BOOTSTRAP AND RANDOMIZATION TESTS
OF SOME NONPARAMETRIC HYPOTHESES

BY

JOSEPH P. ROMANO

TECHNICAL REPORT NO. 270
APRIL 1987

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
BOOTSTRAP AND RANDOMIZATION TESTS
OF SOME NONPARAMETRIC HYPOTHESES

BY

JOSEPH P. ROMANO

TECHNICAL REPORT NO. 270
APRIL 1987

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
BOOTSTRAP AND RANDOMIZATION TESTS
OF SOME NONPARAMETRIC HYPOTHESES

by

Joseph P. Romano

April, 1987
Department of Statistics
Stanford University

In this paper, the asymptotic behavior of some nonparametric tests is studied in situations where both bootstrap tests and randomization tests are applicable. Under fairly general conditions, the tests are asymptotically equivalent in the sense that the resulting critical values and power functions are appropriately close. This implies, among other things, that the difference in the critical functions of the tests, evaluated at the observed data, tends to zero in probability. Randomization tests may be preferable since an exact desired level of the test may be obtained for finite samples. Examples considered are: testing independence, testing for spherical symmetry, testing for exchangeability, testing for homogeneity, and testing for a change point.

Key Words: bootstrap, nonparametric tests, randomization tests, Vapnik-Cervonenkis classes, testing independence, testing equality of distributions, testing for rotational invariance.
SECTION 1
INTRODUCTION

The main goal of this paper is to study the behavior of some nonparametric tests having a common structure. In particular, two methods to simulate a null distribution will be analyzed and compared. The bootstrap method, formulated by Efron (1979), has been shown to be a widely applicable method in testing problems; see Beran (1986) and Romano (1986). In this paper, the problem of testing a nonparametric hypothesis is considered in those situations where certain invariance or randomization ideas apply, thus yielding a randomization distribution as an alternative to the bootstrap distribution. The idea of randomization dates back to Fisher (1935), and then Pitman (1937/8). Both methods are the same in that rejection of a null hypothesis occurs when a common test statistic is large. However, the approaches differ in that critical values are determined by (usually) distinct resampling methods to estimate a null distribution.

The statistical problem considered here has the following fairly general form. Given a sample \( X_1, \ldots, X_n \) of \( S \)-valued random variables, we wish to test the null hypothesis \( H_0 \) that the unknown probability distribution \( P \) on \( S \) generating the data belongs to a certain class \( \Omega_0 \) against the alternative class \( \Omega_1 \). Here, if \( \Omega \) represents the class of all probabilities on \( S \) (equipped with a certain \( \sigma \)-field \( S \)), then \( \Omega_1 \) will typically be \( \Omega - \Omega_0 \). Moreover, \( \Omega_0 \) can be characterized as the set of probabilities \( P \) satisfying \( \tau P = P \) for some mapping \( \tau \) from \( \Omega \) to \( \Omega_0 \). Furthermore, if \( \delta \) is a metric (or possibly a pseudometric) on the space of probabilities on \( S \), \( \Omega_0 \) is specified by \( \delta(P, \tau P) = 0 \). Intuitively, the larger the value of \( \delta(P, \tau P) \), the greater the departure of \( P \) from \( \Omega_0 \). Apparently, any testing problem may be placed in this framework.

Let \( \hat{P}_n \) be the empirical measure of \( X_1, \ldots, X_n \). Then, the proposed test rejects for large values of \( T_n = T_n(X_1, \ldots, X_n) \), where \( T_n \) is of the form

\[
T_n = \frac{1}{n^2} \delta \left[ \hat{P}_n, \tau \hat{P}_n \right],
\]

(1.1)
so that the test rejects when $\tau \hat{P}_n$ is sufficiently far from $\hat{P}_n$.

A typical choice for $\delta$, in the spirit of Kolmogorov Smirnov test statistics, is

$$\delta_V(P, Q) = \sup \left\{ V \in \mathcal{V} : \left| P(V) - Q(V) \right| \right\} \tag{1.2}$$

for some collection of events $V$. It has the advantages of being applicable quite generally (especially for complex data types) and yields tests with excellent power properties. Henceforth, we restrict attention to test statistics $T_n$ given by (1.1) and with $\delta$ given by (1.2).

Next, we give a typical example of the testing problem just described. Further examples will be given in section 2.

Example 1, Testing Independence. Let $X_1, \ldots, X_n$ be i.i.d. $S$-valued random variables and suppose $X_i = (X_{i,1}, \ldots, X_{i,d})$ is made up of $d$ components. The problem is to test the joint independence of the components. To get started, suppose the $j$th component takes values in a space $S_j$, and $S$ is the product space $S = \prod_{j=1}^d S_j$. If $P$ is a probability on $S$, let $P_j$ be the marginal probability on $S_j$ of the $j$th component. In terms of random variables, if $X_i = (X_{i,1}, \ldots, X_{i,d})$ has law $P$ and $X_{i,j}$ takes values in $S_j$, then for $A$ in $S_j$, $P_j(A)$ is the probability that $X_j$ falls in $A$. If $P$ is a probability on $S$ with marginals $P_j$, let $\tau P$ be the product probability $\prod_{j=1}^d P_j$. Note that $\tau P = P$ if and only if $P$ is a product of its marginals. Let $\hat{P}_n$ be the empirical measure of $\left\{ X_i : 1 \leq i \leq n \right\}$ and let $\hat{P}_{n,j}$ be the marginal probability of $\hat{P}_n$ on $S_j$ of the $j$th component. If $A_j$ is a subset of $S_j$, let $\prod_{j=1}^d A_j$ denote the product rectangle of the $A_j$ in $S$. Then, the proposed test rejects for large values of the test statistic $T_n = T_n(X_1, \ldots, X_n)$, where $T_n$ is of the form

$$T_n = n^{1/2} \sup_{A_j \in V_j} \left| \hat{P}_n \left( \prod_{j=1}^d A_j \right) - \prod_{j=1}^d \hat{P}_{n,j}(A_j) \right|$$

and the sup ranges over classes of sets $V_j$ in $S_j$. 


Notice the generality of the testing problem and the flexibility of the choice of test statistic. Specifically, no continuity assumption on the underlying distribution is made. Also, the component spaces can, in fact, be quite general. Indeed, they can be different; some variables could be quantitative (continuous or discrete), while others might be qualitative or categorical. Moreover, $S_j$ itself could be a subset of $\mathbb{R}^k$ for $k>1$. In this case, $X_{i,j}$ itself is a vector of $k$ components, but we do not wish to test the independence of these subcomponents; rather, it is desired to test whether these subcomponents taken jointly are independent from, say, another component. Finally, the results allow for a choice in the collection of sets $V_j$ defining the test statistic. For example, three-dimensional data $(W_i, Y_i, Z_i)$ in $\mathbb{R}^3$ might be given and it is desired to test whether $(W_i, Y_i)$ and $Z_i$ are independent. Then, $S_1$ is the plane and several reasonable choices for $V_1$ exist: all lower left-hand quadrants, all half-spaces, or all ellipses, for example. In summary, no assumptions will be made on the underlying probability law $P$ of the data. However, as will be clearly stated in section 2, the class of sets $V_j$ chosen must be a Vapnik-Cervonenkis class.

1.1 Bootstrap Test.

Let $J_n(P)$ be the law of $T_n(X_1, \ldots, X_n)$ when $X_1, \ldots, X_n$ are i.i.d. $P$. In order to obtain a critical value for a test based on $T_n$, $J_n(P)$ must be approximated for $P \in \Omega_\delta$; that is, $J_n(\tau P)$ must be approximated. The bootstrap procedure is to estimate $J_n(\tau P)$ by $J_n(\tau \hat{P}_n)$, and then use the corresponding critical value from this estimated sampling distribution.

We formally define a bootstrap critical value as follows. Let $J_n(t, P) = P\left[T_n(X_1, \ldots, X_n) \geq t\right]$; that is, $J_n(\cdot, P)$ is the survival function of $T_n(X_1, \ldots, X_n)$ when $X_1, \ldots, X_n$ are i.i.d. $P$. For $\alpha \in (0, 1)$, let

$$b_{n,L}(\alpha, P) = \inf \left\{ t : J_n(t, P) \leq \alpha \right\}$$

and

$$b_{n,U}(\alpha, P) = \sup \left\{ t : J_n(t, P) \geq \alpha \right\}.$$
If $\hat{\alpha}_n$ is any estimate of $\alpha$, let $b_n(\alpha, \hat{\alpha}_n)$ be any random variable lying between $b_{n,L}(\alpha, \hat{\alpha}_n)$ and $b_{n,U}(\alpha, \hat{\alpha}_n)$. In particular, let $\hat{\alpha}_n = \tau \hat{P}_n$, where $\hat{P}_n$ is the empirical measure of $X_1, \ldots, X_n$. Then, the level $\alpha$ bootstrap test rejects when $T_n > b_n(\alpha, \hat{\alpha}_n)$. The random variable $b_n(\alpha, \hat{\alpha}_n)$ is called a bootstrap critical value.

