

Finally we consider part (ii). Because of part (iii), it is sufficient to show that

$$\begin{aligned} &\sup_{C\in\mathcal{C}} |\langle w_n(C,\cdot)\rangle(1) - \widetilde{\mu}_n(C)| \\ &= \sup_{C\in\mathcal{C}} |\sum_{i=1}^n \mu(\Delta(X_{(i)}))\mathbb{1}_C(X_{(i)})\{\frac{n^2}{n_1n_2}\widehat{p}_i(1-\widehat{p}_i) - 1\}| \\ &\le \sum_{i=1}^n \mu(\Delta(X_{(i)}))|\frac{n^2}{n_1n_2}\widehat{p}_i(1-\widehat{p}_i) - 1| \to 0 \text{ a.s.} \end{aligned}$$

This last expression, however, can be treated in much the same way as

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{f(X_{(i)})}\left|\frac{n^2}{n_1n_2}\widehat{p}_i(1-\widehat{p}_i) - 1\right|,$$

in the proof of Lemma 3.4. □

4 Weak convergence under H_0: property (α)

Let $\mathcal{B} \subset \mathcal{B}_0$ be the indexing class for the random measures M_n, u_n and w_n defined by (3.2), (3.15) and (3.16) respectively, and consider the space $\ell_\infty(\mathcal{B})$ as the space of trajectories of these measures. To prove the convergence in distribution in $\ell_\infty(\mathcal{B})$ one needs the convergence of the finite-dimensional distributions and the asymptotic equicontinuity property, studied in the empirical process context, e.g., in Pollard (1990), see Theorem 10.2, and Sheehy and Wellner (1992). This property follows from Lemma 4.2 below (in combination with Lemmas 3.3–3.5), which in turn follows from appropriate exponential inequalities.

The first lemma of this section provides these inequalities. Consider the process

$$\xi(t) = \sum_{t_i \le t} \gamma_i(\delta_i - \widehat{p}_i). \tag{4.1}$$

with $\mathcal{F}_0$-measurable coefficients γ_i, $i = 1, \ldots, n$. The process ξ is a martingale and

$$\langle \xi \rangle(t) = \sum_{t_i \le t} \gamma_i^2 \widehat{p}_i(1 - \widehat{p}_i) \le \Gamma(t)/4 \;\text{ with }\; \Gamma(t) = \sum_{t_i \le t} \gamma_i^2.$$

Lemma 4.1 (i) *The process* $\{\exp(\lambda\xi(t) - \frac{\lambda^2}{8}\Gamma(t)),\ 0 \le t \le 1\}$ *is a supermartingale and* $I\!E[\exp(\lambda\xi(t) - \frac{\lambda^2}{8}\Gamma(t)) \mid \mathcal{F}_0] \le 1$.
(ii) *We have for* $z \ge 0$

$$I\!P\{|\xi(1)| > z \mid \mathcal{F}_0\} \le 2e^{-2z^2/\Gamma(1)}. \tag{4.2}$$

Corollary 4.1 *For* $z \ge 0$

$$I\!P\left\{|M_n(B) - M_n(C)| > z\left(\frac{n^2}{n_1 n_2}\right)^{1/2} \mid \mathcal{F}_0\right\} \le 2\exp(-2z^2/P_n(B\Delta C)),$$

$$I\!P\left\{|u_n(B) - u_n(C)| > z\left(\frac{n^2}{n_1 n_2}\right)^{1/2} \mid \mathcal{F}_0\right\} \le 2\exp(-2z^2/\mu_n(B\Delta C)),$$

$$I\!P\left\{|w_n(B) - w_n(C)| > z\left(\frac{n^2}{n_1 n_2}\right)^{1/2} \mid \mathcal{F}_0\right\} \le 2\exp(-2z^2/\widetilde{\mu}_n(B\Delta C)).$$

Proof Take in inequality (4.2), γ_i equal to $1\!\!1_B(X_{(i)}) - 1\!\!1_C(X_{(i)})$ multiplied by $n^{-1/2}$, $(nf_n(X_{(i)}))^{-1/2}$ and $(\mu(\Delta(X_{(i)})))^{1/2}$, respectively. ∎

Proof of Lemma 4.1 The proof follows the well-known pattern. We give it here, though briefly, because the references we know about, represent the exponential inequality for a martingale in Bennett's form (see, e.g., Freedman (1975) or Shorack and Wellner (1986, pp. 899-900) rather than in Hoeffding's form (4.2). Observe that

$$I\!E[e^{\lambda\gamma_i(\delta_i-\widehat{p}_i)}|\mathcal{F}_{i-1}] = e^{-\lambda\gamma_i\widehat{p}_i}e^{\ln(e^{\lambda\gamma_i}\widehat{p}_i+1-\widehat{p}_i)} \leq e^{\frac{\lambda^2\gamma_i^2}{8}},$$

which can be found by expanding the ln, as a function of $\lambda\gamma_i$, up to the second term and observing that the second derivative is bounded by 1/4. Therefore

