*The Annals of Mathematical Statistics*
1969, Vol. 40, No. 1, 1–23

# A DISTRIBUTION FREE VERSION OF THE SMIRNOV TWO SAMPLE TEST IN THE $p$-VARIATE CASE[1]

BY P. J. BICKEL

*University of California, Berkeley*

**1. Introduction.** One of the classic problems of the theory of nonparametric inference is testing whether two samples come from the same or different populations. If the observations are univariate and we suppose only that the parent populations are governed by a continuous distribution function, genuinely distribution free and asymptotically consistent tests have been proposed by Smirnov, Von Mises, and Lehmann among others. For a review of the enormous literature that has grown up about these and related tests we refer to [1] and [3]. The multivariate case seems to have been studied far less fully. The obvious generalization of the univariate methods leads to procedures which are not distribution free (see [17]). Lehmann [12] has proposed a genuinely distribution free test consistent against all alternatives in this instance also, but his method involves post experimental randomization as an intrinsic factor and is therefore not fully satisfactory. Rosenblatt in [14] has given an ingenious solution to the goodness of fit problem in the multivariate case but his device does not seem to carry over to the two sample problem. Recently, David and Fix [2] and Sen and Chatterjee [16] have remarked that the classical permutation principle of Fisher, as used for instance by Wald and Wolfowitz in [19], when applied to rank tests in multivariate problems leads to procedures which are distribution free and which can, in principle at least, be tabled.

Our approach consists of combining this remark with the generalization of the classical Smirnov test to obtain a procedure which we show is distribution free and consistent against all alternatives. To achieve this aim we require a theorem on the "conditional" convergence of the empirical distribution function which is analogous to known results on permutation distributions (cf. [19], [8]). This theorem is then applied to yield known results of Kiefer and Wolfowitz [11] and Dudley [6].

The paper is organized as follows: Section 2 contains a formal statement of the model and our proposed procedure. In Section 3 we introduce the stochastic process considerations we need, state the main theorems on convergence of stochastic processes ((3.1) and (3.2)) and derive the consistency of our procedure from these as well as the previously mentioned results of Dudley and Kiefer and Wolfowitz. Section 4 deals with some auxiliary results on sampling from a finite population necessary for the proof of Theorem 3.1. Two of these results are due to Hájek ([8] and [9]), and one to Wald, Wolfowitz, Noether and Hoeffding

Received 11 September 1967.

[1] This paper was prepared with the partial support of the Office of Naval Research Contract N00014-67-A-0114-0004.

(see e.g. [7]). The others, Lemmas 4.3, 4.6, and 4.7, are new and may be of independent interest. The proof of Theorem 3.1 is completed in Sections 5, 6, and 7.

**2. The model and the procedure.** Let $\mathbf{X}_1, \mathbf{X}_2, \cdots; \mathbf{Y}_1, \mathbf{Y}_2, \cdots$, be two sequences of $p$-vectors (observations) defined on a measurable space $(\Omega, \mathcal{A})$. For simplicity we take $\Omega$ to be $[R^p]^\infty \times [R^p]^\infty$ where $[R^p]^\infty$ is the space of ordered sequences of ordered $p$ tuples of real numbers. Let $\mathcal{A}$ be the usual product Borel field on $\Omega$, and let $\mathbf{X}_i$, $\mathbf{Y}_i$ be the usual projection maps.

We suppose one of a family of probability measures $P_\theta$ to hold on $(\Omega, \mathcal{A})$, where $\theta$ ranges over $\Theta = \{(F, G): F, G$ probability distribution functions on $R^p\}$ and if, $\theta = (F, G)$, $P_\theta$ is the product probability measure on $(\Omega, \mathcal{A})$ which makes the $\mathbf{X}_i$ independent and identically distributed according to $F$ and the $\mathbf{Y}_i$ similarly independent and identically distributed according to $G$. Let $\Theta_0 = \{\theta \varepsilon \Theta: \theta = (F, F)$ for some $F\}$. The problem we are interested in is testing the hypothesis, $H: \theta \varepsilon \Theta_0$ vs. $K: \theta \not\varepsilon \Theta_0$, consistently at level $\alpha$.

Formally we seek a sequence of (critical) functions $\varphi_{m,n} : [R^p]^m \times [R^p]^n \rightarrow [0, 1]$ which are Borel measurable such that for any given $0 < \alpha \leqq 1$

$$E_\theta(\varphi_{m,n}(\mathbf{X}_1, \cdots, X_m, \mathbf{Y}_1, \cdots, \mathbf{Y}_n)) \leqq \alpha \tag{2.1}$$

for all $\theta \varepsilon \Theta_0$, and

$$\lim_{m,n\to\infty} E_\theta(\varphi_{m,n}(\mathbf{X}_1, \cdots, \mathbf{X}_m, \mathbf{Y}_1, \cdots, \mathbf{Y}_n)) = 1 \tag{2.2}$$

for $\theta \not\varepsilon \Theta_0$, where $E_\theta$ indicates that expectation is taken under the assumption that $\theta$ holds.

We begin with some definitions. Let $F_m(\mathbf{t}, \omega)$ denote the empirical cumulative distribution of $\mathbf{X}_1(\omega), \cdots, \mathbf{X}_m(\omega)$ for $\omega \varepsilon \Omega$, and $\mathbf{t} \varepsilon R^p$. Thus

$$F_m(\mathbf{t}, \omega) = m^{-1} \textstyle\sum_{j=1}^m I_{A(\mathbf{t})}(\mathbf{X}_j(\omega)) \tag{2.3}$$

where $A(\mathbf{t}) = \{\mathbf{s} \varepsilon R^p: \mathbf{s} \leqq \mathbf{t}\}$ and $I_A(\mathbf{s})$ is the indicator function of a set $A$ in $R^p$. As is customary $\mathbf{s} \leqq \mathbf{t}$ refers to coordinatewise inequality between $\mathbf{s} = (s_1, \cdots, s_p)$ and $\mathbf{t} = (t_1, \cdots, t_p)$.

Similarly, define

$$G_n(\mathbf{t}, \omega) = n^{-1} \textstyle\sum_{j=1}^n I_{A(\mathbf{t})}(\mathbf{Y}_j(\omega)) \tag{2.4}$$

and

$$H_N(\mathbf{t}, \omega) = \lambda_N F_m(\mathbf{t}, \omega) + (1 - \lambda_N)G_n(\mathbf{t}, \omega), \tag{2.5}$$

where $N = m + n$, $\lambda_N = mN^{-1}$.

In the sequel we shall frequently refer to conditional probabilities and conditional expectations given $H_N$, written $P_\theta[\cdot \mid H_N]$ and $E_\theta(\cdot \mid H_N)$. We use this convenient notation to refer to the (regular) conditional probabilities of events in $\mathcal{A}$ and conditional expectations of functions measurable $\mathcal{A}$ when the conditioning sigma field

$$\mathcal{B} = \{B \varepsilon \mathcal{A}: (\mathbf{x}_1, \cdots, \mathbf{x}_m, \mathbf{x}_{m+1}, \cdots; \mathbf{y}_1, \cdots, \mathbf{y}_n, \mathbf{y}_{n+1}, \cdots) \varepsilon B$$
$$\Leftrightarrow (P(\mathbf{x}_1), \cdots, P(\mathbf{x}_m), \mathbf{x}_{m+1}, \cdots; P(\mathbf{y}_1), \cdots, P(\mathbf{y}_n), \mathbf{y}_{n+1}, \cdots) \varepsilon B\}$$

for any $P$ which is a map of the set $\{\mathbf{x}_1, \cdots, \mathbf{x}_m, \mathbf{y}_1, \cdots, \mathbf{y}_n\}$ onto itself. This is clearly the information contained in $H_N$, viz. the "values" of the first $m$ and $n$ observations from our two samples without any identification of their "sample" origin. Clearly we can choose the regular conditional probability $P^{\mathcal{B}}(\omega, \cdot)$ so that it depends on $\omega$ only through $H_N(\cdot, \omega)$.

It is evident that $H_N$ is a sufficient statistic for $\theta \varepsilon \Theta_0$ if only $(\mathbf{X}_1, \cdots, \mathbf{X}_m, \mathbf{Y}_1, \cdots, \mathbf{Y}_n)$ has been observed, in the sense that if $A$ is a Borel set in $[(R^p)]^m \times [(R^p)]^n$

$$P_\theta[(\mathbf{X}_1, \cdots, \mathbf{X}_m, \mathbf{Y}_1, \cdots, \mathbf{Y}_n) \varepsilon A \mid H_N]$$

can be chosen independent of $\theta$ for $\theta \varepsilon \Theta_0$. In fact if $\theta \varepsilon \Theta_0$ given $H_N$ there are a finite number of equally likely values that $(\mathbf{X}_1, \cdots, \mathbf{X}_m, \mathbf{Y}_1, \cdots, \mathbf{Y}_n)$ can assume.

In what follows we shall always assume that $P_\theta[\cdot \mid H_N]$ is computed under $\theta \varepsilon \Theta_0$ and we shall accordingly drop the subscript $\theta$ in all subsequent conditional laws, probabilities, and expectations.

Finally for any two distribution functions $F$ and $G$ on $R^p$ we define the Kolmogorov ($L_\infty$) distance between $F$ and $G$ as usual by,

$$d(F, G) = \sup_{\mathbf{t}} |F(\mathbf{t}) - G(\mathbf{t})|. \tag{2.6}$$

We can now define $\varphi_{m,n}$ by,

$$\begin{aligned} \varphi_{m,n}(\mathbf{X}_1(\omega), \cdots, \mathbf{X}_m(\omega), \mathbf{Y}_1(\omega), \cdots, \mathbf{Y}_n(\omega)) &= 1 \\ &\quad \text{if } d(F_m(\cdot, \omega), G_m(\cdot, \omega)) > c_N(H_N(\cdot, \omega)) \\ &= \gamma_N(H_N(\cdot, \omega)) \\ &\quad \text{if } d(F_n(\cdot, \omega), G_n(\cdot, \omega)) = c(H_N(\cdot, \omega)) \\ &= 0 \text{ otherwise,} \end{aligned} \tag{2.7}$$

where $c$ and $\gamma$ are uniquely determined by the condition that they be the smallest such numbers satisfying,

$$E\{\varphi(\mathbf{X}_1, \cdots, \mathbf{X}_m, \mathbf{Y}_1, \cdots, \mathbf{Y}_n) \mid H_N(\cdot, \omega)\} = \alpha. \tag{2.8}$$

This sequence clearly satisfies (2.1) and it follows from Theorems 3.1 and 3.2 of the next section that it also satisfies (2.2).

We remark that, at least in principle, the distribution of the test statistic under the null hypothesis can be tabled.

Suppose for simplicity that $F(\cdot)$ is continuous. We may assume that under $\theta_0 = (F, F)$ with probability one all the coordinates of the $\mathbf{X}$'s and $\mathbf{Y}$'s are distinct. Let

$$\mathbf{X}_i(\omega) = (X_{i1}(\omega), \cdots, X_{ip}(\omega))$$

and define $\mathbf{Y}_i$ similarly. Let $R_{ij}(\omega)$ be the rank of $X_{ij}(\omega)$ among the set of numbers $[\{X_{kj}(\omega)\}, 1 \leqq k \leqq m] \cup [\{Y_{kj}(\omega)\}: 1 \leqq k \leqq n]$.