In Romano (1986), such a bootstrap procedure is applied to several examples (testing goodness of fit, testing independence, testing for spherical symmetry, etc.), and it is established that

$$P_0 \left[ T_n > b_n(\alpha, \hat{\alpha}_n) \right] \to \alpha \quad \text{as} \quad n \to \infty \quad (1.3)$$

for any $P_0$ in $\Omega_0$, and such tests are consistent against all alternatives.

In this paper, this general testing problem is specialized to certain cases where invariance principles apply, thus leading to a randomization procedure as a competitor to the bootstrap procedure.

1.2 Randomization Test.

The following is assumed. As the notation suggests (borrowed from Hoeffding, 1952), the objects considered are defined for an infinite sequence of positive integers $n$ in anticipation of some asymptotic results. The observation $\chi_n$ takes values in a sample space $S^{(n)}$. Typically, $\chi_n$ is a vector of $n$ iid $S$-values random variables. Let $G_n$ be a group of transformations of $S^{(n)}$ onto itself. For now, assume $G_n$ is finite with $M_n$ elements. We assume the hypothesis implies that the distribution $P^{(n)}$ of $\chi_n$ is invariant under $G_n$; that is, for every $g$ in $G_n$, $g\chi_n$ and $\chi_n$ have the same distribution. Let $T_n$ be any real-valued test statistic defined on $S^{(n)}$. For every $x$ in $S^{(n)}$, let

$$T_n^{(1)}(x) \leq T_n^{(2)}(x) \leq \cdots \leq T_n^{(M_n^*)}(x)$$

be the ordered values of $T_n(gx)$ for all $g$ in $G_n$. Given a number $\alpha$ in $(0, 1)$, let $k_n = k_n(\alpha)$ be defined by $k_n = M_n - \left[ M_n \alpha \right]$, where $\left[ t \right]$ denotes the largest integer less than or equal to $t$. Let $M_n^+(x)$ and $M_n^0(x)$ be the numbers of values $T_n^{(j)}(x)$, ($j = 1, \ldots, M_n$) which are greater than
Define 

\[ a_n(x) = \frac{M_n \alpha - M_n^+(x)}{M_n^0(x)} \]  

(1.4)

Let \( \phi_n(x) \) be the test function equal to 1 if \( T_n(x) > T_n^{(k_x)}(x) \), 0 if \( T_n(x) < T_n^{(k_x)}(x) \), and equal to \( a_n(x) \) if \( T_n(x) = T_n^{(k_x)}(x) \). Define \( r_n(\alpha, \bar{x}_n) = T_n^{(k_x)}(\bar{x}_n) \) to be a randomization critical value.

Then, for any \( P^{(n)} \) which is invariant under \( G_n \), \( E_{P^{(n)}}[\phi_n(X_n)] = \alpha \). Such a test will be referred to as a randomization test.

It is a well-known argument why the test \( \phi_n \) has exact level \( \alpha \). In particular, let \( G_n x \) be the \( G_n \)-orbit of \( x \) in \( S \); that is, \( G_n x \) is the set \( G_n x = \{ g x \mid g \in G_n \} \). Note that \( G_n \) defines an equivalence relation on \( S^{(n)} \) by the rule that \( x \) and \( y \) are \( G_n \)-equivalent if \( y = g x \) for some \( g \) in \( G_n \). Now, conditional on \( \bar{x}_n \in G_n x \), the test statistic is equally likely to be any of the values \( T_n^{(j)}(x) \), \( 1 \leq j \leq M_n \). Hence, a conditional level \( \alpha \) test has been constructed for each \( x \), yielding a test with unconditional level \( \alpha \) as well. It is now apparent that \( a_n(x) \) has been so defined as to achieve exact level \( \alpha \), conditional on \( G_n x \), admitting for the possibility of ties or the possibility that \( \alpha M_n \) is not an integer.

The randomization distribution of \( T_n \) will be denoted by \( J_n(P^{(n)} \mid G_n \bar{x}_n) \). That is, \( J_n(P^{(n)} \mid G_n \bar{x}_n) \) is the conditional distribution of \( T_n(\bar{x}_n) \) under \( P^{(n)} \) given that \( \bar{x}_n \) falls in \( G_n \bar{x}_n \). Thus, if \( P^{(n)} \) is invariant under \( G_n \), \( J_n(P^{(n)} \mid G_n \bar{x}_n) \) is the random distribution assigning equal mass to each of the \( M_n \) values \( T_n(\bar{g}_j \bar{x}_n) \). As a consequence, \( J_n(P^{(n)} \mid G_n \bar{x}_n) \) does not depend on \( P^{(n)} \) if \( P^{(n)} \) is invariant under \( G_n \).

The connection with the bootstrap set-up should be apparent. When \( \bar{x}_n \) is a vector of \( n \) i.i.d. variables with distribution \( P \), then \( P^{(n)} = P^n \) and \( J_n(P^n \mid G \bar{x}_n) \) is actually independent of \( P \) for \( P \) in \( \Omega_0 \). It may be worthwhile to point out that, in this case, when \( \Omega_0 \) consists of those \( P \) with \( P^n \) invariant under \( G_n \), a potential choice for a test statistic is \( T_n \) given by (1.1) with \( \tau \hat{\rho}_n = E[\hat{\rho}_n \mid G_n \bar{x}_n] \). Similar to \( J_n(t, P) \), define \( J_n(t, P^{(n)} \mid G_n \bar{x}_n) \) to be the conditional
probability that \( T_n(\bar{x}_n) \) is greater than or equal to \( t \) given that \( \bar{x}_n \) falls in \( G_{n,\bar{x}_n} \) and \( \bar{x}_n \) has distribution \( P^{(n)} \).

Both bootstrap and randomization tests reject for large values of \( T_n \). The difference is that critical values are determined by referring to distinct distributions \( J_n(\tau_{\hat{P}_n}) \) and \( J_n(P_0^0 \mid G_{n,\bar{x}_n}) \), where \( P_0 \) is any distribution in \( \Omega_0 \). Actually, to be certain this conditional distribution makes sense, assume \( P_0 \) is chosen so that the set \( G_{n,\bar{x}_n} \) lies in the support of \( P_0^0 \). In any case, it is always understood that \( J_n(P_0^0 \mid G_{n,\bar{x}_n}) \) refers to the distribution of \( T_n(\bar{y}_n) \) when \( \bar{y}_n \) is uniformly distributed over \( G_{n,\bar{x}_n} \). To remove the arbitrariness of a choice of \( P_0 \), an alternative notation to \( J_n(P_0^0 \mid G_{n,\bar{x}_n}) \) might be \( J_n((\tau\hat{P}_n)^n \mid G_{n,\bar{x}_n}) \).

Example 1, Testing Independence, continued. The sample space \( S^{(n)} \) is \( S^n \), where \( S \) is as explained in example 1. An element \( \bar{x}_n = (x_1, \ldots, x_n) \) in \( S^n \) is therefore made up of \( d \) components, so that \( x_i = (x_{i,1}, \ldots, x_{i,d}) \) with \( x_{i,j} \) in \( S_j \). Let \( \pi_j = \pi_{j}^n, 1 \leq j \leq n !, \) be the \( n ! \) permutations of \( \{1, \ldots, n\} \). Given integers \( i_1, \ldots, i_d \), each between 1 and \( n ! \), let \( g_{i_1}, \ldots, i_d \in G_n \) be defined to transform \( \bar{x}_n \) into \( \bar{y}_n \), where \( \bar{y}_n \) has \( k \)th component \( y_k \) (in \( S \)) given by \( y_{k,j} = x_{\pi_{j}^n,k,j} \). Then, for any \( g \) in \( G_n \) and any \( P \) in \( \Omega_0 \), \( gP^n = P^n \); that is, the distribution of \( g\bar{x}_n \) is the same as \( \bar{x}_n \) if \( \bar{x}_n \) has distribution \( P^n \). In words, under the hypothesis of independence, we can, for each \( j \), permute the data values in \( S_j \) with each other to form a new data set which has the same distribution as the original data set.

To illustrate the difference in the techniques in the example of testing independence, the following algorithms may be employed to carry out the procedures.