$$I\!E[e^{\lambda\gamma_i(\delta_i-\widehat{p}_i)-\frac{\lambda^2}{8}\gamma_i^2}|\mathcal{F}_{i-1}] \leq 1$$

which proves (i). Now

$$\begin{aligned} I\!P\{|\xi(1)| > z|\mathcal{F}_0\} &= I\!P\{e^{\lambda\xi(1)-\frac{\lambda^2}{8}\Gamma(1)} > e^{\lambda z-\frac{\lambda^2}{8}\Gamma(1)}|\mathcal{F}_0\} \\ &+I\!P\{e^{-\lambda\xi(1)-\frac{\lambda^2}{8}\Gamma(1)} > e^{\lambda z-\frac{\lambda^2}{8}\Gamma(1)}|\mathcal{F}_0\} \leq 2e^{\frac{\lambda^2}{8}\Gamma(1)-\lambda z}. \end{aligned}$$

Minimization of this bound in λ leads to (4.2). ∎

The next lemma is the main step towards the asymptotic equicontinuity property of our random measures. For the rest of this section we assume our indexing class to be a Vapnik-Chervonenkis (VC) class (see, e.g., Dudley (1978)). Before we proceed formally let us remind again under what distributions this asymptotic equicontinuity will be obtained. All the three random measures considered are functions of $\{X_{(i)}\}_1^n$ (or of P_n) and of $\{\delta_i\}_1^n$ which is independent of P_n and has distribution described in Section 3 – see (3.4) and (3.5). We consider the distributions of M_n, u_n and w_n induced in $\ell_\infty(\mathcal{B})$ by the distribution of $\{\delta_i\}_1^n$ with P_n fixed and call them conditional distributions given P_n, or given $\mathcal{F}_0$. Observe that with this construction there is no need to care about possible non-measurability of P_n as a random element in $\ell_\infty(\mathcal{B})$ nor to require 'enough measurability' of $\mathcal{B}$. So, in most of the cases $\mathcal{B}$ will be required to be just a VC class.

There are two further reasons, specific for the two-sample problem, to use indexing classes no wider than a VC class. The first is, that though we have to study weak convergence under a fairly simple sequence of distributions there are several different distances induced by different distributions on $[0,1]^m$ which occur in the inequalities of the above corollary. We would need to make assumptions on covering numbers $N(\cdot, Q)$ of $\mathcal{B}$ in each of these distances, that is, for Q being P_n, μ_n or $\widetilde{\mu}_n$, $n \in I\!N$, which would be, from the point of view of applications, inconvenient. However, for VC classes we have a uniform-in-Q bound for $\ln N(\cdot, Q)$ – see Dudley (1978), Lemma 7.13, or van der Vaart and Wellner (1996), p. 86, or the proof of Lemma 4.2.

below – and this makes any VC class an appropriate indexing class for each of M_n, u_n and w_n. The second reason is this: though M_n is not the process of eventual interest for the two-sample problem since it is not asymptotically distribution free, we want it to have a limit in distribution for each $P \in \mathcal{P}$. Therefore the indexing class $\mathcal{B}$ should be P-pregaussian for each $P \in \mathcal{P}$ (see, e.g., Sheehy and Wellner (1992)). However, if the class is pregaussian for all P then it must be a VC class (Dudley, 1984, Theorem 11.4.1). Though our $\mathcal{P}$ is more narrow than the class of all distributions (on $[0,1]^m$), still it seems wide enough to motivate the choice of $\mathcal{B}$ being a VC class.

Let us formulate now the next lemma. For a finite (non-negative) measure Q on $\mathcal{B}_0$ and some subclass $\mathcal{B} \subset \mathcal{B}_0$, let $\mathcal{B}'(\varepsilon, Q) = \{(A,B) \in \mathcal{B} \times \mathcal{B} : Q(A \triangle B) \leq \varepsilon\}$. Call $\{M_n\}_{n\geq 1}$ *conditionally asymptotically equicontinuous, uniformly over the discrete distributions*, (CAEC$_{ud}$) if for any $\delta > 0$

$$(4.3) \quad \lim_{\varepsilon \downarrow 0} \limsup_{n\to\infty} \sup_{P_n} I\!P\{ \sup_{(A,B)\in\mathcal{B}'(\varepsilon,P_n)} |M_n(A) - M_n(B)| > \delta | \mathcal{F}_0\} = 0,$$

where P_n runs over all discrete distributions on $[0,1]^m$, concentrated on at most n points. Call $\{u_n\}_{n\geq 1}$ and $\{w_n\}_{n\geq 1}$ *CAEC$_{ud}$* if for these sequences a property similar to (4.3) holds with $\mathcal{B}'(\varepsilon, P_n)$ replaced by $\mathcal{B}'(\varepsilon, \mu_n)$ and $\mathcal{B}'(\varepsilon, \widetilde{\mu}_n)$, respectively. See Sheehy and Wellner (1992).