Similarly define $S_{ij}(\omega)$ as the rank of $Y_{ij}$. Let $\mathbf{R}_i = (R_{i1}, \cdots, R_{ip})$ and define $\mathbf{S}_i$ similarly. Define

$$F_m^R(\mathbf{t}, \omega) = m^{-1} \sum_{i=1}^m I_{A(\mathbf{t})}(\mathbf{R}_i(\omega)).$$

Then if $H_N^R(\mathbf{t}, \omega)$, $G_n^R(\mathbf{t}, \omega)$ are defined in analogy to $H_N$ and $G_n$, clearly, under $\theta_0$, with probability one,

$$\mathfrak{L}(d(F_m, G_n) \mid H_N) = \mathfrak{L}(d(F_m^R, G_n^R) \mid H_N^R)) \tag{2.9}$$

where $\mathfrak{L}(Z)$ denotes the law of a random variable $Z$. Now $H_N^R$ takes on exactly $(N!)^{p-1}$ distinct values and for each value of $H_N^R$, $F_m^R$ takes on at most $\binom{N}{m}$ possible distinct values.

**3. Limiting behavior of $F_n$ given $H_N$ and statement of principal theorems.** We begin this section by introducing a more general framework for our results.

Let $P_{1N}, \cdots, P_{NN}$ be $N$ points in $R^p$, where we allow repetitions among the points.

In accord with our usage in the last section let $H_N$ be the distribution function of the measure which assigns mass $N^{-1}$ to each point $P_{iN}$.

Consider the experiment of choosing at random $m$ of the points $P_{1N}, \cdots, P_{NN}$. Again in accord with previous usage let $F_m$ be the distribution function of the measure which assigns mass $m^{-1}$ to each of the $m$ selected points, and let $G_n$ be the distribution function of the measure which assigns mass $n^{-1}$ to each of the $n$ remaining points where we suppose $m + n = N$. As we noted in the previous section if $P_{1N}, \cdots, P_{NN}$ denote the values of the "pooled" sample of $\mathbf{X}_1, \cdots, \mathbf{X}_m, \mathbf{Y}_1, \cdots, \mathbf{Y}_n$ where the $\mathbf{X}$'s and $\mathbf{Y}$'s originate from a common distribution $F$, then the conditional distribution of the empirical distribution function of the $\mathbf{X}$'s is precisely that of $F_m$ constructed above. However, for the first two theorems of this section we shall make no assumptions on how the points $P_{1N}, \cdots, P_{NN}$ are obtained. For convenience we introduce a suitable probability space $(\tilde{\Omega}, \tilde{\mathfrak{A}}, \tilde{P})$ with members $\tilde{\omega}$, and write $F_m(\mathbf{t}, \tilde{\omega})$ to indicate the dependence of $F_m$ on $\mathbf{t} \,\varepsilon\, R^p$ as a distribution function, and on the particular subset of $P_{1N}, \cdots, P_{NN}$ selected by the random experiment through $\tilde{\omega}$. In this formulation $H_N$ is a function of $\mathbf{t}$ only and is not stochastically determined. Of course,

$$G_n(\mathbf{t}, \tilde{\omega}) = n^{-1}\{NH_N(\mathbf{t}) - mF_m(\mathbf{t}, \tilde{\omega})\}. \tag{3.1}$$

The triple $(\tilde{\Omega}, \tilde{\mathfrak{A}}, \tilde{P})$ whose formal description we leave to the reader is so chosen that $F_m(\cdot, \cdot)$ is a separable stochastic process on $R^p$ whose finite dimensional joint laws are determined in accord with the experiment we have just described. We take $\{m\}$ and $\{N - m\}$ to be unbounded sequences of natural numbers whose dependence on $N$ we suppress. In all future theorems when we say $N \to \infty$ we mean that $m$ and $n = N - m$ both tend to $\infty$ as well. Without loss of generality we suppose that all the processes $F_m(\cdot, \cdot)$ are defined on the same probability space $(\tilde{\Omega}, \tilde{\mathfrak{A}}\, \tilde{P})$.

Define the stochastic process $X_N(\mathbf{t}, \tilde{\omega})$ on $(\tilde{\Omega}, \tilde{\mathfrak{A}})$ and $R^p$ by

$$\begin{aligned} X_N(\mathbf{t}, \tilde{\omega}) &= [mn/N]^{\frac{1}{2}}[F_m(\mathbf{t}, \tilde{\omega}) - G_n(\mathbf{t}, \tilde{\omega})] \\ &= [mn^{-1}]^{\frac{1}{2}}N^{\frac{1}{2}}[F_m(\mathbf{t}, \tilde{\omega}) - H_N(\mathbf{t}, \tilde{\omega})]. \end{aligned} \tag{3.2}$$

Let $H$ be the distribution function of a probability measure on $R^p$ (which we shall also refer to as $H$). We follow Dudley [5] and define $Q_H$ to be the set of real valued functions $\alpha(\cdot)$ on $R^p$ which:

(i) Are continuous except possibly on hyperplanes of the form $\{(t_1, \cdots t_p): t_j = \text{constant}\}$ which have positive $H$-measure.

(ii) On hyperplanes as above are continuous from below with limits from above, where "below" and "above" refer to strict inequality in all coordinates.

(iii) Have limits at points of the boundary of the compactification $\bar{R}^p$ of $R^p$ where $\bar{R} = [-\infty, +\infty]$.

Endow $Q_H$ with the supremum norm making it a separable Banach space and let $\Sigma_H$ be the topological Borel field of $Q_H$. Then ([5]) there exists a probability measure $\mu_H$ on $(Q_H, \Sigma_H)$ such that the map $X(\mathbf{t}, \alpha)$ defined for $\alpha \in Q_H$ and $\mathbf{t} \in R^p$ and given by,

$$X(\mathbf{t}, \alpha) = \alpha(\mathbf{t}), \tag{3.3}$$

is Gaussian with mean 0 and the stochastic process $X(\mathbf{t}, \cdot)$ has covariance structure given by,

$$\operatorname{Cov}(X(\mathbf{t}, \cdot), X(\mathbf{t}', \cdot)) = H(\mathbf{t} \wedge \mathbf{t}') - H(\mathbf{t})H(\mathbf{t}'), \tag{3.4}$$

where $\mathbf{t} \wedge \mathbf{t}'$ is the vector $(\min(t_1, t_1'), \cdots, \min(t_p, t_p'))$. Finally let $\mathfrak{M}(R^p)$ be the set of all bounded real valued functions on $R^p$ endowed with the supremum norm, $\|\cdot\|$. We require that $(\tilde{\Omega}, \tilde{\mathfrak{A}})$ have the property that $\{\tilde{\omega}: F_m(\cdot, \tilde{\omega}) = v\} \in \tilde{\mathfrak{A}}$ for every $v \in \mathfrak{M}(R^p)$. We shall say that a real valued functional $h$ on $\mathfrak{M}(R^p)$ is *regular for H* (a distribution function) if $h$ is continuous in the supremum norm at all points of a Borel subset $S$ of $Q_H$ such that $\mu_H(S) = 1$. Note that for any functional $h$ the function $h(X_N(\cdot, \tilde{\omega}))$ is measurable in $\tilde{\omega}$ in view of our assumptions on $(\tilde{\Omega}, \tilde{\mathfrak{A}})$. We can now state our main result.

THEOREM 3.1 *Suppose that* $\|H_N - H\| \to 0$ *for some distribution function H. Then, for every h regular for H,*

$$\Delta\{\mathfrak{L}(h(X_N(\cdot, \cdot))), \mathfrak{L}_H(h(X(\cdot, \cdot)))\} \to 0 \tag{3.5}$$

*as* $N \to \infty$, *where* $\mathfrak{L}(h(X_N(\cdot, \cdot)))$ *is the law of* $h(X_N(\cdot, \cdot))$ *under* $\tilde{P}$, $\mathfrak{L}_H(h(X(\cdot, \cdot)))$ *is the law of* $h(X(\cdot, \cdot))$ *under H and* $\Delta$ *is the Lévy-Prohorov distance between laws (distribution functions) on the real line.*

In the terminology of (Dudley [6]) the measures induced by the processes $X_N$ on $\mathfrak{M}(R^p)$ converge weak* to $\mu_H$. We postpone the proof of this theorem to the next four sections.

Suppose now that $\mathbf{X}_1, \cdots, \mathbf{X}_m, \mathbf{Y}_1, \cdots, \mathbf{Y}_n$ are random vectors defined as in the previous section and (with a slight abuse of notation) $F_m$, $G_n$, and $H_N$ are the empirical cumulative distribution functions of the **X**'s, the **Y**'s and the pooled sample respectively. Define the stochastic process $X_N$ in terms of $F_m$ and $G_n$ by equation (3.2). Note that these are all (including $H_N$) functions on $R^p \times \Omega$.

We shall say that a real valued functional $h$ is *regular for the two sample problem and* $H$ if $h$ is regular for $H$ and furthermore $h(X_N(\cdot, \omega))$ is a measurable function of $\omega$ on $\Omega$ for all $N$. Note that $\omega$ is permitted to vary in $H_N$ also with this definition. We have (using the notation of the previous section),

THEOREM 3.2. *Suppose that* $\theta = (F, F)$. *Let* $h$ *be a functional regular for the two sample problem and* $F$. *Then,*

$$(3.6) \qquad \Delta\{\mathcal{L}(h(X_N(\cdot, \cdot) \mid H_N)), \mathcal{L}_F(h(X(\cdot, \cdot)))\} \to 0 \qquad \text{a.s. } P_\theta .$$

PROOF. In view of our characterization of the conditional distribution of $F_m$ given $H_N$ if $\theta = (F, F)$ Theorem 3.2 is an immediate consequence of Theorem 3.1 and,

$$(3.7) \qquad P_{(F,F)}[\|H_N - F\| \to 0] = 1.$$

Assertion (3.7) is nothing else than the Glivenko-Cantelli theorem for vector variables. Theorem 3.2 is proved.

Put more prosaically our theorem states that for almost all sample sequences the conditional distributions of $h(X_N(\cdot, \cdot))$ given $H_N$ (which do not depend on $\theta$) converge to the same limiting distribution, that of $h(X(\cdot, \cdot))$, when $\mu_F$ holds on $(Q_F, \Sigma_F)$. The final dependence on $\theta$ of course appears because of the dependence of $H_N$ on $\theta$. It seems worth noting that the restriction of regularity for the two sample problem on $h$ may be dropped if the $\sigma$-field $\mathfrak{A}$ is completed with respect to $P_\theta$ since we have remarked that given $H_N$, $F_m$ and $G_n$ can take on only finitely many values and consequently the conditional distribution of $h$ given $H_N$ is well defined. We can now establish,

THEOREM 3.3. *The sequence of tests* $\{\varphi_{m,n}\}$ *satisfies* (2.1) *and* (2.2).

PROOF. As we noted previously, (2.1) clearly holds. Let $h(q) = \|q\|$. $h$ is continuous on all of $\mathfrak{M}(R^p)$ and easily satisfies (i).

Now,

$$(3.8) \qquad h(X_N(\cdot, \omega)) = \{mnN^{-1}\}^{\frac{1}{2}} \|F_m - G_n\|.$$

Then for every distribution function $H$ on $R^p$, $\alpha > 0$ there exists a unique finite smallest $c(H, \alpha)$, $0 \leqq \gamma(H, \alpha) \leqq 1$, such that,

$$(3.9) \quad \mu_H\{h(X(\cdot, \cdot)) > C(H, \alpha)\} + \gamma(H, \alpha)\mu_H\{h(X(\cdot, \cdot)) = c(H, \alpha)\} = \alpha.$$

From (3.9) and (3.6) we see that if $\theta = (F, F)$,

$$(3.10) \qquad \{mn(m + n)^{-1}\}^{\frac{1}{2}} c_N(H_N) \to c(F, \alpha) \qquad \text{a.s. } P_\theta$$

where $c_N(H_N)$ is given by (2.7), (2.8).