1. Computing the Bootstrap distribution.

   Step 1. Given the sample \( \bar{x}_n \) (that is, given an observation with empirical \( \hat{P}_n \) and marginal empiricals \( \hat{P}_{n,j} \)), sample independent observations \( X_{i,j}^* = (X_{i,1,j}^*, \ldots, X_{i,d,j}^*) \) for \( i=1, \ldots, n \) so that, for each \( j \), \( X_{i,j}^* \) are i.i.d. \( \hat{P}_{n,j} \). That is, for each \( j \), the \( X_{i,j}^* \) bootstrap variables are sampled
from the original data values \( \{ X_{i,j}, 1 \leq i \leq n \} \) with replacement.

**Step 2.** The \( X^*_i \)'s make up a new data set in \( S^n \), say \( \mathbf{x}^*_n = (X^*_1, \ldots, X^*_n) \). The data set \( \mathbf{x}^*_n \) is called a bootstrap sample. Repeat step 1 \( B \) times to generate data sets \( \mathbf{x}^*_n, b \) for \( b = 1, \ldots, B \). For each new data set, compute \( T_n(\mathbf{x}^*_n, b) \).

**Step 3.** The distribution assigning equal mass to these \( B \) values serves as a stochastic approximation to \( J_n(\hat{p}_n) \). As \( B \to \infty \), the error in approximation tends to zero in an appropriate sense.

**II. Computing the Randomization distribution.**

**Step 1.** A new data set is constructed as in step 1 for the bootstrap procedure, but this time, for each \( j \), \( X^*_i, j \) is sampled from the values \( \{ X_{i,j}, 1 \leq i \leq n \} \) without replacement. This yields a data set \( \mathbf{x}^*_n \) analogous to the bootstrap samples, except that the marginal empiricals for this data set (unlike the bootstrap) necessarily are the same as for the original data set.

**Step 2.** Repeat step 1 \( B \) times to generate independent data sets \( \mathbf{x}^*_n, b \) for \( b = 1, \ldots, B \). For each data set, compute \( T_n(\mathbf{x}^*_n, b) \).

**Step 3.** The distribution assigning equal mass to these \( B \) values serves as a stochastic approximation to \( J_n(P_0^* | G_n, x_n) \).

In order to compute the exact randomization distribution, step 2 may be modified. Notice that each new data set \( \mathbf{x}^*_n, b \) may be expressed as \( \mathbf{x}^*_n, b = g \mathbf{x}_n \) for some \( g \) in \( G_n \). So, one may choose \( g \)'s with or without replacement from \( G_n \). The case of choosing \( B = M_n \) \( g \)'s without replacement from \( G_n \) corresponds to exact evaluation of the randomization distribution. Unfortunately, in the case of testing independence, the number of \( g \)'s one needs for an exact evaluation is \((n!)^{d-1}\), so this approach may not be practical. (The reason the exponent is \( d-1 \) and not \( d \) is that we may keep one coordinate fixed and permute the remaining ones without affecting the randomization distribution.) The results obtained in this paper will apply even when bootstrap and randomization distributions must be approximated as just described; see
section 2.3.

The following point may help to understand the conceptual distinction between the bootstrap and randomization procedures. The bootstrap distribution may be viewed as an unconditional approximation to the null distribution of the test statistic while the randomization distribution may be viewed as a conditional distribution of the test statistic. In the notation previously defined, \(J_n \left[ \tau^B_n \right] = J_n (P^n_0 | S)\), where \(P_0\) is any member of \(\Omega_0\). If it were the case that \(J_n (P^n_0 | G_n x_n)\) did depend on the actual \(P\) in \(\Omega_0\), an alternative or combined approach might be to approximate the conditional distribution \(J_n (P^n | G_n x_n)\) by a bootstrap procedure, say \(J_n \left[ \tau^B_n | G x_n \right]\). In this way, the randomization distribution may be considered a conditional bootstrap distribution. In the case described here, the conditioning is done in such a way so that \(J_n \left[ \tau^B_n | G x_n \right]\) is precisely \(J_n (P^n_0 | G x_n)\) with no error.

Before outlining the balance of the paper and summarizing the mathematical results, it seems important to motivate the type of results sought. After all, a (perhaps) reasonable test statistic is available and the randomization null distribution yields a valid \(\alpha\) level test. However, the fact that the randomization procedure yields an exact level \(\alpha\) test does not warrant its use. For example, it may be the case that the ordered values of \(T_n (g x_n)\) may contain many duplicated, so that the test might be randomized (i.e., based on a coin flip) with high probability; that is, \(T_n (x_n) = T_n^{(k_x)} (x_n)\) with high probability. (as defined in (1.4)) with high probability. This can actually happen if one chooses the class of sets \(V\) in (1.2) in a bad way, as shown by the following trivial example.

Suppose \(X_1, \ldots, X_n\) are i.i.d. \(P\) and take values in the unit circle \(S\). We wish to test the null hypothesis that \(P\) is the uniform distribution \(P_0\). The test statistic is (1.1) with \(\tau^P = P_0\) for all \(P\) and \(\delta\) is given by (1.2) for some collection of sets \(V\) on \(S\). The probability distribution \(P^n_0\) of \(x_n = (X_1, \ldots, X_n)\) is invariant under rotations. That is, each \(X_i\) may be rotated a fixed amount \(\theta_i\) to form a new data set \(x_n\) with distribution \(P^n_0\). This is a slight generalization of the previous set-up because \(G_n\) is not finite, but this presents no problem. It
should be clear that the bootstrap distribution and the randomization distribution are the same; in fact, both are the same as the distribution of the test statistic $J_n(P_0)$ (and are not random) because $\Omega_0$ consists of just $P_0$. Now, if $V$ consists of all (measurable) subsets of $S$, it is clear there is always a finite set $\bar{V}$ in $V$ containing all the data points for which $\bar{P}_n(\bar{V}) = 1$ but $P_n(\bar{V}) = 0$, and so $T_n = n^{1/2}$ with probability one. The same is true for any transformed data set $z_n$ as described above.

The problem in the previous example is not due to the inadequacies of the bootstrap or randomization methods, but rather in the choice of test statistic stemming from the fact that $V$ is too large. The example will be generalized in section 2 (example 3, testing for rotational invariance) with appropriate choices made for $V$. The point of the example is that, even if optimality considerations are set aside, some theory is needed to justify the use of the proposed tests.

The results obtained here may be summarized as follows. The bootstrap and randomization distributions are uniformly close in the following sense. If $z_n$ has distribution $P_0^R$ with $P_0$ in $\Omega_0$, then

$$\sup t \left| J_n(t, \tau \bar{P}_n) - J_n(t, P_0^R \mid G_n z_n) \right| \to 0 \quad \text{in probability}.$$  \hfill (1.5)

Moreover, each distribution, say $J_n(t, \tau \bar{P}_n)$ may be approximated by a strictly increasing continuous distribution, say $J(t, P_0)$ which is not random and depends only on $P_0$; that is, we also have

$$\sup t \left| J_n(t, \tau \bar{P}_n) - J(t, P_0) \right| \to 0 \quad \text{in probability}$$ \hfill (1.6)

and

$$\sup t \left| J_n(t, P_0) - J(t, P_0) \right| \to 0$$ \hfill (1.7)

as well. Thus, the difference in corresponding critical values tends to zero in probability:

$$d_n(\alpha, \tau \bar{P}_n) - r_n(\alpha, z_n) \to 0 \quad \text{in probability}.$$ \hfill (1.8)
Moreover, analogous results hold for the power of the tests under alternatives because critical values are still determined under the null hypothesis. In particular, (1.8) still holds if $P$ is not in $\Omega_0$. Moreover, (1.5) is also true in $\chi_n$ has distribution $P^n$. Actually, even more is true. The critical values for both procedures tend to the common finite value $J(\alpha, \tau P)$ in probability. This implies that the difference in the critical functions of the tests, evaluated at the observed data, tends to zero in probability. Also, the probability that the randomization test is randomized tends to zero. It also easily follows that both tests are consistent. In the same way, one can study the power functions of the tests against general alternatives $Q_n$ appropriately defined to yield a limiting power value less than one. For instance, suppose $Q_n$ satisfies $\delta_{\nu}(Q_n, P_0) = O(n^{1/2})$ for some $P_0$ in $\Omega_0$. Then, if $T_n$ has a limiting continuous distribution under $Q_n$ and (1.8) holds, then the power of both tests tends to the same value. The result is that the power functions of both tests may be said to be asymptotically equivalent. Hence, the randomization test may be preferable since it has exact level $\alpha$ for finite samples.

In section 2, these results are made clear, and a general methodology for proving these claims is developed. Several examples are introduced in section 3 for which the results apply. Details of the proofs are given in the appendix.