Lemma 4.2 *Let $\mathcal{B} \subset \mathcal{B}_0$ be a VC class. Then under the null hypothesis $P_1 = P_2$, all three sequences $\{M_n\}_{n\geq 1}$, $\{u_n\}_{n\geq 1}$ and $\{w_n\}_{n\geq 1}$ are $CAEC_{ud}$.*

Proof As above, let $N(\varepsilon, Q)$ denote the covering number of the class $\mathcal{B}$ in the pseudo-metric $d(A,B) = Q(A\triangle B)$ and let α denote the index of the VC class $\mathcal{B}$. Then for all Q and some constant K depending on α, and depending on Q only through $Q([0,1]^m)$

$$N(\varepsilon, Q) \leq K \left(\tfrac{1}{\varepsilon}\right)^{\alpha-1}, \ 0 < \varepsilon < 1$$

(see, e.g., van der Vaart and Wellner (1996), pp. 85–86 and Theorem 2.6.4, and Dudley (1978) Lemma 7.13). Now we can apply this bound to $N(\varepsilon, P_n)$, $N(\varepsilon, \mu_n)$ and $N(\varepsilon, \widetilde{\mu}_n)$ and use the inequalities of Corollary 4.1 and the classical chaining argument, see Dudley (1978) Section 5, but chain down to ∞. We present the proof for M_n; that for u_n and w_n is similar and will be omitted. (In the proof for u_n we will assume that $\mu_n([0,1]^m) \leq 2$, say. This is sufficient for our needs.)

Take $0 < \varepsilon_0 < 1$, to be specified later on and set $\varepsilon_{i+1} = \varepsilon_i^2$, $i = 0,1,2,\ldots$. For $B \in \mathcal{B}$, denote approximating sets corresponding to ε_i with B_i, so $P_n(B\triangle B_i) < \varepsilon_i$. Then

$$(4.4) \qquad I\!P\{ \sup_{(A,B)\in\mathcal{B}'(\varepsilon_0,P_n)} | M_n(A) - M_n(B)| > \delta | \mathcal{F}_0\}$$

$$\le \mathbb{P}\{ \sup_{(A,B)\in\mathcal{B}'(\varepsilon_0,P_n)} | M_n(A_0) - M_n(B_0)| > \tfrac{\delta}{2}|\mathcal{F}_0\}$$
$$+ 2\mathbb{P}\{ \sup_{B\in\mathcal{B}} | M_n(B_0) - M_n(B)| > \tfrac{\delta}{4}|\mathcal{F}_0\}.$$

Using Corollary 4.1 the first term on the right in (4.4) can be bounded from above by

$$\begin{aligned}
&2N^2(\varepsilon_0, P_n)\exp\left(\frac{-2\delta^2}{4}\frac{n_1n_2}{n^2}\frac{1}{3\varepsilon_0}\right)\\
&= 2\exp\left(2\ln N(\varepsilon_0,P_n) - \frac{\delta^2}{6\varepsilon_0}\frac{n_1n_2}{n^2}\right)\\
&\le 2\exp\left(2\ln N(\varepsilon_0,P_n) - \frac{\delta^2}{12\varepsilon_0}p_0(1-p_0)\right),
\end{aligned}$$

where for the last inequality n is taken large enough. Now taking ε_0 small enough, this last expression is not larger than

$$2\exp\left(2\alpha\ln\tfrac{1}{\varepsilon_0} - \frac{\delta^2}{12\varepsilon_0}p_0(1-p_0)\right) < \tfrac{\delta}{2}.$$

Now consider the probability in the second term on the right in (4.4). We have for small enough ε_0 and large enough n

$$\begin{aligned}
&\mathbb{P}\{\sup_{B\in\mathcal{B}}|M_n(B_0) - M_n(B)| > \tfrac{\delta}{4}|\mathcal{F}_0\}\\
\le\ & \sum_{j=0}^{\infty}\mathbb{P}\left\{\sup_{B\in\mathcal{B}}\ |M_n(B_j) - M_n(B_{j+1})| > 2\left(\frac{\alpha\varepsilon_j\ln(1/\varepsilon_{j+1})}{p_0(1-p_0)}\right)^{1/2}|\mathcal{F}_0\right\}\\
\le\ & \sum_{j=0}^{\infty}N^2(\varepsilon_{j+1},P_n)\times\\
&\qquad \sup_{B\in\mathcal{B}}\mathbb{P}\left\{|M_n(B_j) - M_n(B_{j+1})| > 2\left(\frac{\alpha\varepsilon_j\ln(1/\varepsilon_{j+1})}{p_0(1-p_0)}\right)^{1/2}|\mathcal{F}_0\right\}\\
\le\ & 2\sum_{j=0}^{\infty}N^2(\varepsilon_{j+1},P_n)\exp\left(\frac{-2\cdot 4\alpha\varepsilon_j\ln(1/\varepsilon_{j+1})\frac{n_1n_2}{n}}{2p_0(1-p_o)\varepsilon_j}\right)\\
\le\ & 2\sum_{j=0}^{\infty}\exp(2\ln N(\varepsilon_{j+1},P_n) - 3\alpha\ln(1/\varepsilon_{j+1}))\\
\le\ & 2\sum_{j=0}^{\infty}\exp(-\alpha\ln(1/\varepsilon_{j+1})) = 2\sum_{j=0}^{\infty}\varepsilon_{j+1}^{\alpha} < \varepsilon_0 < \tfrac{\delta}{4}.
\end{aligned}$$