Now, suppose $\theta = (F, G)$, $F \neq G$. Without loss of generality let us assume that

$mN^{-1} \to \lambda$, where $0 \leqq \lambda \leqq 1$. Define,

$$H = \lambda F + (1 - \lambda)G. \tag{3.11}$$

By the Glivenko-Cantelli theorem in this case,

$$P_\theta[\|F_m - F\| \to 0] = P_\theta[\|G_n - G\| \to 0] = 1 \tag{3.12}$$

and hence,

$$P_\theta[\|H_N - H\| \to 0] = 1. \tag{3.13}$$

An application of Theorem 3.1 identical to that used to obtain Theorem 3.2 and (3.10) now enables us to conclude that,

$$\{mn(m + n)^{-1}\}^{\frac{1}{2}} c_N(H_N) \to c(H, \alpha) \qquad \text{a.s. } P_\theta . \tag{3.14}$$

On the other hand by (3.12),

$$\|F_m - G_n\| \to \|F - G\| > 0 \qquad \text{a.s. } P_\theta . \tag{3.15}$$

Of course, (3.14) and (3.15) imply that,

$$\varphi_{m,n}(\mathbf{X}_1, \cdots, \mathbf{Y}_1, \cdots) \to 1 \qquad \text{a.s. } P_\theta \tag{3.16}$$

if $F \neq G$. Theorem 3.3 follows.

There are many possible choices of $\varphi_{m,n}$ satisfying (2.1) and (2.2), some perhaps based on statistics having a limiting distribution independent of $F$ if $\theta = (F, F)$. Certainly plausible statistics whose distribution is more invariant than that of $d(F_m, G_n)$ can readily be defined (see for instance Weiss [20] and Vincze [18]). We hope to investigate this question in a later paper. We remark that Theorems 3.2 and 3.3 are of some interest even in the case $p = 1$ since the test $\varphi_{m,n}$ which for $F$ continuous is the usual Smirnov test is also distribution free when the underlying distribution is discrete, viz. when there are ties. Unfortunately, and this is, of course, the principal difficulty in higher dimensions, the limiting value of the cut-off point of the test depends on $F$ if $F$ is not continuous.

From Theorem 3.3 we immediately deduce,

THEOREM 3.4. *If $h$ satisfies the hypotheses of Theorem* 3.2 *and* $\theta = (F, F)$ *then*

$$\Delta\{\mathcal{L}_\theta(h(X_N(\cdot, \cdot))), \mathcal{L}_F(h(X(\cdot, \cdot)))\} \to 0. \tag{3.17}$$

PROOF. Let,

$$\Psi_N(s, H_N) = E(\exp is\, h(X_N(\cdot, \cdot)) \mid H_N). \tag{3.18}$$

Then,

$$P_\theta[\Psi_N(s, H_N) \to E_F\{\exp is\, h(X(\cdot, \cdot))\} \ \forall\, s] = 1. \tag{3.19}$$

Theorem 3.4 follows upon applying the dominated convergence theorem to the variables $\Psi_N(s, H_N)$. From Theorem 3.4 we can derive Dudley's Theorem 2 [6]. Let,

$$\tilde{X}_m(\mathbf{t}, \omega) = m^{\frac{1}{2}}(F_m(\mathbf{t}, \omega) - F(\mathbf{t}, \omega)). \tag{3.20}$$

We shall say *h is regular for the one sample problem and F* if $h$ is regular for $F$ and $h(X_m(\cdot, \omega))$ is measurable in $\omega$.

THEOREM 2. *If h is regular for the one sample problem and F and* $\theta = (F, F)$ *then,*

$$\mathcal{L}_\theta\{h(\tilde{X}_m(\cdot, \cdot))\} \rightarrow \mathcal{L}_{\mu_F}\{h(X(\cdot, \cdot))\} \quad \text{as} \quad m \rightarrow \infty. \tag{3.21}$$

PROOF. By Proposition 1 of [6] we need only consider $h$ continuous on $\mathfrak{M}(R^p)$. For such an $h$, if $\theta = (F, F)$,

$$\mathcal{L}_\theta\{h(X_N(\cdot, \cdot))\} \rightarrow \mathcal{L}_\theta\{h(X_m(\cdot, \cdot))\} \qquad \text{as } n \rightarrow \infty, \tag{3.22}$$

($m$ fixed) by the Glivenko-Cantelli theorem. Therefore any limit law stochastic or substochastic of the sequence $\mathcal{L}_\theta\{h(\tilde{X}_m(\cdot, \cdot))\}$ as $m \rightarrow \infty$ is a limit law of the sequence $\mathcal{L}_\theta\{h(X(\cdot, \cdot))\}$ as $m$, $n$ both $\rightarrow \infty$. Theorem 2 follows. Of course, Theorem 3.3 could have been derived from Theorem 2.

From Theorem 2 one can, of course, derive Theorem 2 of [11] by applying the former to the functionals $h(q) = \|q\|$ and $h(q) = \|\max(q, 0)\|$. The limiting law of our test statistics $\{mn(m + n)^{-1}\}^{\frac{1}{2}} \|F_m - G_n\|$ under $\theta = (F, F)$ given $H_N$ is then identified as being the same as the distribution given in [11], and Theorem 1 of [11] and 1-$m$ of [10] may be employed to give some information about $c(F, \alpha)$. It seems plausible that one can obtain bounds similar to those of [10] and [11] for the conditional distribution of $\{mn(m + n)^{-1}\}^{\frac{1}{2}}d(F_m, G_n)$ in terms of $H_N$ but we have not attempted to do so. Finally, we note that Theorems 2 and 3.2 imply that the expressions for the limit law of $m^{\frac{1}{2}}\|F_m - F\|$ if $p = 1$ given by (Schmid [15]) apply in our case also.

**4. Some results on sampling without replacement.** The lemmas of this section play a fundamental role in the proof of Theorem 3.1 given in the next three sections. In order to place Lemmas 4.6 and 4.7 in the scheme of things it may be worthwhile to read Sections 6 and 7 first.

Let $V_{1N}, \cdots, V_{NN}$ be a sample taken in order from a finite population $c_{1N}, \cdots, c_{NN}$ without replacement, viz. $P[(V_{1N}, \cdots, V_{NN}) = (c_{i_1N}, \cdots, c_{i_NN})] = [N!]^{-1}$ for every permutation $(i_1, \cdots, i_N)$ of $(1, \cdots, N)$. Suppose $\sum_{i=1}^N c_{iN} = 0$. Then, we have,

LEMMA 4.1. (Hájek [8]). *Let* $V_{1N}, \cdots, V_{NN}$, $\{c_{iN}\}$ *be as above. Let* $\{d_{iN}\}$, $1 \leqq i \leqq N$, *be another sequence of constants such that* $\sum_{i=1}^N d_{iN} = 0$.

*Suppose* $N^{-1}\sum_{i=1}^N c_{iN}^2 \rightarrow \sigma^2$, $\sum_{i=1}^N d_{iN}^2 \rightarrow K$. *Then,*

$$\mathcal{L}(\textstyle\sum_{i=1}^N d_{iN}V_{iN}) \rightarrow \eta(0, \sigma^2K) \tag{4.1}$$

*if,*

$$\text{(a)}\quad N^{-\frac{1}{2}}\max|c_{iN}| \rightarrow 0, \qquad \text{(b)}\quad \max|d_{iN}| \rightarrow 0, \tag{4.2}$$

*and*

$$N^{-1}\textstyle\sum_{\{|\delta_{Nij}|>\tau\}}\delta_{Nij}^2 \rightarrow 0 \tag{4.3}$$

*for every* $\tau > 0$, *where* $\delta_{Nij} = c_{iN}\, d_{jN}$.

Another useful result also due to (Hájek [9]) is,

LEMMA 4.2. (Hájek). *Let* $\{V_{iN}\}$, $\{c_{iN}\}$ *be as above. Suppose* $S_j = \sum_{i=1}^{j} V_{iN}$. *Then,*

$$P[\max_{1 \leqq j \leqq N} |S_j| \geqq \epsilon] \leqq E|S_n|^r \{\epsilon(1 - nN^{-1})\}^{-r} \tag{4.4}$$

*for every* $r \geqq 0$, $1 \leqq n \leqq N$, $\epsilon > 0$.

This lemma is of little value if $n$ is close to $N$. To deal with that situation we need inequalities of the type given in Lemma 4.3.

LEMMA 4.3. *Under the assumption of Lemma* 4.2 *suppose* $kl \leqq n < k(l + 1)$ *where* $k \geqq 2$, $N \geqq n$, $l \geqq 1$. *Then there exists* $c(k, r)$ *such that,*

$$P[\max_{1 \leqq j \leqq n} |S_j| \geqq \epsilon] \leqq c(k, r)\epsilon^{-r} E|S_l|^r. \tag{4.5}$$

PROOF. Suppose $n = kl + d$. Then,

$$P[\max_{1 \leqq j \leqq n} |S_j| \geqq \epsilon] \leqq P[\max_{1 \leqq j \leqq kl} |S_j| \geqq \epsilon/2] + P[\max_{kl < j \leqq n} |S_{kl} + (S_j - S_{kl})| \geqq \epsilon/2]. \tag{4.6}$$

But,

$$P[\max_{kl < j \leqq n} |S_{kl} + (S_j - S_{kl})| \geqq \epsilon/2] \leqq P[|S_{kl}| \geqq \epsilon/4] + P[\max_{1 \leqq j \leqq d} |S_j| \geqq \epsilon/4]. \tag{4.7}$$

Hence, if $0 \leqq d < k \leqq kl$ we have,

$$P[\max_{1 \leqq j \leqq n} |S_j| \geqq \epsilon] \leqq 3P[\max_{1 \leqq j \leqq kl} |S_j| \geqq \epsilon/4], \tag{4.8}$$

and we need establish the lemma only for $n = kl$. We proceed by induction. If $k = 2$,

$$P[\max_{1 \leqq j \leqq 2l} |S_j| \geqq \epsilon] \leqq P[\max_{1 \leqq j \leqq l} |S_j| \geqq \epsilon/2] + P[|S_l| \geqq \epsilon/4] + P[\max_{1 \leqq j \leqq l} |S_j| \geqq \epsilon/4] \leqq 3P[\max_{1 \leqq j \leqq l} |S_j| \geqq \epsilon/4] \leqq 3.8^r \epsilon^{-r} E|S_l|^r \tag{4.9}$$

by Lemma 4.2. The induction step is given by (4.6) and (4.7) with $d = l$.

We will also need,

LEMMA 4.4. (Wald-Wolfowitz-Noether-Hoeffding.) *Suppose the* $c_{iN}$ *satisfy the hypotheses of Lemma* 4.1 *as well as* (4.2) (a), $S_n$ *is as in Lemma* 4.2, *and* $n/N \to \lambda$, $0 < \lambda < 1$. *Then as* $N \to \infty$,

$$E(N^{-k} S_n^{2k}) \to \lambda^k \sigma^{2k} \mu_{2k} \tag{4.10}$$

*where* $\mu_{2k}$ *is the* $2k$th *central moment of the standard normal distribution.*

A proof of Lemma 4.4 may be found in [7] pp 237–239.