It should be pointed out that the program developed here is similar to that carried out by Hoeffding (1952). He obtained similar results for randomization tests based on test statistics arising from optimal parametric procedures. For example, he obtains results for the permutation test of whether a correlation is zero based on the optimal test statistic in the case of Gaussian data. In this case, the randomization test can only be extended to consider a null hypothesis with distributions having independent marginals. In contrast, the problems considered here are tackled from a purely nonparametric point of view. In addition, comparisons with the bootstrap are made.

Of course, one could compare bootstrap and randomization tests in those type of problems considered by Hoeffding. In particular, these are problems where the test statistics have Gaussian or Chi-squared limiting distributions, and bootstrap and randomization distributions
will also behave as approximately Gaussian or Chi-squared. By combining known results about bootstrap tests (Beran, 1986) and randomization tests (Hoeffding, 1952), analogous results would be obtained to show such tests are asymptotically equivalent in the sense described in this paper. In the problems studied here, the asymptotic theory for randomization tests is developed and compared with that of bootstrap tests (Romano, 1986). Moreover, the distributions of the test statistics considered in this paper cannot be approximated by a simple asymptotic distribution, further showing the power of simulation techniques.
SECTION 2
ASYMPTOTIC RESULTS

In this section, we outline the justification for the claims made in section 1, followed by several examples in section 3. Details of the proofs will appear in the appendix.

To summarize the problem, we now focus on the i.i.d. case. Slight extensions will be presented in examples 4 and 5. Given a sample $X_1, \ldots, X_n$ of i.i.d. $S$-valued random variables, we wish to test the null hypothesis that $\tau P = P$ for some specified $\tau$. The test statistic $T_n$ is given by (1.1) with $\delta$ given by (1.2). Furthermore, we will assume the choice of sets $V$ in the definition of $\delta$ is a countable Vapnik-Cervonenkis (V.C.) class of subsets of $S$. The restriction to countable classes of sets is to avoid measurability problems and may be weakened, but there do not appear to be any applications which warrant the need to do so. The reader may consult Romano (1986) for some discussion on the choice of $V$ in the context of the testing problem considered here.

To study the asymptotic behavior of the bootstrap and randomization tests, it is first helpful to review the bootstrap. All details may be found in Romano (1986).

In order for the bootstrap to succeed, the distribution of the test statistic $J_n(P)$ must be smooth as $P$ varies. But, smoothness in $J_n(P)$ can be traced to smoothness of the mapping $\tau$. It should be clear that the bootstrap will fail if $\tau$ is not smooth. To see why, if we wish to test the null hypothesis that $P$ belongs to some specified $\Omega_0$, one could always define $\tau$ to be $\tau P = P$ if $P$ is in $\Omega_0$, and $\tau P = P_0$ otherwise, for some $P_0$ in $\Omega_0$. In such case, bootstrap critical values will always be determined under $P_0$, so if the true underlying law is actually $P$ different from $P_0$ but still in $\Omega_0$, then there is little hope that actual Type I error of the test can be controlled.

The following smoothness condition on $\tau$ holds for the problems considered in this paper. It is assumed that $\tau$ is differentiable in the following sense: if $P \in \Omega_0$, then there exists a mapping $f(\cdot, \cdot, P)$ on $S \times V$ so that
\[ \tau Q(V) = \tau P(V) + \int f(x, V, P) d(Q - P) + o(1) \cdot \tau P \cdot \tau V \] (2.1)

as \( |Q - P| \cdot \tau V \to 0 \) and \( |P - P_0| \cdot \tau V \to 0 \) for some \( P_0 \) in \( \Omega_0 \). Here, \( \cdot \cdot \cdot | \cdot \cdot \cdot \tau V \) is the supremum norm in \( L_\infty(V) \), the metric space of real-valued bounded functions on \( V \). In order to analyze \( J_n(P_0) \), consider the process \( S_n(\cdot) \), given by

\[ S_n(V) = n^{1/2} \left[ \hat{\beta}_n(V) - \tau \hat{\beta}_n(V) \right]. \] (2.2)

Regard \( S_n \) as a random variable on \( L_\infty(V) \). Then, the test statistic \( T_n \) is just the norm of the process \( S_n \). Letting

\[ \psi(x, V, P) = 1(x \in V) - f(x, V, P) \] (2.3)

shows that the test statistic behaves approximately as \( |Z_n(\cdot)| \cdot \tau V \), where

\[ Z_n(V) = \int \psi(x, V, P_0) d(\hat{\beta}_n - P_0). \] (2.4)

Of course, \( Z_n \) is just the empirical process indexed by the class of functions \( F_V(P_0) = \{ \psi(\cdot, V, P_0), V \in V \} \). Because of the linear structure here (and assumptions on the functions \( \psi(\cdot, V, P_0) \)), \( Z_n \) is approximately a mean 0 Gaussian process \( Z \) indexed by \( V \) with covariance function

\[ \text{Cov} \left[ Z_n(V), Z_n(W) \right] = \int \psi(x, V, P_0) \psi(x, W, P_0) dP_0(x). \] (2.5)

If this approximation is valid as \( P_0 \) varies as well, then the bootstrap will succeed. In fact, uniformity in \( P \) can often be expressed in the following way. Define the metric (or possibly a pseudometric if \( V \) is not large enough) \( d_V(P, Q) \) between probabilities \( P \) and \( Q \) to be the supremum of \( |P(V) - Q(V)| \) over sets \( V \) in \( V \) and \( V \leq V \). Then, the following condition typically holds.

**Condition A.** Fix \( P_0 \) in \( \Omega_0 \). If \( P_n \) is in \( \Omega_0 \) with \( d_V(P_n, P_0) \to \infty \), then \( J_n(\cdot, P_n) \) converges weakly to a continuous strictly increasing \( J(\cdot, P_0) \).
The approach to verifying Condition A then consists in analyzing the process $S_n$ defined in (2.2). So, let $L_n(P)$ be the law of $S_n$ (as a r.v. on $L_\infty(V)$) based on $n$ observations from $P$. From the smoothness of $\tau$ via the approximation by $Z_n$ discussed above, Condition A is implied by the following.

**Condition B.** Fix $P_0$ in $\Omega_0$. If $P_n$ is in $\Omega_0$ with $d_V(P_n, P_0) \to 0$, then

$$\rho\left[L_n(P_n), L(P_0)\right] \to 0 \quad \text{as} \quad n \to \infty,$$

where $L(P_0)$ is the distribution of a mean 0 Gaussian process $Z$ with covariance given by (2.5), and $\rho$ is any metric metrizing weak convergence of probabilities on $L_\infty(V)$. Moreover, assume $Z$ has its paths in a separable subset of $L_\infty(V)$.

**Remark 2.1.** Condition B has been verified for several examples in Romano (1986). This will be made clear in several specific examples in section 3.

**Remark 2.2.** The assumption that the limit process $Z$ has separable support is needed to yield that the supremum of $|Z|_\infty$ has a continuous strictly increasing c.d.f.

**Remark 2.3.** Uniformity in weak convergence described by the metric $d_V$ may seem inappropriate at first. From the empirical processes literature (e.g., Pollard, 1984, section VII.5), one can expect the limit process $Z$ has uniformly continuous paths in the sense that if $m_V(V_n, V) \to 0$, then $Z(V_n) \to Z(V)$, where $m_V$ is the metric (or possible pseudometric) on $V$ given by

$$m_V^2(V, W) = \int \left[\psi(x, V, P_0) - \psi(x, W, P_0)\right]^2 dP_0(x).$$

Note that this metric depends on $P_0$ but the metric $d_V$ does not. In all the examples considered, it happens that $m_V$ and $d_V$ are equivalent metrics, so we have chosen to state Conditions A and B for $d_V$. 
Of course, Condition A implies that the bootstrap is valid in the sense (1.3). Furthermore, if \( b(\alpha, P) \) denotes the upper \( \alpha \)-quantile of \( J(\cdot, P) \), then the bootstrap critical value \( b_n(\alpha, \hat{P}_n) \) tends to \( b(\alpha, P_0) \) in probability if \( P_0 \) (assumed to be in \( \Omega_0 \)) is the true law. To study the consistency of the bootstrap test, assume:

**Condition C.** The map \( \tau \) is continuous in the following sense. For any sequence \( P_n \) and any \( P \), if \( d_\mathcal{V}(P_n, P) \to 0 \), then \( d_\mathcal{V}(\tau P_n, \tau P) \to 0 \).