Hence, since the 'n large enough' requirements do not depend on P_n,

$$\limsup_{n\to\infty}\sup_{P_n}\mathbb{P}\ \sup_{(A,B)\in\mathcal{B}'(\varepsilon_0,P_n)}|M_n(A) - M_n(B)| > \delta|\mathcal{F}_0\} \le \delta,$$

which gives (4.3). ∎

We are now prepared to formulate the statement on weak convergence of M_n, u_n and w_n under the null hypothesis. What mainly remains to be proved is the finite-dimensional convergence in distribution, which we will obtain via the martingale central limit theorem. Let '$\overset{\mathcal{D}(P_n)}{\to}$' denote convergence in distribution under the sequence of conditional distributions, given P_n.

Theorem 4.1 *Let $n_1/n \to p_0 \in (0,1)$. For any VC class $\mathcal{B} \subset \mathcal{B}_0$ and any $P \in \mathcal{P}$ we have for a.a. sequences $\{P_n\}_{n\geq 1}$*

$$M_n \overset{\mathcal{D}(P_n)}{\to} W_P \quad (n \to \infty)$$

in the space $\ell_\infty(\mathcal{B})$.
If condition (C4) *holds, then also*

$$u_n \overset{\mathcal{D}(P_n)}{\to} W \quad (n \to \infty)$$

in $\ell_\infty(\mathcal{B})$; and, if condition (3.17) *of Lemma 3.5 holds for $\mathcal{C} = a(\mathcal{B})$, then also*

$$w_n \overset{\mathcal{D}(P_n)}{\to} W \quad (n \to \infty)$$

in $\ell_\infty(\mathcal{B})$. Moreover, since $\mathcal{B}$ is a VC class the limiting processes W_P and W are bounded and uniformly continuous with respect to $Q(\cdot \triangle \cdot)$, for Q being P and μ, respectively.

Proof First note that Lemma 4.2 in conjunction with Lemma 3.3 yields that for any $\delta > 0$ and a.a. sequences $\{P_n\}_{n\geq 1}$

$$\text{(4.5)} \qquad \lim_{\varepsilon\downarrow 0} \ \limsup_{n\to\infty} \ \mathbb{P}\{ \sup_{(A,B)\in\mathcal{B}'(\varepsilon,P)} |M_n(A) - M_n(B)| > \delta | \mathcal{F}_0\} = 0.$$

Similarly Lemmas 4.2 and 3.4 lead to (4.5) with P replaced by μ and M_n by u_n; also Lemmas 4.2 and 3.5 give (4.5) with P replaced by μ and M_n by w_n (note that the conditions of Lemma 3.5 also hold for $\mathcal{C} = a'(\mathcal{B})$). This settles the proper asymptotic equicontinuity for the three processes.

Now we turn to the convergence of the finite dimensional distributions. For a sequence of martingales $\xi_n(t) = \sum_{t_i\leq t} \gamma_{in}(\delta_i - \widehat{p}_i)$, the convergence $\xi_n \overset{\mathcal{D}}{\to} \xi$ in $D[0,1]$ to a Brownian motion with covariance $F(s \wedge t)$ is equivalent to the conditions:

1) $\langle \xi_n \rangle(t) \xrightarrow{\mathbb{P}} F(t)$ for all $t \in [0,1]$, and

2) $\sum_{i=1}^{n} \gamma_{in}^2 1\!\!1\{\gamma_{in}^2 > \varepsilon\}\widehat{p}_i(1-\widehat{p}_i) \xrightarrow{\mathbb{P}} 0$ for all $\varepsilon > 0$

(the Lindeberg condition), see, e.g., Liptser and Shiryayev (1981) or Shorack and Wellner (1986, pp. 894 and 895). To obtain the finite-dimensional convergence for $M_n(B_1, \cdot), \cdots, M_n(B_k, \cdot)$, according the Cramér-Wold device, it is sufficient to prove the convergence for all linear combinations $\lambda_1 M_n(B_1, \cdot) + \cdots + \lambda_k M_n(B_k, \cdot)$, which leads to the choice

$$\gamma_{in} = \left(\frac{n}{n_1 n_2}\right)^{1/2} \sum_{j=1}^{k} \lambda_j 1\!\!1_{B_j}(X_{(i)}),$$

but since $1\!\!1_B \cdot 1\!\!1_C = 1\!\!1_{B\cap C}$ the choice of just one indicator is sufficient.