Before giving the main results of this section we state and prove an auxiliary lemma. Our techniques in this and the succeeding lemma are similar to those used by Hájek in [9] and by Kiefer and Wolfowitz in [11].

LEMMA 4.5. *Let* $\{V_{iN}\}$, $1 \leqq i \leqq N$, *be as above. Then,*

$$P[\textstyle\sum_{k=r+1}^{r+m} V_{kN} < g(V_{1N}, \cdots, V_{rN}) \mid V_{1N}, \cdots, V_{rN}] \leqq [g(V_{1N}, \cdots, V_{rN}) + mS_r(N-r)^{-1}]^{-2} m(1 - m(N-r)^{-1})(N-r)^{-1} \textstyle\sum_{i=1}^{N} c_{iN}^2 \tag{4.11}$$

*for any* $g(V_{1N}, \cdots, V_{rN})$ *such that* $g(V_{1N}, \cdots, V_{rN}) + mS_r(N-r)^{-1} < 0$.

PROOF. Given $V_{1N}, \cdots, V_{rN}$ the $V_{kN}$ with $k > r$ are distributed as a sample without replacement from the population of size $N - r$ whose members $c_{iN}'$, $i = 1, \cdots, N - r$, are the $c_{iN}$ which have not appeared among the $V_{iN}$ with $i \leqq r$. Then

$$E(V_{(r+1)N} \mid V_{iN}, 1 \leqq i \leqq r) = -S_r(N-r)^{-1}. \tag{4.12}$$

Chebyshev's inequality applied to the left-hand side of (4.11) gives,

$$P[\textstyle\sum_{k=i+1}^{r+m} V_{kN} \leqq g(V_{1N}, \cdots, V_{rN}) \mid V_{1N}, \cdots, V_{rN}] \leqq [g(V_{1N}, \cdots, V_{rN}) + mS_r(N-r)^{-1}]^{-2} (1 - (m-1)/(N-r-1))m \operatorname{Var}\{V_{r+1,N} \mid V_{1N}, \cdots, V_{rN}\}. \tag{4.13}$$

Finally,

$$\begin{aligned} &\operatorname{Var}\{V_{r+1,N} \mid V_{1N}, \cdots, V_{rN}\} \\ &= (N-r)^{-1}\{\textstyle\sum_{i=1}^{N} (c_{iN} + S_r(N-r)^{-1})^2 - \sum_{i=1}^{r} (V_{iN} + S_r(N-r)^{-1})^2\} \\ &= (N-r)^{-1}\{\textstyle\sum_{i=1}^{N} c_{iN}^2 - \sum_{i=1}^{r} V_{iN}^2 - S_r^2(N-r)^{-1}\} \\ &\leqq (N-r)^{-1} \textstyle\sum_{i=1}^{N} c_{iN}^2 . \end{aligned} \tag{4.14}$$

□

The final results of this section are further extensions of Lemmas 4.2 and 4.3 which are central to our argument. Let $S$ be a fixed subset with $n$ members ($n \leqq N$) of the set of all ordered $p$-tuples $(i_1, \cdots, i_p)$ where the $i_j$ may range independently over $1, \cdots, N$. Now, relabel $V_{1N}, \cdots, V_{nN}$ as $V_{(i_1,\cdots,i_p)}$ where $(i_1, \cdots, i_p)$ ranges over $S$. We shall call such sets $S$, *configurations*. Define,

$$S_{(j_1,\cdots,j_p)} = \textstyle\sum_{(i_1,\cdots,i_p)\varepsilon[A(\mathbf{j})]\cap S} V_{(i_1,\cdots,i_p)} \tag{4.15}$$

where $\mathbf{j} = (j_1, \cdots, j_p)$ is a vector with natural number entries and $A(\mathbf{j})$ is defined in Section 2.

Define,

$$a(\gamma, \epsilon) = [1 - n(N-n)^{-1} - \gamma]\epsilon \tag{4.16}$$

and

$$b(\gamma, \epsilon) = 1 - n(N-n)^{-1}[a(\gamma, \epsilon)]^{-2} \textstyle\sum_{i=1}^{N} c_{iN}^2 . \tag{4.17}$$

Choose $\gamma$ so that $\gamma \leqq 1 - n(N-n)^{-1}$ and $b(\gamma, \gamma^s\epsilon) > 0$ for $0 \leqq s \leqq p - 2$.

We have,

LEMMA 4.6 *Under the above assumptions, if $p \geqq 2$, $n < N$, and $r > 0$,*

$$(4.18)\quad P[\max\{|S_{(i_1,\cdots,i_p)}|: 1 \leqq i_j \leqq N, j = 1, \cdots, p\} \geqq \epsilon] \leqq \{\textstyle\prod_{s=0}^{p-2} b^{-1}(\gamma, \gamma^s\epsilon)\}\, E|S_n|^r[\gamma^{p-1}\epsilon]^{-r}.$$

PROOF. We prove the lemma for $p = 2$. The general case follows by an induction which we leave to the reader.

Let $t_1$ be the smallest $i$ such that $S_{(i,j)} \geqq \epsilon$ for some $1 \leqq j \leqq N$ if such an $i$ exists. Otherwise let $t_1 = \infty$. Let $t_2$ be the largest $j$, $1 \leqq j \leqq N$, such that $S_{(t_1,j)} \geqq \epsilon$, if $t_1 < \infty$. Otherwise let $t_2 = \infty$. Then,

$$\begin{aligned} &P[\max\{S_{(i,j)} : 1 \leqq i \leqq N, 1 \leqq j \leqq N\} \geqq \epsilon] \\ (4.19)\quad &= \textstyle\sum_{k=1}^N P[t_1 = k] \\ &\leqq \textstyle\sum_{k=1}^N P[t_1 = k, S_{(N,t_2)} < \gamma\epsilon] + P[S_{(N,t_2)} \geqq \gamma\epsilon, t_1 < \infty] \\ &\leqq \textstyle\sum_{k=1}^N P[S_{(N,t_2)} < \gamma\epsilon, t_1 = k] + P[\max_{1\leqq j\leqq N} S_{(N,j)} \geqq \gamma\epsilon]. \end{aligned}$$

But,

$$\begin{aligned} (4.20)\quad &P[S_{(N,t_2)} < \gamma\epsilon, t_1 = k] \\ &= \int_{[t_1=k]} P[S_{(N,t_2)} < \gamma\epsilon \mid V_{(i,j)}, 1 \leqq i \leqq k, 1 \leqq j \leqq N, (i,j) \,\varepsilon\, S]\, dP \end{aligned}$$

where $P[\cdot \mid \cdot]$ denotes as usual a conditional probability. For convenience denote the conditioning event in (4.20) by $A$.

Let $n(i, j)$ be the number of summands in $S_{(i,j)}$ and define,

$$(4.21)\quad \rho(i, j) = (n(N, j) - n(i, j))[N - n(i, N)]^{-1}.$$

By definition of $t_2$ we have $S_{(t_1,N)} \leqq S_{(t_1,t_2)}$ and therefore,

$$\begin{aligned} (4.22)\quad P[S_{(N,t_2)} < \gamma\epsilon \mid A] &\leqq P[(S_{(N,t_2)} - S_{(t_1,t_2)} \\ &\quad + S_{(t_1,N)}\rho(t_1, t_2) < \gamma\epsilon - (1 - \rho(t_1, t_2))S_{(t_1,t_2)} \mid A]. \end{aligned}$$

Applying Lemma 4.5 to the right-hand side of (4.22) we obtain,

$$\begin{aligned} &P[S_{(N,t_2)} < \gamma\epsilon \mid A] \\ &\leqq [\gamma\epsilon - (1 - \rho(t_1, t_2))S_{(t_1,t_2)}]^{-2}[1 - (n(N, t_2) \\ (4.23)\quad &- n(t_1, t_2) - 1)(N - n(t_1, N) - 1)^{-1}](N - n(t_1, N))^{-1} \\ &\{\textstyle\sum_{i=1}^N c_{iN}^2\}(n(N, t_2) - n(t_1, t_2)) \\ &\qquad\qquad \text{if } \gamma\epsilon - (1 - \rho(t_1, t_2))S_{(t_1,t_2)} < 0. \end{aligned}$$

But,

$$(4.24)\quad 0 \leqq \rho(i, j) \leqq n[N - n]^{-1}$$

and $S_{(t_1,t_2)} \geqq \epsilon$. From these remarks we conclude,

$$P[S_{(N,t_2)} < \gamma\epsilon \mid A]$$

$$\leqq \{[\gamma - (1 - n(N-n)^{-1})]\epsilon\}^{-2} n(N-n)^{-1} \textstyle\sum_{i=1}^N c_{iN}^2, \quad \text{for } \gamma \leqq 1 - n(N-n)^{-1}, \tag{4.25}$$

Substituting (4.25) in (4.20) we get,

$$P[S_{(N,t_2)} < \gamma\epsilon, t_1 = k] \leqq P[t_1 = k]\{[\gamma - (1 - n(N-n)^{-1})]\epsilon\}^{-2} n(N-n)^{-1} \textstyle\sum_{i=1}^N c_{iN}^2. \tag{4.26}$$

Finally, from (4.19) and (4.26)

$$P[\max\{S_{(i,j)} : 1 \leqq i \leqq N, 1 \leqq j \leqq N\} \geqq \epsilon]$$

$$\leqq P[\max_{1 \leqq j \leqq N} S_{(N,j)} \geqq \gamma\epsilon] + [a(\gamma, \epsilon) n (N-n)^{-1} \textstyle\sum_{i=1}^N c_{iN}^2] \cdot P[\max\{S_{(i,j)} : 1 \leqq i \leqq N, 1 \leqq j \leqq N\} \geqq \epsilon]. \tag{4.27}$$

Applying (4.27) to sampling from the population $\{-c_{1N}, \cdots, -c_{NN}\}$ adding the resulting inequalities and transposing we obtain,

$$P[\max\{|S_{(i,j)}| : 1 \leqq i \leqq N, 1 \leqq j \leqq N\} \geqq \epsilon] \leqq b^{-1}(\gamma, \epsilon) P[\max_{1 \leqq j \leqq N} |S_{(N,j)}| \geqq \gamma\epsilon]. \tag{4.28}$$

The proof for $p = 2$ is finished upon applying Lemma 4.2 to the right hand side of (4.28).□

It is important to note that the right hand side of (4.18) depends only on the $c_{iN}$ and $n$ and is independent of the configuration chosen.

Finally Lemma 4.7 extends Lemma 4.3. We use the notation of Lemma 4.6.