Condition C is evidently weak and is easy to check. It seems doubtful Condition A could hold in any reasonable example without C holding anyway. Conditions A and C imply the bootstrap test is consistent against any alternative. This should be clear because Conditions A and C imply that if \( P \) is the true distribution (whether or not in \( \Omega_0 \)), then the bootstrap critical value tends to \( b(\alpha, \tau P) \) in probability, but under an alternative the test statistic \( T_n \) tends to \( \infty \) in probability. In summary, we have:

**Proposition 2.1.** Condition B implies Condition A, which implies (1.3). If Condition C holds as well, then for any \( P \) satisfying \( \delta_\mathcal{V}(P, \tau P) > 0 \), we have

\[
P \left[ T_n > b_n(\alpha, \tau \hat{P}_n) \right] \to 1 \quad \text{as} \quad n \to \infty.
\]  

(2.6)

We now proceed to analyzing the randomization test. Since, in fact, Condition B holds in all the examples we will consider, the methodology used will depend on already having verified Condition B. In particular, if we fix \( P_0 \) in \( \Omega_0 \), we already know quite a bit about the unconditional distribution \( J_n(\cdot, P_0) \) of the test statistic; in fact, \( J_n(\cdot, P_0) \) has a continuous, strictly increasing weak limit \( J(\cdot, P_0) \) so that both (1.6) and (1.7) are true. Hence, to verify (1.5) it suffices to show

\[
\sup_t \left| J_n(t, P_0^0 | G_n X_n) - J(t, P_0) \right| \to 0 \quad \text{in probability.}
\]  

(2.7)

Roughly speaking, this means we must show that the conditional distributions of \( T_n \) given the
σ-field generated by the partition of $G_n$-orbits converge weakly to the same limit as the unconditional distribution of $T_n$. The following elementary condition, due to Hoeffding (1952) implies (2.7) is true.

**Condition D.** Let $G_n$ and $G_n'$ be random transformations which are uniformly distributed over $G_n$ and independent of the observation $\xi_n$. Here, $\xi_n$ has distribution $P^n$ which need not be invariant under $G_n$. Then, $T_n(G_n\xi_n)$ and $T_n(G_n'\xi_n)$ are asymptotically independent, each with a continuous increasing limiting cdf $J(\cdot, Q)$.

It follows from Hoeffding (1952, Theorem 3.2) and the assumptions on $J(\cdot, Q)$ that Condition D implies

$$
\sup_t \left| \int_t P_0^{(n)} | G_n \xi_n ) - J(t, Q) \right| \to 0 \quad \text{in probability,}
$$

(2.8)

where $P_0^{(n)}$ is any distribution invariant under $G_n$. Actually, Hoeffding proves a pointwise (for fixed $t$) result, but a stronger statement is possible when (as is the case here) the limit distribution $J(\cdot, Q)$ is known to be continuous. Hence, in the case $P^{(n)} = P_0^n$ with $P_0$ in $\Omega_0$, $Q$ necessarily equals $P_0$ and (2.7) holds. Moreover, in the case $P^{(n)} = P^n$ but $P$ is not in $\Omega_0$, Condition D will still be verified to imply (2.8) with $Q = \tau P$, yielding consistency results about the test.

To obtain the validity of condition D, consider the process

$$
S_n(\xi_n, V) = n^{1/2} \left[ \hat{F}_n(\xi_n, V) - \tau \hat{F}_n(\xi_n, V) \right],
$$

(2.9)

where $\hat{F}_n(\xi_n, V)$ is the empirical measure of an observation $\xi_n$ from $S^{(n)}$ evaluated at the set $V$. Regard $\left[S_n(G_n\xi_n, \cdot), S_n(G_n'\xi_n, \cdot)\right]$ as a random variable on the product space of $L_\infty(V)$ with itself. The joint distribution of $\left[T_n(G_n\xi_n), T_n(G_n'\xi_n)\right]$ is just the joint distribution of $\left[|S_n(G_n\xi_n, \cdot)|_V, |S_n(G_n'\xi_n, \cdot)|_V\right]$. Hence, Condition D is trivially implied by the following.

**Condition E.** Suppose $\xi_n$ has distribution $P^{(n)}$. Then, $S_n(G_n\xi_n, \cdot)$ and $S_n(G_n'\xi_n, \cdot)$ are
asymptotically independent each with law $L(Q)$.

Remark 2.4. In order to verify Condition E in the case $P^{(n)}$ is invariant under $G_n$, one may replace $G_n$ by the identity transformation.

Now consider the case $P^{(n)} = P^n_0$ and suppose $P_0$ is in $\Omega_0$. From what we already know, $S_n(x_n, \cdot)$ and $S_n(G_n x_n, \cdot)$ each have weak Gaussian limits $L(P_0)$, so when considered jointly on the product space, the random variable

$$R_n = R_n(x_n, G_n) = [S_n(x_n, \cdot); S_n(G_n x_n, \cdot)]$$

is uniformly tight. Therefore, all we need do is analyze the finite dimensional distributions of the process $R_n$. Using the differentiability (2.1) will help establish the joint asymptotic Gaussianity of the limiting finite dimensional distributions of $R_n$, and a covariance calculation should determine the required independence.

In the case $P^{(n)} = P^n$ and $P$ is not in $\Omega_0$, Condition E is still verified with $Q=\tau P$. Sometimes, such a calculation is superfluous for the following reason. In general, the behavior of the randomization distribution only depends on the value of the $G_n$-orbit, $G_n x_n$, of $x_n$. When $P$ is not in $\Omega_n$, it may be the case that the distribution of (the orbit-valued random variable) $G_n x_n$ has the same distribution under some $P_0$ in $\Omega_0$ as under $P$. In this case, the appropriate choice of $\tau$ should satisfy $\tau P = P_0$. The result is that the analysis of the randomization distribution under an alternative hypothesis $P$ is the same as under $P_0$.

In summary, we have the following.

Proposition 2.2. Condition E implies Condition D. If Condition D holds when $P^{(n)} = P^n$ for some $P$ not in $\Omega_0$ with $Q = \tau P$, then the randomization critical value $r_n(\alpha, x_n)$ tends to $J(\alpha, \tau P)$ in probability. Hence, if Condition C holds as well, the randomization test is consistent.
We are now in a position to more fully discuss a comparison of the two procedures in the i.i.d. case. Assume Conditions A-C hold and D holds when \( P^{(n)}=P^0 \). For any fixed \( P \) (in \( \Omega_0 \) or not), the difference between the randomization and bootstrap distributions tends to 0 uniformly in \( t \); that is, Propositions 2.1 and 2.2 yield

\[
\sup_t \left| J_n(t, \tau \hat{x}_n) - J_n(t, P^0_0 \mid G_n \xi_n) \right| \to 0 \quad \text{in } P^n-\text{probability}, \tag{2.10}
\]

where \( P_0 \) is any member of \( \Omega_0 \). It follows (using 2.8) that the difference in critical values tends to zero in \( P^n \)-probability. This implies that the probability that the tests differ (in whether or not to reject the null hypothesis) tends to zero. Furthermore, both tests are consistent against all alternatives. However, such a result is practically useless in comparing the power of the two tests since presumably one could construct many consistent tests. In order to obtain a more useful result, a more refined analysis is needed. Consider a sequence of alternatives \( Q_n \) to \( \Omega_0 \) and study the asymptotic power (if it exists) against such a sequence. In order to get an interesting limit for the asymptotic power, it should be clear that one needs to get close to \( \Omega_0 \) and, in fact, \( Q_n \) should satisfy \( \delta_{\nu}(Q_n, \tau Q_n) = O(n^{-1/2}) \). The next proposition gives conditions when the bootstrap and randomization distributions are uniformly close under general sequences of alternatives \( Q_n \), and when the limiting power of each test against \( Q_n \) is the same.

**Proposition 2.3.** Assume Conditions (A)-(C) hold. Let \( Q_n \) be a sequence of alternatives to \( \Omega_0 \) satisfying \( \delta_{\nu}(Q_n, Q_\infty) \to 0 \). Assume Condition (E) holds with \( P^{(n)}=Q_n^n \) and \( Q=\tau Q_\infty \). Then,

\[
\sup_t \left| J_n(t, \tau \hat{x}_n) - J_n(t, P^0_0 \mid G_n \xi_n) \right| \to 0 \quad \text{in } Q_n^n-\text{probability}, \tag{2.11}
\]

where \( P_0 \) is any member of \( \Omega_0 \). Also, the corresponding critical values satisfy

\[
b_n(\alpha, \tau \hat{x}_n) - r_n(\alpha, \xi_n) \to 0 \quad \text{in } Q_n^n-\text{probability.} \tag{2.12}
\]

Furthermore, suppose the limiting distribution of \( T_n \) based on a sample of size \( n \) from \( Q_n \) is
continuous; that is, there exists a continuous function \( H(t) \) so that

\[
Q_n^a(T_n \geq t) \rightarrow H(t).
\]

Then, the asymptotic power of both tests against the sequence of alternatives \( Q_n \) is

\[
H \left[ b(\alpha, \tau Q_\infty) \right].
\]

**Remark 2.5.** The function \( H(t) \) could be identically 1, yielding the test is consistent against the sequence of alternatives \( Q_n \).