Now according to Lemma 3.3 condition 1) is satisfied (even almost surely) with $F(t) = EW_P^2(B \cap A_t) = P(B \cap A_t)$, whereas condition 2) is trivially satisfied. For u_n, according to Lemma 3.4, condition 1) holds also a.s., with $F(t) = EW^2(B \cap A_t) = \mu(B \cap A_t)$ and condition 2) as well (see (3.18)), while for w_n, again, a.s.-versions of both conditions follow from Lemma 3.5 with the same F as for u_n. This completes the proof of the finite-dimensional convergence of M_n, u_n and w_n and hence of the theorem. ∎

5 Weak convergence under alternatives: properties (β) and (γ)

According to Theorem 4.1 the processes u_n and w_n have property (α) (see Section 2). In this section we need to study if they also have the properties (β) or (γ); that is, we will study the weak convergence of u_n and w_n, as well as M_n, under alternatives (C3). Since these are contiguous alternatives the asymptotic equicontinuity follows and we need only to study the finite-dimensional convergence of these processes. The usual way to do this is to study the joint weak convergence of each of our processes with the logarithm of the likelihood ratio

$$L_n = \sum_{i=1}^{n} \left[\delta_i \ln \frac{dP_1}{dP}(X_{(i)}) + (1-\delta_i) \ln \frac{dP_2}{dP}(X_{(i)})\right]$$

and then to apply LeCam's Third Lemma, see e.g., Shorack and Wellner (1986, p. 156). It is well known that

$$Z_n = L_n - \sqrt{\frac{n}{n_1 n_2}} \sum_{i=1}^{n} (\delta_i - \frac{n_1}{n}) h(X_{(i)}) \tag{5.1}$$

converges in probability to a constant c under the distribution P^n.

It would be, however, more consistent with the presentation in Section 4 if we consider, instead of L_n, the logarithm of the likelihood ratio of the conditional distributions, given $\mathcal{F}_0$, which is

$$L'_n = L_n - \ln I\!E[e^{L_n}|\mathcal{F}_0]$$

and to study the joint convergence of our processes and L'_n under a.a. sequences of the conditional hypothetical distribution $I\!P\{\cdot|\mathcal{F}_0\}$. As shown in Urinov (1992), for L'_n it is again true that

$$Z'_n = L'_n - \sqrt{\frac{n}{n_1 n_2}} \sum_{i=1}^{n} (\delta_i - \frac{n_1}{n}) h(X_{(i)})$$

satisfies $Z'_n \stackrel{I\!P}{\to} c$, still under P^n. However, it is not true, as far as we understand it, that condition (C3) implies convergence of z'_n to a constant in $I\!P\{\cdot|\mathcal{F}_0\}$ for a.a. sequences $\{P_n\}_{n\geq 1}$. Hence, though it is possible to show the convergence in this sense to the appropriate limits if L'_n is replaced by its leading first term (see the proof of Theorem 5.1 below), the eventual statement of convergence is true under the unconditional distributions P^n and $P_1^{n_1} \times P_2^{n_2}$ only.

Write $H(t) = \int_{A_t^c} h dP$ and let

$$g(x) = h(x) - \frac{H(t(x))}{P(A^c_{t(x)})}$$

where $t(x)$ is defined as in (3.3). Remark that the linear operator that maps h into g is norm preserving (though not one-to-one since it annihilates constant functions):

$$\int g^2 dP = \int (h - \int h dP)^2 dP = \int h^2 dP (= ||h||^2). \tag{5.2}$$

Now denote with Z a $N(0, ||h||^2)$ random variable (Z will be the limit of $L_n - z_n$; cf. also (2.2)) such that (W_P, Z) is jointly Gaussian, that is, for any finite collection of $B_1, \cdots, B_k \in \mathcal{B}$ the vector $(W_P(B_1), \cdots, W_P(B_k), Z)$ is Gaussian, and let $\mathrm{Cov}(W_P(B), Z) = \int_B g dP$. Similarly, let (W, Z) be jointly Gaussian with $\mathrm{Cov}(W(B), Z) = \int_B g f^{1/2} d\mu$. Let '$\stackrel{\mathcal{D}}{\to}$' and '$\stackrel{\tilde{\mathcal{D}}}{\to}$' denote convergence in distribution under P^n and $P_1^{n_1} \times P_2^{n_2}$, respectively.

Theorem 5.1 *If the class $\mathcal{B} \subset \mathcal{B}_0$ is such that $M_n \stackrel{\mathcal{D}}{\to} W_P$ and/or $u_n \stackrel{\mathcal{D}}{\to} W$ $(n \to \infty)$ in $\ell_\infty(\mathcal{B})$, then*

$$(5.3)\qquad M_n \overset{\widetilde{\mathcal{D}}}{\to} W_P + \int_{\cdot} g dP \quad (n \to \infty)$$

and/or

$$(5.4)\qquad u_n \overset{\widetilde{\mathcal{D}}}{\to} W + \int_{\cdot} g f^{1/2} d\mu \quad (n \to \infty)$$

in $\ell_\infty(\mathcal{B})$.