LEMMA 4.7. *Under the assumptions of Lemma* 4.6 *there exist* $c(k, r)$, $\lambda(k)$ *such that, if* $kl \leqq n < k(l+1)$, $k \geqq 3$, $r > 0$, *and* $l \geqq 1$, *then*

$$P[\max\{|S_{(i_1,\cdots,i_p)}| : 1 \leqq i_j \leqq N, j = 1, \cdots, p\} \geqq \epsilon] \leqq c(k, r)(1 - \lambda(k)\epsilon^{-2} \textstyle\sum_{i=1}^N c_{iN}^2)^{-(p-1)} \epsilon^{-r} E|S_l|^r. \tag{4.29}$$

PROOF. With our new notation the argument is in fact simpler than that given for Lemma 4.3. Suppose $n = kl + d$, with $0 \leqq d < k \leqq l$. Divide $S$ into $(k+1)$ subsets $S^{(1)}, \cdots, S^{(k+1)}$ such that $S^{(1)}, \cdots, S^{(k)}$ have $l$ members each while $S^{(k+1)}$ has $d$ members. Define,

$$S_{(j_1,\cdots,j_p)}^{(t)} = \textstyle\sum_{(i_1,\cdots,i_p)\epsilon[A(\mathbf{j})]\cap S^{(t)}} V_{(i_1,\cdots,i_p)} \tag{4.30}$$

where $\mathbf{j} = (j_1, \cdots, j_p)$ for $t = 1, \cdots, k+1$. Then,

$$S_{(i_1,\cdots,i_p)} = \textstyle\sum_{j=1}^{k+1} S_{(i_1,\cdots,i_p)}^{(j)}, \tag{4.31}$$

and

$$P[\max\{|S_{(i_1,\cdots,i_p)}| : 1 \leqq i_j \leqq N, j = 1, \cdots, p\} \geqq \epsilon] \leqq \textstyle\sum_{t=1}^{k+1} P[\max\{|S_{(i_1,\cdots,i_p)}^{(t)}| : 1 \leqq i_j \leqq N, j = 1, \cdots, p\} \geqq \epsilon(k+1)^{-1}]. \tag{4.32}$$

Since each $S^{(t)}$ is a configuration with $\leqq l$ members, we have for suitable $\gamma$, by Lemma 4.6,

$$(4.33)\quad \begin{aligned} &P[\max\{|S_{(i_1,\cdots,i_p)}|: 1 \leqq i_j \leqq N, j = 1, \cdots, p\} \geqq \epsilon] \\ &\leqq (k+1)\,[\textstyle\prod_{s=0}^{p-2} (1-(k-1)^{-1}) \\ &\quad [(1-(k-1)^{-1}-\gamma)\,\gamma^s\epsilon/(k+1)]^{-2} \textstyle\sum_{i=1}^N c_{iN}^2)^{-1}] \\ &\quad \cdot\{E|S_l|^r\}\gamma^{-r(p-1)}\epsilon^{-r}(k+1)^r. \end{aligned}$$

(It is clear that the bound on the right hand side of (4.18) continues to be valid if the number of members of $S$ is $\leqq n$.) Since $k \geqq 3$ we may take $\gamma = \frac{1}{4}$ and hence $\lambda(k) = [16(k-1)(k+1)^2]^{-1}$ and $c(k, r) = (k+1)^{r+1}4^{r(p-1)}$ will do. If $l < k$ a further obvious decomposition of $S^{(k+1)}$ will complete the proof.

Both Lemma 4.3 and this lemma will be used for $k = 3$ only.

**5. Proof of Theorem 3.1: Initial considerations and enumeration of cases.** It follows from the work of Dudley (Theorem 1 and Proposition 1 of [6]) that to establish Theorem 3.1 we need only show,

(i) *Finite dimensional convergence*:

$$(5.1)\quad \mathcal{L}(X_N(\mathbf{t}^{(1)}, \cdot), \cdots, X_N(\mathbf{t}^{(r)}, \cdot)) \to \mathcal{L}_H(X(\mathbf{t}^{(1)}, \cdot), \cdots, X(\mathbf{t}^{(r)}, \cdot))$$

*as* $N \to \infty$, *for all* $\mathbf{t}^{(1)}, \cdots, \mathbf{t}^{(r)} \in R^p$. (As usual, $\mathcal{L}(X_N(\mathbf{t}^{(1)}, \cdot), \cdots)$ is the joint law of the variable $X_N(\mathbf{t}^{(1)}, \cdot), \cdots, X_N(\mathbf{t}^{(r)}, \cdot)$ when $\tilde{P}$ holds while $\mathcal{L}_H(X(\mathbf{t}^{(1)}, \cdot), \cdots)$ is the law of the variables $X(\mathbf{t}^{(1)}, \cdot), \cdots, X(\mathbf{t}^{(r)}, \cdot)$ when $\mu_H$ is the measure on $Q_H$.)

(ii) *Tightness*: For every $\epsilon > 0$, $\delta > 0$, there exists a compact subset $K$ of $Q_H$ such that,

$$(5.2)\quad \tilde{P}[\tilde{\omega}: X_N(\cdot, \tilde{\omega}) \in K^\delta] \geqq 1 - \epsilon$$

*for all* $N$ *sufficiently large* ($m$, $n$ *large*), *where*

$$K^\delta = \{f \in \mathfrak{M}(R^p): \|f - g\| \leqq \delta \quad \text{for some} \quad g \in K\}.$$

(We need not introduce outer probabilities because our processes $X_N$ take on only a finite number of values.)

Proof of (i). If $A$ is a Borel subset of $R^p$ let us define

$$(5.3)\quad F_m(A, \tilde{\omega}) = \textstyle\iint_A dF_m(\mathbf{t}, \tilde{\omega}).$$

Similarly, we shall talk about $G_n(A, \cdot)$, $H_N(A, \cdot)$ and $H(A)$. Then, (5.1) will follow if we show that for any $s$ disjoint rectangles $A_1, \cdots, A_s$ of $R^p$ such that $0 < H(A_i) < 1$, $1 \leqq i \leqq s$, we have,

$$(5.4)\quad \mathcal{L}(X_N)(A_1, \cdot), \cdots, X_N(A_s, \cdot)) \to \mathcal{L}(X(A_1), \cdots, X(A_s)) \qquad \text{as } m, n \to \infty,$$

where $X(A_1), \cdots, X(A_s)$ are Gaussian with mean 0 and Cov $(X(A_i), X(A_j)) = H(A_i \cap A_j) - H(A_i)H(A_j)$.

Let

$$(5.5) \qquad Z_i = (nm^{-1})^{\frac{1}{2}} (m+n)^{-\frac{1}{2}} \quad \text{if } P_{iN} \text{ has been selected among the sample of } m,$$
$$= -(mn^{-1})^{\frac{1}{2}} (m+n)^{-\frac{1}{2}}, \qquad \text{otherwise.}$$

Then,

$$(5.6) \quad \tilde{P}[X_N(A_i, \cdot) = q_i, 1 \leqq i \leqq s] = \tilde{P}[\textstyle\sum_{P_{jN} \varepsilon A_i} Z_j = q_i, 1 \leqq i \leqq s].$$

Let

$$(5.7) \qquad c_{iN} = (nm^{-1})^{\frac{1}{2}}, \qquad 1 \leqq i \leqq m,$$
$$= -(mn^{-1})^{\frac{1}{2}}, \qquad m+1 \leqq i \leqq m+n.$$

Let $V_{1N}, \cdots, V_{NN}$ be obtained as in Lemma 4.2. Then,

$$\tilde{P}[\textstyle\sum_{P_{jN} \varepsilon A_i} Z_j = q_i, 1 \leqq i \leqq s]$$
$$(5.8) \qquad = P[N^{-\frac{1}{2}} \textstyle\sum_{i=1}^{NH_N(A_i)} V_{iN} = q_1, \cdots,$$
$$N^{-\frac{1}{2}} \textstyle\sum_{i=NH_N(\Sigma_1^{k-1} A_j)+1}^{NH_N(\Sigma_1^k A_j)} V_{iN} = q_k, \cdots,$$
$$N^{-\frac{1}{2}} \textstyle\sum_{i=NH_N(\Sigma_1^{s-1} A_j)+1}^{NH_N(\Sigma_1^s A_j)} V_{iN} = q_s].$$

From (5.8)

$$(5.9) \qquad \mathcal{L}(\textstyle\sum_{i=1}^s \lambda_i X_N(A_i, \cdot)) = \mathcal{L}(\textstyle\sum_{i=1}^N d_{iN} V_{iN})$$

where $d_{iN} = N^{-\frac{1}{2}}(\lambda_k - \sum_{k=1}^s \lambda_k H_N(A_k))$ if $NH_N(\sum_{j=1}^{k-1} A_j) \leqq i \leqq NH_N(\sum_{j=1}^k A_j)$ for $1 \leqq k \leqq s$,

$= 0$ otherwise.

It readily follows that $c_{iN}$, $d_{iN}$ as given satisfy (4.2) and (4.3). Then (5.4) is a consequence of (5.9) and Lemma 4.2.

Proof of (ii). We prove assertion (ii) in four steps. We shall show that the assertion holds,

(a) For $H$ which are continuous and such that at least one marginal of $H$ is the uniform distribution on [0, 1];

(b) For $H$ which are continuous and such that at least one marginal of $H$ is strictly increasing in [0, 1];

(c) For $H$ which are continuous;

(d) For arbitrary $H$.

The next section deals with cases (a) and (b) while the final section has cases (c) and (d).

**6. Proof of Theorem 3.1: Cases (iia) and (iib).** We note first that in all cases it suffices to prove the result for sequences $H_N$ such that all the $P_{iN}$ fall in $I^p = \{\mathbf{t}: ||t|| \leqq 1\}$ where $||t|| = \max_{1 \leqq i \leqq p} |t_i|$. To see this let $q$ be any $1 - 1$ map of $R$ onto $(0, 1)$ which is a homeomorphism.

If $\{P_{iN}\}$, $\{H_N\}$, $H$ are given let $\{P_{iN}^*\}$ be the sequence of points in $I^p$ obtained by transforming each coordinate of $P_{iN}$ by $q$, and let $H_N^*$ be the associated sequence of distribution functions. Clearly if $||H_N - H|| \to 0$, then $||H_N^* - H^*|| \to 0$ where $H^*$ is the distribution function of a measure concentrated on $I^p$ such that, if $q^{-1}(\mathbf{t}) = (q^{-1}(t_1), \cdots, q^{-1}(t_p))$,

$$H^*(\mathbf{t}) = H(q^{-1}(\mathbf{t})) \tag{6.1}$$

for $\mathbf{t}$ in the interior of $I^p$. It is easy to see now that if $h$ is any functional regular for $H$, then there exists a functional $h^*$ regular for $H^*$ such that,

$$\mathcal{L}(h(X_N(\cdot, \cdot))) = \mathcal{L}(h^*(X_N^*(\cdot, \cdot))) \tag{6.2}$$

and

$$\mathcal{L}_H(h(X(\cdot, \cdot))) = \mathcal{L}_{H^*}(h^*(X(\cdot, \cdot))). \tag{6.3}$$

Our assertion follows and in the sequel we shall limit ourselves to sequences $\{P_{iN}\}$ falling in $I^p$. For cases (a), (b) and (c), we conclude from Proposition 2 of [6] that, in view of (i), to establish (ii) we need only show that,

$$\limsup_{\delta \to 0} \limsup_N P[\sup \{|X_N(\mathbf{s}, \cdot) - X_N(\mathbf{t}, \cdot)|: ||\mathbf{s} - \mathbf{t}|| \leqq \delta, \mathbf{s}, \mathbf{t} \in I^p\} \geqq \epsilon] = 0 \quad \text{for every} \quad \epsilon > 0. \tag{6.4}$$

PROOF OF (a). We need a lemma.