**Remark 2.6.** In order to compute the limiting distribution of \( T_n \) under \( Q_n^a \) (or show its existence), consider the process \( S_n \) defined in (2.2), where \( \hat{P}_n \) is the empirical based on a sample of size \( n \) from \( Q_n \). Suppose \( Q_n \) satisfies an approximation like \( Q_n = P_0 + \Delta n^{-1/2} \), where \( \Delta \) is some element in \( L_\infty(V) \). The differentiability condition (2.1) shows \( S_n \) is approximately a Gaussian process with covariance (2.5), but this time with mean \( \Delta \). Thus, the limiting distribution of \( T_n \) under \( Q_n^a \) is the distribution of the supremum of such a process. It should be the case that such a distribution is necessarily continuous as in the case \( \Delta = 0 \), and perhaps a proof of this could be based on the proof of Proposition 2 in Beran and Millar (1986a). In any case, it is clear that the above proposition holds as long as \( H \) is continuous at \( b(\alpha, \tau Q_\infty) \) and the difference in asymptotic power of the two procedures could only be as large as the jump in \( H \) at \( b(\alpha, \tau Q_\infty) \).

**Remark 2.7.** Why use the tests considered in this paper anyway? The power properties of supremum or Kolmogorov-Smirnov distance type tests have been well-studied in several problems; see [5] and [8], for example. It is clear that these properties hold quite generally in problems of the type considered in this paper. To be more specific, consider a sequence of alternatives \( Q_n \) to \( Q_0 \) satisfying an approximation like \( Q_n \approx P_0 + \Delta \varepsilon_n \). Then, the argument in Remark 2.6 shows the power of the bootstrap and randomization tests against \( Q_n \) tends to 1 if \( (\varepsilon_n/n^{1/2}) \rightarrow \infty \). Moreover, in the case \( \varepsilon_n = n^{1/2} \), the power can be made arbitrarily close to 1 if
$|\Delta|$ is large. To see why, consider the following simple argument. Choose $V_0$ so $\Delta(V_0) \neq 0$ and let $\Delta_0 = \Delta(V_0)$. Then, the power of the bootstrap test (for example) is bounded below by

$$Q_n \left[ n^{1/2} \left| \hat{\Delta}_n(V_0) - \tau \hat{\Delta}_n(V_0) \right| \geq b_n(\alpha, \tau \hat{\Delta}_n) \right].$$

The distribution of the random variable on the left side of the inequality tends to the distribution of the absolute value of a real-valued Gaussian random variable with mean $\Delta_0$ (and variance which only depends on $P_0$ and $V_0$ and not the original choice of $\Delta$). Also, the right side of the inequality tends to $b(\alpha, P_0)$ in probability. Hence, the limiting power is bounded below by

$$P \left[ |Z + \Delta_0| \geq \sigma b(\alpha, P_0) \right],$$

for some $\sigma = \sigma(V_0, P_0)$. Now, increase $\Delta_0$.

In summary, the limiting power of supremum tests against a sequence of alternatives converging to $\Omega_0$ at the $n^{1/2}$ rate is not degenerate. In typical smooth problems (as in the examples to come), this is the best obtainable rate.
SECTION 3
EXAMPLES

In this section, the methodology of the previous section is applied to several examples. To avoid measurability problems, the collection of sets \( V \) in the definition of the test statistic will always be countable. This may be weakened, but there does not seem to be a need to do so. From section 1, we have seen that \( V \) can not be chosen to be too large. An appropriate weak restriction is that \( V \) be a Vapnik-Cervonenkis (V.C.) class of sets. For a discussion of V.C. classes and the choice of \( V \), see section 2 of Romano (1986). Tacit is the assumption that we can choose \( V \) large enough to be a V.C. class and so that \( \delta_V \) is indeed a metric. Actually, many such choices for \( V \) exist in typical applications, as the ones discussed below. In general, the results may apply to testing \( \delta_V(P, \tau P) = 0 \).

Example 1, Testing Independence, continued.

Proposition 3.1. Assume \( V_j \) (in the definition of the test statistic \( T_n \)) is a countable V.C. class in \( S_j \). Conditions A-E all hold when \( P^{(n)} = P^n \) for any fixed \( P \) (whether \( P \) is in \( \Omega_0 \) or not). Thus, both bootstrap and randomization tests are consistent and asymptotically equivalent in the sense (2.10). Moreover, suppose \( Q_n \) is any sequence of alternatives to \( \Omega_0 \) satisfying \( d_V(Q_n, Q_\infty) \rightarrow 0 \) for some \( Q_\infty \). Then, Condition E holds for \( P^{(n)} = Q_n^n \) and \( Q = \tau Q_\infty \), so that (2.11) and (2.12) follow.

Example 2, Testing for Rotational Invariance. The problem is to test whether the underlying probability distribution on \( S = \mathbb{R}^p \) belongs to the class \( \Omega_0 \) of rotationally invariant or spherically symmetric distributions. Let \( S_r \subset S \) be the sphere of radius \( r \) and center 0. If \( X = (X_1, \ldots, X_p) \) has probability distribution \( P \) on \( \mathbb{R}^p \), then \( P \) is completely specified by the marginal distribution \( P_R \) of \( R = \left( \sum_{i=1}^p X_i^2 \right)^{1/2} \) and the conditional distribution \( P_{X|R} \) of \( X \) given
Of course, if \( P \) is spherically symmetric, then \( P_{X \mid R \sim r} \) is always uniformly distributed on \( S_r \). Let \( \tau P \) be the distribution \( Q \) in \( \Omega_0 \) such that \( P_R = Q_R \). Then, the proposed test statistic becomes

\[
T_n = \frac{1}{n} \max_{V \in \mathcal{V}} \left| \hat{P}_n(V) - \tau \hat{P}_n(V) \right|
\]  \hspace{1cm} (3.1)

To see that a randomization test is applicable, we must identify the appropriate class of transformations \( G_n \) on \( S^p \). If \( \theta \) is a point on \( S_1 \) and \( x \) is a point on \( S \), let \( g_{\theta \cdot x} \) be the new point \( \theta \cdot x \cdot \frac{1}{p} \), where \( \cdot \cdot \cdot \cdot \cdot \) is the usual euclidean norm on \( \mathbb{R}^p \). Then,

\[
G_n = \left\{ (g_{\theta_1}, \ldots, g_{\theta_n}); \theta_i \in S \right\}
\]

so an element \((g_{\theta_1}, \ldots, g_{\theta_n})\) of \( G_n \) transforms a point \( x_n = (X_1, \ldots, X_n) \) in \( S^n \) to the point \((g_{\theta_1} X_1, \ldots, g_{\theta_n} X_n)\).

Notice that \( G_n \) is not finite except in the case \( p = 1 \). Thus, the description of the test given in section 1 is not quite accurate. In this case, \( J_n (P^X \mid G_n x_n) \) still denotes the conditional distribution of \( T_n \) (when \( x_n \) has distribution \( P^n \)) given that \( x_n \in G_n x_n \). The randomization test then refers to an appropriate critical value from \( J_n (P_0^n \mid G_n x_n) \), where \( P_0 \) is some distribution in \( \Omega_0 \). The important fact is that \( J_n (P_0^n \mid G_n x_n) \) is the distribution of \( T_n (G_n x_n) \), where \( x_n \) is any element of \( G_n x_n \) and \( T_n \) is uniformly distributed over \( G_n \). The methodology described in section 2 for comparing the bootstrap and randomization tests is then applicable. The only technicality involved is showing that Hoeffding's Condition D implies (2.7), but an easy generalization of his proof shows the argument carries over as long as \( J_n (P^n \mid G_n x_n) \) is the distribution of \( T_n (x_n) \) when \( x_n \) is uniform over \( G_n x_n \). In general, one needs to be able to put a uniform probability distribution on \( G_n \). In the example here, it is clear how to do this.

To see the difference between the bootstrap and randomization methods in this situation, the following algorithm may be employed.

**I. Computing the Bootstrap distribution.**

**Step 1.** Given the sample \( x_n = (X_1, \ldots, X_n) \) with \( R_i = |X_i|_p \), let \((R_1^*, \ldots, R_n^*)\) be \( n \) variables
uniformly chosen with replacement from \((R_1, \ldots, R_n)\). Let \((\theta_1^*, \ldots, \theta_n^*)\) be \(n\) i.i.d. points uniformly distributed over \(S_1\). Let \(x_n^*\) be a new data set with \(x_n^* = (X_1^*, \ldots, X_n^*)\) and \(X_j^* = R_j^* \cdot \theta_j^*\).