The proof of this theorem is deferred to the second half of this section, but to explain the nature of the function g already here, let us remark that the leading term of L_n and L_n' has the following explicit representation (see (4.1)):

$$(5.5)\qquad \sum_{i=1}^{n}\left(\delta_i - \frac{n_1}{n}\right) h(X_{(i)}) = \sum_{i=1}^{n}(\delta_i - \widehat{p}_i) g_n(X_{(i)}),$$

where $g_n(x) = h(x) - (\int_{A^c_{t(x)}} h dP_n)/P_n(A^c_{t(x)})$ has the same form as the function g only with P replaced by the empirical distribution P_n. The equality (5.5) can be derived from (3.8) or verified directly.

Now let us consider whether it follows from this theorem that u_n has property (β). Let $\mathbf{Q}_{u_n}$ and $\widetilde{\mathbf{Q}}_{u_n}$ denote the distributions of u_n under P^n and $P_1^{n_1} \times P_2^{n_2}$ respectively, and let $\mathbf{Q}$ and $\widetilde{\mathbf{Q}}$ denote the distributions of W and $W + \int_{\cdot} g f^{1/2} d\mu$ respectively.

Theorem 5.2 *If the indexing class* $\mathcal{B}$ *generates* $\mathcal{B}_0$*, then for each sequence of alternatives satisfying* (C3)

$$d(\mathbf{Q}_{u_n}, \widetilde{\mathbf{Q}}_{u_n}) \to d(\mathbf{Q}, \widetilde{\mathbf{Q}}) = \lambda \quad (n \to \infty).$$

Hence, Theorems 4.1 and 5.2 show, that if $\mathcal{B}$ is a VC class generating $\mathcal{B}_0$ and (C4) holds, then u_n is a strong $\mathcal{P}$-solution of the two-sample problem.

Remark that the process M_n also possesses property (β). It only lacks property (α).

Let us now consider w_n. To find out what is the limiting covariance between $w_n(B)$ and L_n we need to study the limit of the expression

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_B(X_{(i)}) g_n(X_{(i)}) \sqrt{n\mu(\Delta(X_{(i)}))}\, \widehat{p}_i(1-\widehat{p}_i)$$

where the multipliers $\widehat{p}_i(1-\widehat{p}_i)$ are not essential from the point of view of convergence. On the unit interval, i.e. $m = 1$, it can be proved that

$$(5.6)\qquad \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_B(X_i) g(X_i) \sqrt{n\mu(\Delta(X_i))}\, \mathbb{P} \;\to\; k \int_B g f^{-1/2} dP$$

$$(5.7) \qquad = k \int_B g f^{1/2} d\mu,$$

with $k = \frac{3}{4}\frac{\sqrt{\pi}}{2}$, using, e.g., the general method presented in Borovikov (1987). It follows, heuristically speaking, from the fact that the $n\mu(\Delta(X_i))$ behave 'almost' as independent random variables each with a Gamma(2) distribution with scale parameter $2f(X_i)$, and so k stands for the moment of order $\frac{1}{2}$ of a Gamma(2) distribution with scale parameter 2. However, in the unit cube, $[0,1]^m$, we will need to keep (5.6) for some $k < 1$ as an assumption.

Let (W', Z') be again jointly Gaussian with the same marginal distributions as that of W and Z, but with covariance $\mathrm{Cov}(W'(B), Z') = k \int_B g f^{1/2} d\mu$.

Theorem 5.3 *If the class $\mathcal{B} \subset \mathcal{B}_0$ is such that $w_n \xrightarrow{D} W$ in $\ell_\infty(\mathcal{B})$ and if (5.6) is true, then*

$$(5.7) \qquad w_n \xrightarrow{\widetilde{D}} W + k \int_{\cdot} g f^{1/2} d\mu.$$

Let $\widetilde{\mathbf{Q}}^{(k)}$ be the distribution of the right hand side of (5.7). If $\mathcal{B}$ generates $\mathcal{B}_0$ then

$$(5.8) \qquad d(\mathbf{Q}_{w_n}, \widetilde{\mathbf{Q}}_{w_n}) \to d(\mathbf{Q}, \widetilde{\mathbf{Q}}^{(k)}) = 2\Phi\left(\frac{1}{2}k||h||\right) - 1.$$

From (5.8) it follows that under the conditions of Theorem 5.3 the process w_n certainly possesses property (γ) although not property (β) because $\Phi(\frac{1}{2}k||h||) < \Phi(\frac{1}{2}||h||)$. So, w_n is a weak $\mathcal{P}$-solution of the two-sample problem.

Finally, we present the postponed proofs of Theorems 5.1 and 5.2. The proof of Theorem 5.3 is much the same and will therefore be omitted.