LEMMA 5.1 *Let* $A(\mathbf{s}, \delta) = \{\mathbf{t} \in I^p: \mathbf{t} \geqq \mathbf{s}, ||\mathbf{t} - \mathbf{s}|| \leqq \delta\}$. *Then* (6.4) *holds if*

$$\lim_{\delta \to 0} \delta^{-p} \limsup_N \sup_{\mathbf{s} \in I^p} \tilde{P}[\sup \{|X_N(\mathbf{s}, \cdot) - X_N(\mathbf{t}, \cdot)|: \mathbf{t} \in A(\mathbf{s}, \delta)\} \geqq \epsilon] = 0. \tag{6.5}$$

PROOF. Without loss of generality consider the particular sequence $\delta_m = m^{-1} \to 0$. Decompose $I^p$ into the $m^p$ cubes whose vertices are lattice points of the form $m^{-1}(j_1, \cdots, j_p)$, $0 \leqq j_i \leqq m$. Then,

$$\begin{aligned} &\tilde{P}[\sup \{|X_N(\mathbf{s}, \cdot) - X_N(\mathbf{t}, \cdot)|: ||\mathbf{s} - \mathbf{t}|| \leqq \delta, \mathbf{s}, \mathbf{t} \in I^p] \geqq \epsilon] \\ &\leqq \tilde{P}[\textstyle\bigcup_{j=1}^{m^p}[\sup |X_N(\mathbf{e}_j, \cdot) - X_N(\mathbf{s}, \cdot)|: \mathbf{s} \in A(\mathbf{e}_j, \delta)\} \geqq \epsilon/2] \\ &\leqq \textstyle\sum_{j=1}^{m^p} \tilde{P}[\sup \{|X_N(\mathbf{e}_j, \cdot) - X_N(\mathbf{s}, \cdot)|: \mathbf{s} \in A(\mathbf{e}_j, \delta)\} \geqq \epsilon/2] \end{aligned} \tag{6.6}$$

where $\mathbf{e}_1, \cdots, \mathbf{e}_{m^p}$ are the "smallest" vertices of the $m^p$ cubes we are considering.□

Let $B(\mathbf{s}, \mathbf{t}) = \{\mathbf{v}: \mathbf{s} < \mathbf{v} \leqq \mathbf{t}\}$ if $\mathbf{t} > \mathbf{s}$, where $<$ corresponds to strict inequality for all coordinates. More generally, let $T$ be the set of all sub collections of the natural numbers $1, \cdots, p$. Denote such a collection by $\{i_1, \cdots, i_r\}$. If $\mathbf{t} \geqq \mathbf{s}$ let

$$B(\mathbf{s}, \mathbf{t}, \{i_1, \cdots, i_r\}) = \{\mathbf{v} \,\varepsilon\, I^p: \mathbf{v} = (v_1, \cdots, v_p),$$
$$\mathbf{v} \leqq \mathbf{t}, v_k > s_k \text{ if } k \,\varepsilon\, \{i_1, \cdots, i_r\}, v_k \leqq s_k \text{ otherwise}\}.$$

Thus $B(\mathbf{s}, \mathbf{t}) = B(\mathbf{s}, \mathbf{t}, \{1, \cdots, p\})$. Then,

$$A(\mathbf{t}) - A(\mathbf{s}) = \sum_{\{i_1,\cdots,i_r\}\varepsilon T} B(\mathbf{s}, \mathbf{t}, \{i_1, \cdots, i_r\}). \tag{6.7}$$

It follows that to prove (6.5) we need only establish

$$\lim_{\delta\to 0} \delta^{-p} \limsup_{m,n} \sup_{\mathbf{s}\varepsilon I^p} \tilde{P}[\sup\{|X_N(B(\mathbf{s}, \mathbf{t}, (i_1, \cdots, i_r\}), \cdot)|: \mathbf{t} \,\varepsilon\, A(\mathbf{s}, \delta)\} \geqq \epsilon] = 0 \tag{6.8}$$

for every $\{i_1, \cdots, i_r\}$, $\epsilon > 0$.

We give the argument for $\{i_1, \cdots, i_r\} = \{1, \cdots, p\}$. The same proof holds for the general case.

We return to the notation of Section 4 but suppose throughout that $c_{iN}$ is given by (5.7). We have,

$$\tilde{P}[\sup\{|X_N(B(\mathbf{s}, \mathbf{t}), \cdot)|: \mathbf{t} \,\varepsilon\, A(\mathbf{s}, \delta)\} \geqq \epsilon] \leqq P[\max N^{-\frac{1}{2}}\{|S_{i_1,\cdots,i_p}|: 1 \leqq i_j \leqq N, 1 \leqq j \leqq p\} \geqq \epsilon] \tag{6.9}$$

where $\boldsymbol{\delta} = (\delta, \cdots, \delta)$ and the subset $S$ of Lemma 4.6 is defined as follows.

Move the $NH_N(B(\mathbf{s}, \mathbf{s} + \boldsymbol{\delta}))$ points in $B(\mathbf{s}, \mathbf{s} + \boldsymbol{\delta})$ whose position is given by $H_N$ to points on the lattice of points of the form $(i_1, \cdots, i_p)$ with $1 \leqq i_j \leqq NH_N(B(\mathbf{s}, \mathbf{s} + \boldsymbol{\delta}))$, $1 \leqq j \leqq p$, in such a way as to preserve the strict order relationships between the coordinates of these $NH_N(B(\mathbf{s}, \mathbf{s} + \boldsymbol{\delta}))$ points. It is permissible to move ties to distinct points. Let $S$ be the set of all points $(i_1, \cdots, i_p)$ which are obtained in this fashion from the points $\{P_{iN}\}$. (The inequality in (5.18) is an equality if there are no duplicates among the points $P_{iN}$.)

Now, if $\limsup nN^{-1} \leqq \delta$ it is easy to see that,

$$\liminf_N b(\gamma, N^{\frac{1}{2}}\epsilon) \geqq 1 - \delta(1 - \delta)^{-1}[(1 - \delta(1 - \delta)^{-1} - \gamma)\epsilon]^{-1}. \tag{6.10}$$

If $H$ has a uniform marginal then $H(B(\mathbf{s}, \mathbf{s} + \boldsymbol{\delta})) \leqq \delta$ and by assumption,

$$\sup\{H_N(B(\mathbf{s}, \mathbf{s} + \boldsymbol{\delta})): \mathbf{s} \,\varepsilon\, I^p\} \to \sup\{H(B(\mathbf{s}, \mathbf{s} + \boldsymbol{\delta}): s \,\varepsilon\, I^p\} \leqq \delta. \tag{6.11}$$

Applying Lemma 4.5 to the right hand side of (6.9) and using (6.10), (6.11) we obtain

$$\begin{aligned} &\limsup_N \sup_{\mathbf{s}\varepsilon I^p} \tilde{P}[\sup\{|X_N(B(\mathbf{s}, \mathbf{t}), \cdot)|: \mathbf{t} \,\varepsilon\, A(\mathbf{s}, \delta)\} \geqq \epsilon] \\ &\leqq \gamma^{-r(p-1)} \epsilon^{-r} \limsup_N N^{-\frac{1}{2}} E|S_{n_1}|^r \\ &\cdot \textstyle\prod_{s=0}^{p-2} [1 - \delta(1 - \delta)^{-1} [(1 - \delta(1 - \delta)^{-1} - \gamma)\gamma^s\epsilon]^2]^{-1} \end{aligned} \tag{6.12}$$

for suitable $\gamma$, where $n_1 = NH_N(B(\mathbf{s}, \mathbf{s} + \delta))$. But by Lemma 4.4 if (6.11) holds,

$$\limsup_N N^{-k} E(S_{n_1})^{2k} \leqq \delta^k \mu_{2k}. \tag{6.13}$$

From (6.12) and (6.13) we conclude that in case (a),

$$(6.14)\quad \limsup_N \sup_{\mathbf{s}\varepsilon I^p} P[\sup\{|X_N(B(\mathbf{s}, \mathbf{t}), \cdot|: \mathbf{t}\,\varepsilon\, A(\mathbf{s}, \delta)\} \geqq \epsilon] \leqq \epsilon^{-2k} 4^{k(p-1)}\, \delta^k \mu_{2k}(1 + o(\delta))$$

for $\delta$ so small that we can take $\gamma = \frac{1}{2}$. Taking $k = p + 1$ we see that (6.5) is satisfied. Theorem 3.1 is proved in case (a).

Proof for Case (b). Without loss of generality suppose $q(t) = H(t, \infty, \cdots, \infty)$ is continuous and strictly increasing on [0, 1]. Let $P'_{iN}$ be the point obtained by transforming the first co-ordinate of $P_{iN}$ by $q$ and leaving the others unchanged. If we now define $H_N'$, $F_m'$, $G_n'$, and $X_N'$ suitably it is clear that,

$$(6.15)\qquad ||H_N' - H'|| \to 0$$

where $H'(t, \infty, \cdots, \infty) = t$ for $0 \leqq t \leqq 1$. Therefore, from case (a), the process $X_N'$ satisfies the tightness condition (6.4). But,

$$(6.16)\quad \sup\{|X_N(\mathbf{s}, \cdot) - X_N(\mathbf{t}, \cdot)|: ||\mathbf{s} - \mathbf{t}|| \leqq \delta\} = \sup\{|X_N'(\mathbf{q}^{-1}(\mathbf{s}), \cdot) - X_N'(\mathbf{q}^{-1}(\mathbf{t}), \cdot)|: ||\mathbf{s} - \mathbf{t}|| \leqq \delta\}$$

where

$$\mathbf{q}^{-1}((t_1, \cdots, t_p)) = (q^{-1}(t_1), t_2, \cdots, t_p).$$

In view of (6.16) and the continuity of $q^{-1}$ we see that the theorem holds in case (b).

**7. Proof of Theorem 3.1: Cases (iic) and (iid).** If $q(t)$ is as above we say $s$ is a *point of increase* of $q$ if $t \neq s \Rightarrow q(t) \neq q(s)$. We say $\mathbf{s}$ is a *point of increase* if its first co-ordinate is a point of increase of $q$. An examination of the arguments used for case (b) will show that they serve to establish the theorem even if $q$ is not strictly increasing as long as the $P_{iN}$ are all points of increase. We need only note that the sup on the right hand side of (6.16) can be taken just over the set of all $\mathbf{s}$, $\mathbf{t}$ whose first coordinates are points of increase of $q$.

If $N$ is fixed, let $A^{(1)}$ be the set of all $P_{iN}$ which are points of increase and $A^{(2)}$ be the set of all the other $P_{iN}$. Call the number of members of $A^{(i)}$, $N^{(i)}$, and if $N^{(i)} > 0$ let $H_N^{(i)}$ be the distribution function of the measure assigning mass $[N^{(i)}]^{-1}$ to each point of $A^{(i)}$, while if $N^{(i)} = 0$ let $H_N^{(i)} \equiv 0$. Let $M_N^{(i)}$ be the (random) number of points of the sample which fall in $A^{(i)}$, and if $M_N^{(i)} > 0$ let $F_N^{(i)}$ be the distribution function of the measure which assigns mass $[M_N^{(i)}]^{-1}$ to each point of the sample falling in $A^{(i)}$. Otherwise let $F_N^{(i)} = 0$. Define $G_N^{(i)}$ similarly and let,

$$(7.1)\quad X_N^{(1)}(\mathbf{t}, \cdot) = \{M_N^{(1)}(N^{(1)} - M_N^{(1)})^{-1}\}^{\frac{1}{2}}\,[N^{(1)}]^{\frac{1}{2}}\,[F_N^{(1)}(\mathbf{t}, \cdot) - H_N^{(1)}(\mathbf{t}, \cdot)] \quad \text{if } M_N^{(1)} \text{ and } N^{(1)} - M_N^{(1)} > 0$$
$$\equiv 0 \qquad \text{otherwise.}$$