**Step 2.** A data set \(x_n^*\) constructed as in step 1 is called a bootstrap data set. Repeat step 1 \(B\) independent times to generate data sets \(x_{n,b}^*\) for \(b = 1, \ldots, B\). For each such data set, compute \(T_n(x_{n,b}^*)\).

**Step 3.** The distribution assigning equal mass to these \(B\) values serves as an approximation to \(J_n(\tau \hat{P}_n)\).

**II. Computing the Randomization distribution.**

**Step 1.** A new data set \(x_n^*\) is constructed as in step 1 for the bootstrap procedure, but this time the \(R_j^*\)'s are sampled without replacement from \((R_1, \ldots, R_n)\).

**Step 2.** Repeat step 1 \(B\) independent times to generate samples \(x_{n,b}^*\). Compute \(T_n(x_{n,b}^*)\) for each data set.

**Step 3.** The distribution assigning equal mass to these \(B\) values serves as an approximation to \(J_n(P^n_0 | G_n x_n)\).

To describe the class of sets \(V\) allowed in (3.1), embed the sample space \(S\) into \(S_1 \times \mathbb{R}\), where \(\mathbb{R}\) denotes the nonnegative real numbers. A point \(x\) in \(S\) is identified with the point \((s_1, s_2)\) in \(S_1 \times \mathbb{R}\) if it is at distance \(s_2\) from the origin and \(x/||x||_p = s_1\). In the case \(x\) is the origin, identify it with \((0, 0)\).

**Proposition 3.1.** Assume the collection of sets \(V\) in (3.1) is of the form \(V = V_1 \times V_2\), where \(V_1\) is a (countable) V.C. class in \(S_1\) and \(V_2\) is a (countable) V.C. class in \(\mathbb{R}\). Conditions A-E all hold when \(P^{(n)} = P^n\) for any fixed \(P\) (whether \(P\) is in \(\Omega_0\) or not). Thus, both bootstrap and randomization tests are consistent and asymptotically equivalent in the sense (2.10). Moreover, suppose \(Q_n\) is any sequence of alternatives to \(\Omega_0\) satisfying \(d_V(Q_n, Q_\omega) \rightarrow 0\) for some \(Q_\omega\). Then, Condition E holds for \(P^{(n)} = Q_n^n\) and \(Q = \tau Q_\omega\), so that (2.11) and (2.12) follow.
A slight modification of the previous set-up is needed in some situations. Suppose \( \Omega_0 \) is now specified as the set of probabilities \( P \) satisfying \( \tau_j(P) = P \) for \( 1 \leq j \leq k \), where \( \tau_j \) is a mapping from \( \Omega_0 \). Then, \( P \) lies in \( \Omega_0 \) if and only if \( \max_{1 \leq j \leq k} \delta(P, \tau_j(P)) = 0 \) and the proposed test rejects for large values of \( \max_{1 \leq j \leq k} \delta(\hat{P}_n, \tau_j(\hat{P}_n)) \).

Example 3, Testing whether a probability law is exchangeable. Let \( X_1, \ldots, X_n \) be i.i.d. \( S \)-valued random variables, where \( X_i = (X_{i,1}, \ldots, X_{i,d}) \) is made up of \( d \) components each living in a space \( E \). Let \( P \) be the probability law generating the data. The problem is to test whether \( P \) is exchangeable. That is, if \( D = \{1, \ldots, d\} \) and \( \pi_j: D \to D, 1 \leq j \leq d \), are the \( d! \) permutations of \( D \), then the problem is to test whether the law of \( (X_{i,1}, \ldots, X_{i,d}) \) is the same as the law of \( (X_{i,\pi_j(1)}, \ldots, X_{i,\pi_j(d)}) \) for every \( j \). Given any probability \( P \) on \( S \), let \( \tau_jP \) denote the law of \( (X_{i,\pi_j(1)}, \ldots, X_{i,\pi_j(d)}) \) if \( (X_1, \ldots, X_d) \) has law \( P \). Then, the proposed test statistic takes the form

\[
T_n = n^{\frac{1}{2}} \max_{j} \sup_{V \in \mathcal{V}} \left| \hat{P}_n(V) - \tau_j \tilde{P}_n(V) \right|.
\]

In the case \( d = 2 \), we can delete the max in this expression to get an equivalent test statistic

\[
T_n = n^{\frac{1}{2}} \sup_{V \in \mathcal{V}} \left| \hat{P}_n(V) - \tau \tilde{P}_n(V) \right|,
\]

(3.2)

where \( \tau \hat{P}_n = \frac{1}{2} \left[ \tau_1 \hat{P}_n + \tau_2 \tilde{P}_n \right] \). Note that, when \( d = 2 \), \( \tau P = P \) if and only if \( P \in \Omega_0 \), and the test statistic is of the form (1.1). In this case, \( P \) is exchangeable if \( P(V) = P(V') \) for all (measurable) \( V \), where \( V' \) is the set \( \left\{ (y, x) : x \in E, y \in E, (x, y) \in V \right\} \).

The appropriate group of transformations \( G_n \) for this problem may be described as follows. A transformation \( g \) in \( G_n \) takes an element \( x_n = (X_1, \ldots, X_n) \) to an element \( y_n = (Y_1, \ldots, Y_n) \) if \( Y_i \) is some permutation of \( X_i \). That is, \( Y_n = (X_{i,\pi(i)}, \ldots, X_{i,\pi(d)}) \) for some permutation \( \pi \) of \( D \). The permutation \( \pi \) transforming \( X_i \) may also depend on \( i \) so that \( g \) may
be identified by a vector \( \pi = (\pi^{(1)}, \ldots, \pi^{(n)}) \) where \( \pi^{(k)} \) is some permutation of \( D \). Then, \( G_n \) is the collection of all such \( g \), so \( G_n \) has \( (d!)^n \) elements.

To avoid introducing new notation, we restrict attention to the case \( d = 2 \). Suffice it to say that bootstrap and randomization tests are asymptotically equivalent procedures in the sense described here even if \( d > 2 \). The asymptotics for the bootstrap for \( d > 2 \) are given in Romano (1986).

**Proposition 3.1.** Let \( V \) be a countable V.C. class in the definition of the test statistic \( T_n \) (3.2). Assume \( V \) contains elements \( V \) with \( V \neq V' \). Then, Conditions A-E all hold when \( P^{(n)} = P^n \) for any fixed \( P \) (whether \( P \) is in \( \Omega_0 \) or not). Thus, both bootstrap and randomization tests are consistent and asymptotically equivalent in the sense (2.10). Moreover, suppose \( Q_n \) is any sequence of alternatives to \( \Omega_0 \) satisfying \( d_Y(Q_n, Q_\infty) \to 0 \) for some \( Q_\infty \). Then, Condition E holds for \( P^{(n)} = Q_n^n \) and \( Q = \tau Q_\infty \), so that (2.11) and (2.12) follow.

**Example 4.** \( K \)-sample Test of Homogeneity. The structure of the previous tests can easily be adapted to \( k \) independent samples from possibly different populations. For \( i = 1, \ldots, k \), let \( X_{i,j}, 1 \leq j \leq n_i \) be a sample of \( S \)-valued random variables with probability distribution \( P_i \). The problem is to test the homogeneity hypothesis \( H_0: P_1 = \cdots = P_k \). Let \( \hat{P}_n \) be the empirical measure of all \( n = \sum_{i=1}^k n_i \) observations combined and let \( \hat{P}_{n,i} \) be the empirical measure of the \( i \)-th sample. Then, one possible test statistic for testing \( H_0 \) takes the form

\[
T_n = n^{\frac{1}{2}} \max_{1 \leq i \leq k} \left[ c_{n,i} \sup_{V \in V} \left| \hat{P}_n(V) - \hat{P}_{n,i}(V) \right| \right],
\]

where the \( c_{n,i} \) are constants depending on the sample sizes \( n_i \) and, again, \( V \) is an appropriate class of sets. One possible choice for \( c_{n,i} \) is \((n_i/n)^{1/2}) \). An alternative, but similar, test statistic is

\[
T_n = n^{\frac{1}{2}} \max_{i,j} \left[ c_{n,i,j} \sup_{V \in V} \left| \hat{P}_{n,i}(V) - \hat{P}_{n,j}(V) \right| \right].
\]
Special cases of these tests were proposed by Smirnov (1939) and Kiefer (1959) in the case \( S = \mathbb{R}, \mathbf{V} = \{(-\infty, t]: t \in \mathbb{R}\} \), and the assumption that the underlying probability distribution is continuous. In this case, the test statistic is distribution-free under the null hypothesis. This distribution-free property does not extend to \( S = \mathbb{R}^p \) when \( p > 1 \) even if the underlying population is assumed to be continuous. Bickel (1969) considers the two-sample problem \((k = 2)\) in the case \( S = \mathbb{R}^p \) and \( \mathbf{V} = \{(-\infty, t]: t \in \mathbb{R}^p\} \), and shows that the randomization test is consistent against (fixed) alternatives. Here, the sample space \( S \) is arbitrary and \( \mathbf{V} \) is assumed to be any (countable) V.C. class. No assumptions at all (such as continuity assumptions) are made on the underlying population.