Proof of Theorem 5.1 Since the sequence of alternative distributions $\{P_1^{n_1} \times P_2^{n_2}\}_{n\geq 1}$ is contiguous to the sequence $\{P^n\}_{n\geq 1}$, the CAEC$_{ud}$ property of M_n and/or u_n will be true under the alternative distributions as well. Hence (5.3) and (5.4) will follow if we show the convergence of the finite dimensional distributions of M_n and/or u_n to the proper limits. Let us focus on u_n – the proof for M_n is similar and simpler. The convergence

$$\{u_n(B_j)\}_{j=1}^k \xrightarrow{\widetilde{\mathcal{D}}} \{W(B_j) + \int_{B_j} g f^{1/2} d\mu\}_{j=1}^k$$

will follow from the Cramér-Wold device, the convergence

(5.9) $$\sum_{j=1}^{k} \alpha_j u_n(B_j) + \beta Z_n \xrightarrow{\mathcal{D}} \sum_{j=1}^{k} \alpha_j W(B_j) + \beta Z,$$

where $\{\alpha_j\}_{j=1}^k$ and β are any constants and

$$Z_n = \sqrt{\frac{n}{n_1 n_2}} \sum_{i=1}^{n} h(X_{(i)})(\delta_i - \frac{n_1}{n}),$$

and from LeCam's Third Lemma. To see that (5.9) is true observe that, given P_n, the left hand side is the value of a martingale in t

$$\sqrt{\frac{n}{n_1 n_2}} \sum_{i=1}^{t} \left[\frac{1}{f_n^{1/2}(X_{(i)})} \sum_{j=1}^{k} \alpha_j \mathbb{1}_{B_j}(X_{(i)}) + \beta g_n(X_{(i)}) \right] (\delta_i - \widehat{p}_i)$$

(cf. (4.1)) at the last point $t = n$. Hence if we verify that

(5.10) $$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{f_n^{1/2}(X_{(i)})} \sum_{j=1}^{k} \alpha_j \mathbb{1}_{B_j}(X_{(i)}) + \beta g_n(X_{(i)}) \right]^2 \frac{n^2}{n_1 n_2} \widehat{p}_i(1 - \widehat{p}_i) \\ &\to \int_{[0,1]^m} \left[\frac{1}{f^{1/2}} \sum_{j=1}^{k} \alpha_j \mathbb{1}_{B_j} + \beta g \right]^2 dP \text{ a.s. } (n \to \infty) \end{aligned}$$

for a.a. $\{P_n\}_{n\geq 1}$, then actually '$\xrightarrow{\mathcal{D}(P_n)}$ a.s.' will be proved and hence '$\xrightarrow{\mathcal{D}}$' as well. However, (5.10) will follow from the SLLN if we show that the functions f_n and g_n can be replaced by f and g respectively and use the truncation applied in the proof of Lemma 3.4. We have

$$\sup_{0\leq t\leq 1} \left| \int_{A_t^c} h dP_n - \int_{A_t^c} h dP \right| \to 0 \text{ and } \sup_{0\leq t\leq 1} |P_n(A_t^c) - P(A_t^c)| \to 0,$$

a.s. and hence

$$\sup_{x\in A_{1-\varepsilon}} |g_n(x) - g(x)| \to 0 \text{ a.s. } (n \to \infty),$$

while on $A_{1-\varepsilon}^c$ we have, according to (5.5),

$$\begin{aligned} &\frac{n}{n_1 n_2} \sum_{i=1}^{n} g_n^2(X_{(i)}) \mathbb{1}_{A_{1-\varepsilon}^c}(X_{(i)}) \widehat{p}_i(1 - \widehat{p}_i) \\ &\leq \frac{1}{n-1} \sum_{i=1}^{n} h^2(X_{(i)}) \mathbb{1}_{A_{1-\varepsilon}^c}(X_{(i)}) \to \int_{A_{1-\varepsilon}^c} h^2 dP < \delta \text{ a.s.} \end{aligned}$$

The proof of

$$\int |\frac{1}{f_n^{1/2}} - \frac{1}{f^{1/2}}|^2 dP \to 0 \text{ a.s.}$$

is similar to the one used in Lemma 3.4 and is omitted here. ∎

Proof of Theorem 5.2 If $\mathcal{B}$ generates $\mathcal{B}_0$, then L_n is a linear functional of u_n and, hence, the following two distances in variation are equal:

$$d(P_1^{n_1} \times P_2^{n_2}, P^n) = d(\widetilde{\mathbf{Q}}_{u_n}, \mathbf{Q}_{u_n}).$$

Hence $d(\widetilde{\mathbf{Q}}_{u_n}, \mathbf{Q}_{u_n}) \to 2\Phi(\frac{1}{2}||h||) - 1$. Again since $\mathcal{B}$ generates $\mathcal{B}_0$, the distance in variation between the distributions of W and $W + \int_\cdot gf^{1/2}d\mu$ on $\mathcal{B}$ coincides with the one on the whole $\mathcal{B}_0$. Therefore the log-likelihood statistics of these two Gaussian processes on $\mathcal{B}$ and $\mathcal{B}_0$ coincide and are equal to

$$\int gf^{1/2}dV - \tfrac{1}{2}\int g^2 dP$$

which, because of (5.2), is $N(-\frac{1}{2}||h||^2, ||h||^2)$ when $V = W$ and which is $N(\frac{1}{2}||h||^2, ||h||^2)$ when $V = W + \int_\cdot gf^{1/2}d\mu$. The distance in variation between these two normal distributions is, obviously, $\lambda = 2\Phi(\frac{1}{2}||h||) - 1$. ∎

Acknowledgements. We are grateful to the referees for several comments that led to an improved presentation.