Then,

$$X_N(\mathbf{t},\cdot) = [M_N^{(1)}(N^{(1)} - M_N^{(1)})/mn]^{\frac{1}{2}} (N^{(1)}N^{-1})^{\frac{1}{2}} X_N^{(1)}(\mathbf{t},\cdot)$$
$$(7.2) \qquad + (m/n)^{\frac{1}{2}} N^{\frac{1}{2}}(M_N^{(1)}m^{-1}) - (N^{(1)}N^{-1}) H_N^{(1)}(\mathbf{t})$$
$$+ (m/n)^{\frac{1}{2}} N^{\frac{1}{2}}(M_N^{(2)}m^{-1}) F_m^{(2)}(\mathbf{t},\cdot)$$
$$- (N^{(2)}N^{-1}) H_N^{(2)}(\mathbf{t},\cdot)).$$

It is easy to see that $(N^{(1)}N^{-1}) \to 1$ as $N \to \infty$. Furthermore since $M_N^{(1)}$ has a hypergeometric distribution we can readily calculate

$$(7.3) \quad E\{(mn^{-1})^{\frac{1}{2}} N^{\frac{1}{2}}(M_N^{(1)}m^{-1}) - (N^{(1)}N^{-1})\}^2$$
$$= (N)(N-1)^{-1} - (N^{(1)}N^{(2)})N^{-2} \to 0.$$

From (7.2) we conclude that to establish (6.4) in case (c) we need only show,

$$(7.4) \quad \lim_\delta \limsup_N \tilde{P}[\max\{|X_N^{(1)}(\mathbf{t},\cdot)$$
$$- X_N^{(1)}(\mathbf{s},\cdot)|: ||\mathbf{s}-\mathbf{t}|| \leqq \delta, \mathbf{s},\mathbf{t} \,\varepsilon\, I^p\} \geqq \epsilon] = 0$$

and

$$(7.5) \quad \lim_N \tilde{P}[\max\{(mn^{-1})^{\frac{1}{2}} N^{\frac{1}{2}}|(M_N^{(2)}m^{-1}) F_m^{(2)}(\mathbf{t},\cdot)$$
$$- N^{(2)}N^{-1} H_N^{(2)}(\mathbf{t},\cdot)| \,\mathbf{t}\,\varepsilon\, I^p\} \geqq \epsilon] = 0 \qquad \text{for every} \qquad \epsilon > 0.$$

Now note that given the sequence of variables $M_N^{(1)}$, the "distribution" of the stochastic processes $X_N^{(1)}(\cdot,\cdot)$ is the same as that of a sequence of processes $X_N(\cdot,\cdot)$ based on taking a sample of size $M_N^{(1)}$ from the $N^{(1)}$ points $\{P_{iN}\}$ lying in $A^{(1)}$. Since the $H_N$ function corresponding to these points (which is just $H_N^{(1)}$) tends to $H$ in norm and since all of the $P_{iN}$ in $A^{(1)}$ are points of increase we may conclude from our extension of case (b) that for every $\epsilon' > 0$ there exists a $\delta' > 0$ and natural numbers $m'$ and $n'$ such that if $m' \leqq r \leqq N^{(1)}$, $N^{(1)} - r \geqq n'$, $\delta \leqq \delta'$ then,

$$(7.6) \quad \tilde{P}[\max\{|X_N^{(1)}(\mathbf{s},\cdot) - X_N^{(1)}(\mathbf{t},\cdot)|: \mathbf{s},\mathbf{t}\,\varepsilon\, I^p,$$
$$||\mathbf{s}-\mathbf{t}|| \leqq \delta\} \geqq \epsilon \mid M_N^{(1)} = r] \leqq \epsilon'.$$

Since both $M_N^{(1)}$ and $N^{(1)} - M_N^{(1)} \to \infty$ in probability as $N \to \infty$ we see that (7.6) implies (7.4). Similarly we see that

$$(7.7) \quad \tilde{P}[\sup\{(mn^{-1})^{\frac{1}{2}} N^{\frac{1}{2}}|(M_N^{(2)}m^{-1}) F_m^{(2)}(\mathbf{t},\cdot)$$
$$- N^{(2)}N^{-1} H_N^{(2)}(\mathbf{t},\cdot) \mid :\mathbf{t}\,\varepsilon I^p\} \geqq \epsilon \mid M_N^{(2)} = r] \leqq P[\sup\{|S^{(2)}_{(i_1,\cdots,i_p)}$$
$$+ n(i_1,\cdots,i_p)\, a(r,N)|: 1 \leqq i_j \leqq N^{(2)}, j = 1,\cdots,p\} \geqq \epsilon],$$

$$(7.8) \qquad a(r,N) = N^{-\frac{1}{2}}[N^{(2)}]^{-1}[r(nm^{-1})^{\frac{1}{2}} - (N^{(2)} - r)(mn^{-1})^{\frac{1}{2}}]$$

and $\{S^{(2)}_{(i_1,\cdots,i_p)}\}$ is the collection of partial sums based on a configuration of $N^{(2)}$ points with selection from a finite population with $N^{(2)}$ members whose values are given by,

$$(7.9)\quad c_{iN}^{(2)} = (nm^{-1})^{\frac{1}{2}} N^{-\frac{1}{2}} - a(r, N) \quad \text{for} \quad i = 1, \cdots, r,$$
$$= (mn^{-1})^{\frac{1}{2}} N^{-\frac{1}{2}} - a(r, N) \quad \text{for} \quad i = r + 1, \cdots, N^{(2)}.$$

Then,

$$(7.10)\qquad \textstyle\sum_{i=1}^{N^{(2)}} c_{iN}^{(2)} = 0$$

and

$$(7.11)\qquad \textstyle\sum_{i=1}^{N^{(2)}} [c_{iN}^{(2)}]^2 \leqq N^{-1}\{(rnm^{-1}) + (N^{(2)} - r)(mn^{-1})\}.$$

Let us now apply Lemma 4.7 to the configuration of $A^{(2)}$ with $k = 3$ and $r = 2$, if $N^{(2)} \geqq 3$.

$$(7.12)\quad P[\sup\{|S_{(i_1,\cdots,i_p)}| : 1 \leqq i_j \leqq N^{(2)}, j = 1, \cdots, p\} \geqq \epsilon]$$
$$\leqq c(3, 2)(1 - \lambda(3) \textstyle\sum_{i=1}^{N^{(2)}} [c_{iN}^{(2)}]^2)^{-(p-1)} E(S_l^{(2)})^2$$

where $S_l^{(2)}$ is the sum of a sample of $l$ from the population $\{c_{iN}^{(2)}\}$. Of course,

$$(7.13)\qquad E(S_l^{(2)})^2 \leqq \tfrac{4}{9} \textstyle\sum_{i=1}^{N^{(2)}} [c_{iN}^{(2)}]^2.$$

Now, if $(rm^{-1}) \to 0$ and $(N^{(2)} - r)(n^{-1}) \to 0$ we see that,

$$(7.14)\qquad \textstyle\sum_{i=1}^{N^{(2)}} [c_{iN}^{(2)}]^2 \to 0.$$

Hence, for $N^{(2)} \geqq 3$, it follows that if

$$(7.15)\qquad N^{(2)} a(M_N^{(2)}, N) \to 0,$$

$$(7.16)\qquad M_N^{(2)} m^{-1} \to 0$$

in probability, (7.5) will follow from (7.7) and (7.11)–(7.14). However, (7.15) and (7.16) follow by a calculation like (7.3). Of course, if $N^{(2)} \leqq 3$, (7.5) follows trivially. The proof of case (c) is complete.

PROOF FOR CASE (d). For every $\{i_1, \cdots, i_p\} \varepsilon T$ there exist at most a denumerable set of hyperplanes of the form $\{\mathbf{t}: t_{i_j} = a_{i_j}\}$ which are assigned positive measure by $H$. Let us for simplicity label all of these hyperplanes by $E_1, E_2, \cdots$. Let $E_0 = R^p$. For every $i \geqq 0$, let $K_i$ be the union of all $E_j$ properly included in $E_i$. Let $N_i$ be the numbers of points $P_{jN}$ which fall in $(E_i - K_i)$ and if $N_i > 0$ let $H_{Ni}$ be the distribution function of the measure which assigns mass $[N_i]^{-1}$ to each point $P_{jN}$ which falls in $(E_i - K_i)$. If $N_i = 0$ let $H_{Ni} \equiv 0$. Similarly, let $M_{Ni}$ be the number of points of the sample falling in $(E_i - K_i)$ and define $F_{mi}$ to be distribution functions of the measure assigning mass $[M_{Ni}]^{-1}$ to each such point if $M_{Ni} > 0$ and to be $\equiv 0$ otherwise. Finally let,

$$(7.17)\quad X_{Ni}(\mathbf{t}, \cdot) = \{M_{Ni}(N_i - M_{Ni})^{-1}\}^{\frac{1}{2}} N_i^{\frac{1}{2}}(F_{mi}(\mathbf{t}, \cdot) - H_{Ni}(\mathbf{t}, \cdot)).$$

We begin by proving case (d) when there are only a finite number say $J$ of the $E_i$ and $H(E_i - K_i) > 0$ for $i = 1, \cdots, J$.

Write

$$(7.18)\quad X_N(\mathbf{t},\ \cdot) = \textstyle\sum_{i=1}^{J} \{[M_{Ni}(N_i - M_{Ni})(mn)^{-1}]^{\frac{1}{2}}\ (N_iN^{-1})^{-\frac{1}{2}}\ X_{Ni}(\mathbf{t},\ \cdot) + (mn^{-1})^{\frac{1}{2}}\ N^{\frac{1}{2}}(M_{Ni}m^{-1} - N_iN^{-1})\ H_{Ni}(\mathbf{t})\}.$$

Note first that,

$$(7.19)\qquad H_{Ni}(\mathbf{t}) \to H_i(\mathbf{t})$$

uniformly in $\mathbf{t}\ \varepsilon\ I^p$ where $H_i$ is the distribution function of the measure assigned by $H$ to $E_i - K_i$ divided by $H(E_i - K_i)$. Further, by a calculation such as (7.3), since $N_iN^{-1} \to H(E_i - K_i)$ we see that,

$$(7.20)\qquad M_{Ni}m^{-1} \to H(E_i - K_i)$$

in probability and in fact,

$$(7.21)\qquad (mn^{-1})^{\frac{1}{2}}\ N^{\frac{1}{2}}(M_{Ni}m^{-1} - N_iN^{-1}) = O_p(1),$$

(is bounded in probability).

Since the map $(f_1, \cdots, f_J) \to \alpha_1 f_1 + \cdots + \alpha_J f_J$ of $\mathfrak{M}(I^p) \times \cdots \times \mathfrak{M}(I^p) \to M(I^p)$ is continuous in $\|\cdot\|$ we see that to prove (5.2) we need only show that, for every $\epsilon > 0$, $\delta > 0$, $1 \leqq i \leqq J$, there exists a compact $K \subset Q_H$ such that,

$$(7.22)\qquad \tilde{P}[\tilde{\omega}: X_{Ni}(\cdot, \tilde{\omega})\ \varepsilon\ K^{\delta}] \geqq 1 - \epsilon,$$

and

$$(7.23)\quad \tilde{P}[\tilde{\omega}: (mn^{-1})^{\frac{1}{2}}\ N^{\frac{1}{2}}(M_{Ni}(\tilde{\omega})m^{-1} - N_iN^{-1})\ H_{Ni}(\cdot)\ \varepsilon\ K^{\delta}] \geqq 1 - \epsilon$$

for all $m$, $n$ sufficiently large. Now, (7.19) and (7.21) suffice to establish (7.23).