REFERENCES

BICKEL, P. J. (1969). A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Math. Statist.* **40** 1-23.

BICKEL, P. J. and BREIMAN, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11** 185-214.

BOROVIKOV, V. P. (1987). Limit theorems for statistics that are partial sums of function of spacings. *Theory Probab. Appl.* **32** 86-97.

BRÉMAUD, P. (1981). *Point Processes and Queues. Martingale Dynamics.* Springer, New York.

CABANA, A. and CABANA, E. M. (1994). Goodness of fit and comparison tests of the Kolmogorov-Smirnov type for bivariate populations. *Ann. Statist.* **22** 1447-1459.

CHATTERJEE, S. K. and SEN, P.K. (1964). Non-parametric tests for the bivariate two-sample location problem. *Calcutta Statist. Assoc. Bull.* **13** 18-58.

CHIBISOV, D. M. (1961). On tests of fit based on sample spacings. (in Russian) *Teor. Verojatnost. i Primenen* **6** 354-358.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899-929.

DUDLEY, R. M. (1984). A course on empirical processes. École d'été de probabilité de Saint-Flour XII-1982. *Lecture Notes in Math.* **1097** 1-142. Springer, New York.

DZAPARIDZE, K. O. and NIKULIN, M. S. (1979). The probability distributions of the Kolmogorov and omega-squared statistics for continuous distributions with shift and location parameters. *Zap. Nauchn. Sem. LOMI* **85** 46-74.

FISHER, R. A. (1936). *Statistical Methods for Research Workers.* Edinburgh.

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100-118.

FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalisations of the Wald-Wolfovitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697-717.

GAENSSLER, P. (1983). *Empirical Processes.* IMS, Hayward, Calif.

HENZE, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16** 772-783.

HENZE, N. and PENROSE, M.D. (1999). On the multivariate runs test. *Ann. Statist.* **27** 290-298.

JACOD, J. and SHIRYAYEV, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer, New York.

KHMALADZE, E. V. (1988). An innovation approach to goodness-of-fit tests in $\mathbb{R}^m$. *Ann. Statist.* **16** 1503-1516.

KHMALADZE E. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798-829.

KIM, K.-K. and FOUTZ, R. V. (1987). Tests for the multivariate two-sample problem based on empirical probability measures. *Canad. J. Statist.* **15** 41-51.

LIPTSER, R. SH. and SHIRYAYEV, A. N. (1981). On necessary and sufficient conditions in the functional central limit theorem for semimartingales. *Theory Probab. Appl.* **16** 130-135.

OOSTERHOFF, J. and VAN ZWET, W. R. (1979). A note on contiguity and Hellinger distance. *In Contribution to Statistics.* (J. Jurečková, ed.) 157-166. Reidel, Dordrecht.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* IMS. Hayward, Calif.

PURI, M. L. and SEN, P. K. (1966). On a class of multivariate multisample rank-order tests. *Sankhyā Ser. A.* **28** 353–376.

PURI, M. L. and SEN, P. K. (1969). On a class of rank order tests for the identity of two multiple regression surfaces. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **12** 1-8.

SCHILLING, M. F. (1983). Goodness of fit testing in R^m based on the weighted empirical distribution of certain nearest neighbor statistics. *Ann. Statist.* **11** 1-12.

SCOTT, D. W. (1992). *Multivariate Density Estimation. Theory Practice and Visualisation.* Wiley, New York.

SHEEHY, A. and WELLNER, J. A. (1992). Uniform Donsker classes of functions. *Ann. Probab.* **20** 1983-2030.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

SMIRNOV, N. V. (1939). Estimate of deviation between empirical distribution functions in two independent samples. (in Russian) *Bull. Moscow Univ.* **2** 3-16.

URINOV, I. K. (1992). Test of homogeneity under small grouping. Martingale limit theorems. *Theory Probab. Appl.* **37** 658-672.

VAN DER VAART, A. W. and WELLNER, J. A. (1996) *Weak Convergence and Empirical Processes With Applications to Statistics.* Springer, New York.

WALD, A. and WOLFOWITZ, J. (1944). Statistical tests based on permutations of the observations. *Ann. Math. Statist.* **15** 358-372.

DEPARTMENT OF MATHEMATICS
AND COMPUTING SCIENCE
EINDHOVEN UNIVERSITY OF TECHNOLOGY
P.O. BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS
einmahl@win.tue.nl

SCHOOL OF MATHEMATICS
DEPARTMENT OF STATISTICS
THE UNIVERSITY OF NSW
SYDNEY 2052
AUSTRALIA *estate@maths.unsw.EDU.AU*