For simplicity we shall prove (7.22) when $E_i = \{\mathbf{t}: t_p = 0\}$. The proof of the general case is analogous. Define

$$(7.24)\qquad \tilde{X}_{Ni}(t_1, \cdots, t_{p-1}, \tilde{\omega}) = X_{Ni}(t_1, \cdots, t_{p-1}, 0, \tilde{\omega}).$$

It is easy to see that given $M_{Ni} = m_i$, the stochastic process $\tilde{X}_{Ni}(t_1, \cdots, t_{p-1}, \cdot)$ is distributed as an $X_N(\mathbf{t}, \cdot)$ process based on a configuration of $N_i$ points in $R^{p-1}$ with distribution function $\tilde{H}_{Ni}$ given by

$$(7.25)\qquad \tilde{H}_{Ni}(t_1, \cdots, t_{p-1}) = H_{Ni}(t_1, \cdots, t_{p-1}, 0).$$

Of course,

$$(7.26)\quad \tilde{H}_{Ni}(t_1, \cdots, t_{p-1}) \to \tilde{H}_i(t_1, \cdots, t_{p-1}) = H_i(t_1, \cdots, t_{p-1}, 0)$$

and by construction $\tilde{H}_i$ is continuous on $I^{p-1}$. Since $M_{Ni} \to \infty$ in probability as does $N_i - M_{Ni}$, and since case (c) has been established for all $p$ we conclude that,

$$(7.27)\quad \lim_{\delta} \limsup_{m,n} \tilde{P}[\sup\{|\tilde{X}_{Ni}(\mathbf{s}, \cdot) - \tilde{X}_{Ni}(\mathbf{t}, \cdot)|: \|\mathbf{s} - \mathbf{t}\| \leqq \delta,\ \mathbf{s}, \mathbf{t}\ \varepsilon\ I^{p-1}\} \geqq \epsilon] = 0$$

and hence by Dudley's Proposition 2 [6], for every $\epsilon > 0$, $\epsilon > 0$ there exists a compact subset $K_i$ of $Q\tilde{H}_i$ such that

$$(7.28) \qquad \tilde{P}[\tilde{\omega}: \tilde{X}_{Ni}(\cdot, \tilde{\omega}) \,\varepsilon\, K_i^{\delta}] \geqq \epsilon$$

for all $m$, $n$ sufficiently large. (Note that $K_i^{\delta} \subset \mathfrak{M}(I^{p-1})$.) Now, the map $T$ of $\mathfrak{M}(I^p)$ into $\mathfrak{M}(I^{p-1})$ which is defined by,

$$(7.29) \qquad Tf(t_1, \cdots, t_{p-1}) = f(t_1, \cdots, t_{p-1}, 0)$$

is continuous in $||\cdot||$ and hence (7.28) implies (7.22).

To complete the proof of case (d) we must finally deal with the possibility that some of the $H(E_i - K_i) = 0$ and that $J = \infty$. Let $i_1, i_2, \cdots$ be the sequence of indices such that $H(E_{i_j} - K_{i_j}) > 0$.
Write,

$$(7.30) \quad X_N(\mathbf{t}, \cdot) = \textstyle\sum_{j=1}^{A} X_{Ni_j}(\mathbf{t}, \cdot)\,[(M_{Ni_j}(N_{i_j} - M_{Ni_j}) \\ (mn)^{-1}]^{\frac{1}{2}}\,(N_{i_j}N^{-1})^{-\frac{1}{2}}] + R_{NA}(\mathbf{t}, \cdot).$$

From our previous assertions it is evident that (5.2) will follow in this general case if we can show that for every $\epsilon > 0$, $\delta > 0$, there exists an $A < \infty$ such that

$$(7.31) \qquad \tilde{P}[\sup\{|R_{NA}(\mathbf{t}, \cdot)| : \mathbf{t} \,\varepsilon\, I^p\} \leqq \delta] \geqq 1 - \epsilon$$

for all $m$, $n$ sufficiently large.

Let,

$$(7.32) \qquad g(A) = \textstyle\sum_{i=A+1}^{\infty} H(E_{i_j} - K_{i_j})$$

and $Ng_N(A)$ equal the number of $P_{iN}$ which are not in $\bigcup_{j=1}^{A} (E_{i_j} - K_{i_j})$. Evidently,

$$(7.33) \qquad g_N(A) \to g(A) \quad \text{as } N \to \infty \quad \text{and,}$$

$$(7.34) \qquad g(A) \to 0 \quad \text{as } A \to \infty.$$

We argue as in the proof of (7.5). Let $M_N^A$ be the number of points in the sample which are not in $\bigcup_{j=1}^{A} (E_{i_j} - K_{i_j})$. Then,

$$(7.35) \quad \tilde{P}[\sup\{|R_{NA}(\mathbf{t}, \cdot)| : \mathbf{t} \,\varepsilon\, I^p\ \delta \mid M_N^A = r] \\ \leqq P[\max\{|S_{(i_1,\cdots,i_p)} + n(i_1, \cdots, i_p)a(r, N, A)| : 1 \leqq i_j \leqq Ng_N(A), \\ j = 1, \cdots, p\} \geqq \delta]$$

where $S^A_{(i_1,\cdots,i_p)}$ are partial sums corresponding to a configuration of $Ng_N(A)$ points and sampling from a population of $Ng_N(A)$ numbers $\{c_{iN}(r)\}$ which are given by,

$$(7.36) \quad c^A_{iN}(r) = (nm^{-1})^{\frac{1}{2}}N^{\frac{1}{2}} - a(r, N, A) \qquad \text{for } i = 1, \cdots, r \\ = -(mn^{-1})^{\frac{1}{2}}N^{\frac{1}{2}} - a(r, N, A) \qquad \text{for } i = r + 1, \cdots, Ng_N(A)$$

and

$$(7.37)\quad a(r, N, A) = N^{-\frac{1}{2}}(Ng_N(A))^{-1}\{r(nm^{-1})^{\frac{1}{2}} - [Ng_N(A) - r](mn^{-1})^{\frac{1}{2}}\}.$$

If $Ng_N(A) \geqq 3$ we can apply Lemma 4.7 with $k = 3$ and $r = 2$ to the right hand side of (7.35) and get after some computation,

$$\tilde{P}[\sup\{|R_N(\mathbf{t}, \cdot)| : \mathbf{t} \varepsilon I^p\} \geqq \delta \mid M_N^A]$$

$$(7.38)\quad \leqq \tfrac{4}{9}c(2, 3)[1 - \lambda(3)(\delta + Ng_N(A)|a(M_N^A, N, A)|)^{-2}K(M_N^A, A)]^{-(p-1)}$$

$$\cdot(\delta + Ng_N(A)|a(M_N^A, N, A)|)^{-2}K(M_N^A, A),$$

where

$$(7.40)\quad K(r\, A), = \textstyle\sum_{i=1}^{Ng_N(A)} [c_{iN}(r)]^2.$$

Finally, we have,

$$(7.41)\quad K(M_N^A, A) \leqq \max\{M_N^A m^{-1}, (Ng_N(A) - M_N^A)n^{-1}\}$$

and

$$(7.42)\quad E([Ng_N(A)a(M_N^A, N, A)]^2) = E((mn^{-1})^{\frac{1}{2}}N^{\frac{1}{2}}(M_N^A m^{-1} - g_N(A))^2)$$

$$= g_N(A)(1 - g_N(A))N(N - 1)^{-1} \leqq 2\, g_N(A)$$

for $N$ sufficiently large. From (7.42) we see that,

$$(7.43)\quad \tilde{P}[|[Ng_N(A)]^{-1}a(M_N^A, N, A)| \geqq g_N^{\frac{1}{2}}(A)] \leqq 2g_N^{\frac{1}{2}}(A),$$

$$(7.44)\quad \tilde{P}[M_N^A m^{-1} \geqq g_N(A) + m^{-\frac{1}{2}}] \leqq 2g_N(A)$$

and

$$(7.45)\quad \tilde{P}[(Ng_N(A) - M_N^A)m^{-1} \geqq g_N(A) + n^{-\frac{1}{2}}] \leqq 2g_N(A).$$

By (7.39), (7.41) and (7.43)–(7.45) we obtain,

$$\tilde{P}[\sup\{|R_{NA}(\mathbf{t}, \cdot)| : \mathbf{t} \varepsilon I^p\} \geqq \delta]$$

$$(7.46)\quad \leqq \tfrac{4}{9}c(2, 3)[1 - \lambda(3)(\delta + g_N^{\frac{1}{2}}(A))^{-2}\, g_N(A)$$

$$+ \max(m^{-\frac{1}{2}}, n^{-\frac{1}{2}})]^{-(p-1)}[(g_N(A)$$

$$+ \max(m^{-\frac{1}{2}}, n^{-\frac{1}{2}})](\delta + g_N^{\frac{1}{2}}(A))^{-2} + 6g_N^{\frac{1}{2}}(A),$$

for $Ng_N(A) \geqq 3$. Letting $N \to \infty$ and then $A \to \infty$ we have by (7.33) and (7.34),

$$(7.47)\quad \lim_{A\to\infty} \lim\sup_N \tilde{P}[\sup\{|R_{NA}(\mathbf{t}, \cdot)| : \mathbf{t} \varepsilon I^p\} \geqq \delta] = 0$$

and this is equivalent to requiring (7.31) for $N$ sufficiently large. The same conclusion evidently holds if $Ng_N(A) < 3$. The theorem is proved.

**Acknowledgment.** I am grateful to Volker Strassen for a helpful conversation.

REFERENCES

[1] BARTON, C. L. and MALLOWS, D. E. (1965). Some aspects of the random sequence. *Ann. Math. Statist.* **36** 236–260.

[2] DAVID, F. N. and FIX, E. (1960). Rank correlation and regression in a nonnormal surface. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **6** 177–197. Univ. of California Press.

[3] DARLING, D. A. (1957). The Kolmogorov-Smirnov, Crámer-von Mises tests. *Ann. Math. Statist.* **28** 823–838.

[4] DOOB, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

[5] DUDLEY, R. M. (1965). Gaussian processes on several parameters. *Ann. Math. Statist.* **36** 771–788.

[6] DUDLEY, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

[7] FRASER, D. A. S. (1957). *Nonparametric Methods in Statistics*. Wiley, New York.

[8] HÁJEK, J. (1961). Some extensions of the Wald-Wolfowitz-Noether theorems. *Ann. Math. Statist.* **32** 506–523.

[9] HÁJEK, J. (1965). Extension of the Kolmogorov-Smirnov test to regression alternatives. *Bernoulli-Bayes-Laplace Seminar*. Springer Verlag, Heidelberg.

[10] KIEFER, J. (1961). On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.* **11** 649–660.

[11] KIEFER, J. and WOLFOWITZ, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* **87** 173–186.

[12] LEHMANN, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *Ann. Math. Statist.* **22** 165–179.

[13] PROHOROV, Y. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Prob. Appl.* **1** 157–214.

[14] ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472

[15] SCHMID, P. (1958). On the Kolmogorov Smirnov limit theorem for discontinuous distribution functions. *Ann. Math. Statist.* **29** 1011–1027.

[16] SEN, P. K. and CHATTERJEE, S. K. (1964). Nonparametric tests for the bivariate two sample location problem. *Calcutta Statist. Assn. Bull.* **13** 18–58.

[17] SIMPSON, P. B. (1951). Note on estimation of a bivariate distribution function. *Ann. Math. Statist.* **22** 476–478.

[18] VINIZE, I. (1960). Two sample tests based on order statistics. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 697–705. Univ. of California Press.

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

[20] WEISS, L. (1960). Two sample tests for multivariate distributions. *Ann. Math. Statist.* **31** 159–164.