For simplicity, assume the sample size \( n_i \) is a function of \( n \), where \( n_i = n_i(n) \) is the integer part of \( \lambda_i n \) if \( i < k \) and \( n_k = n - \sum_{i < k} n_i \). Let \( J_n(P_1, \ldots, P_k) \) be the sampling distribution of \( T_n \) when the \( k \) samples have distributions \( P_1, \ldots, P_k \), and let \( J_n(t, P_1, \ldots, P_k) \) be the corresponding survival function. The bootstrap null distribution is then \( J_n(\hat{P}_n, \ldots, \hat{P}_n) \).

To see that a randomization test is applicable, we must identify \( G_n \). The sample space \( S^{(n)} \) is \( S^n \). Think of the observed data \( \mathbf{x}_n \) as a vector of length \( n \), ordered so that the first \( n_1 \) data values of \( \mathbf{x}_n \) are thought of as coming from the first population, the next \( n_2 \) data values from the second, and so on. Let \( \pi \) be a permutation of the integers 1 to \( n \) and if \( \mathbf{x}_n = (X_1, \ldots, X_n) \) is in \( S^n \), let \( g\mathbf{x}_n = (X_{\pi(1)}, \ldots, X_{\pi(n)}) \). Then, \( G_n \) is the collection of all such \( g \). Under the null hypothesis, \( g\mathbf{x}_n \) and \( \mathbf{x}_n \) have the same distribution. As before, let \( J_n(P^{(n)}|G_n, \mathbf{x}_n) \) denote the conditional distribution of the test statistic \( T_n(\mathbf{x}_n) \) given that \( \mathbf{x}_n \) falls in \( G_n \) and \( \mathbf{x}_n \) has distribution \( P^{(n)} \). For the remainder, we focus on the test statistic (3.3). Similar results could be obtained for (3.4).

**Proposition 3.4.** Let \( \mathbf{V} \) be a (countable) V.C. class in (3.3). Assume the sample sizes \( n_i \) satisfy \( n_i/n \to \lambda_i \) as \( n \to \infty \) for some \( \lambda_i \) in \((0, 1)\) and \( c_{n,i} \to c_i \) for some \( c_i > 0 \). Let \( Q^{(n)} = \bigwedge_{j=1}^k Q_n^{(n)} \) be a sequence of possible distributions of the data, so that \( Q_{n,1} \) represents the
distribution of the $i$th sample. Assume $d_{V}(Q_{n,i}, Q_{\infty,i}) \to 0$ as $n \to \infty$, for some probability $Q_{\infty,i}$ on $S$. Then,
\[
\sup_{t} \left| J_{n}(t, \hat{P}_{n}, \ldots, \hat{P}_{n}) - J_{n}(t, P(\cdot) \mid G_{n}X_{n}) \right| \to 0 \quad \text{in } Q^{(n)} \text{ probability},
\]
where $P^{(n)} = P^{n}$ is any distribution made up of $n$ i.i.d. components. Moreover, the corresponding critical values of the bootstrap and randomization tests tend to a common finite limit in probability. Also, bootstrap and randomization tests are consistent against any alternative of the form $Q^{(n)} = \prod_{j=1}^{k} Q_{j}^{n_{j}}$, if the $Q_{j}$ are not all equal.

**Remark 3.1.** Similar remarks to those presented after Proposition 2.3 are applicable in this context as well. In particular, let the hypothesis on $Q^{(n)}$ in Proposition 3.4 be satisfied. If the distribution of the test statistic under $Q^{(n)}$, namely $J_{n}(Q_{n,1}, \ldots, Q_{n,k})$ tends weakly to a continuous distribution, then the power of the randomization and bootstrap tests tends to a common value.

**Example 5, Testing For a Change Point.** Let $X_{n} = (X_{1}, \ldots, X_{n})$ be a sample of $n$ independent random variables taking values in a sample space $S$. The null hypothesis asserts that the $X_{i}$'s have a common (unknown) distribution $P$. The alternative hypothesis asserts that, for some $J$, $\bar{X}_{J} = (X_{1}, \ldots, X_{J})$ are i.i.d. with a distribution $P_{1}$, and $\bar{X}_{J} = (X_{J+1}, \ldots, X_{n})$ are i.i.d. with a different distribution $P_{2}$. Let $\hat{P}_{j}$ be the empirical of the first $j$ observations and let $\hat{Q}_{j}$ be the empirical of all observations but the first $j$. In the spirit of the test statistics considered in this paper, a natural test statistic might be
\[
T_{n} = \max_{1 \leq j \leq n} c_{n,j} \delta_{V}(\hat{P}_{j}, \hat{Q}_{j}),
\]
where $\delta_{V}$ is the metric (1.2) and $c_{n,j}$ is some sequence of norming constants. One possible choice is
\[
c_{n,j} = \left[ \frac{n}{j(n-j)} \right]^{-1/2}.
\]
As in the other examples, both bootstrap and randomization methods apply. In particular, the bootstrap method consists in resampling (conditional on the data) $n$ i.i.d. points from the empirical distribution of the data. The randomization method, on the other hand, consists in generating samples of size $n$ by sampling the data without replacement, or equivalently, transforming $x_n$ into $y_n = (X_{\pi(1)}, \ldots, X_{\pi(n)})$ for some permutation $\pi$ of $\{1, 2, \ldots, n\}$. The methods applied in this paper can be extended to analyze and compare the two techniques. The details are deferred to a forthcoming technical report.

The reader may consult [11] for references on the change-point problem. In [11], this problem is considered in a parametric setting where the data distribution is known to be Gaussian with unknown mean and known variance.
SECTION 4
STOCHASTIC APPROXIMATIONS TO NULL DISTRIBUTIONS

The attractiveness of the testing procedures described in this paper is marred by the burgeoning amount of computation involved. Monte Carlo approximations to bootstrap null distributions are described in Romano (1986). Here, we focus on the problem of implementation of the randomization procedure. Two main difficulties are apparent.

(i). The exact value of the observed test statistic $T_n(x_n)$ may be difficult to obtain because the supremum of $\hat{P}_n - \tau \hat{P}_n$ over the class of sets $V$ may be hard to compute. Instead, one may have to resort to computing the supremum discrepancy between $\hat{P}_n$ and $\tau \hat{P}_n$ over a finite number of search sets. One possibility is to choose these search sets $V_1, \ldots, V_s$ i.i.d. (and independent of $x_n$) according to a probability on $V$.

(ii). The randomization null distribution, which assigns equal mass to the values $T_n(g; x_n)$, $g \in G_n$, may be hard to compute because the number of elements in $G_n$ is too large. Here, one can sample $g_1, \ldots, g_r$ from $G_n$ with or without replacement and approximate the randomization distribution by the distribution assigning equal mass to the values $T_n(g_i; x_n)$, $1 \leq i \leq r$.

Of course, if problem (i) is present, then $T_n(g_i; x_n)$ will have to be approximated as well. One possibility is to use the same search sets $V_1, \ldots, V_s$ for all $i$. An alternative possibility is to allow the search sets to depend on $i$; that is, for each $i$ choose $s$ independent random sets from $V$.

The important fact is the following. The above stochastic approximations do not affect the level of the test, so that the resulting test has exact desired level $\alpha$. This follows because the random vector $\left[ T_n(x_n), T_n(g_1; x_n), \ldots, T_n(g_r; x_n) \right]$ is still exchangeable, even if the $g_i$'s are chosen at random with or without replacement from $G_n$, or if $T_n$ is computed by replacing the supremum over $V$ by a maximum over random search sets. In the case the same $V_j$'s are used
to approximate \( T_n(g; x_n) \) for each \( i \), an easy way to see that the sets \( V_j \) chosen do not change this exact finite sampling result is by conditioning on these chosen sets and then regard the resulting test statistic as a new test statistic in its own right of the form described in section 1.

For the remainder of this section, we focus on the asymptotic behavior of the randomization test employing such stochastic approximations. A general methodology to handle such computational difficulties is described in Beran and Millar (1986b), where it is suggested to approximate a supremum over a collection of sets by a maximum over randomly chosen sets in the context of constructing a confidence set for a measure. In general, their arguments presented in this case extend to cover the approximations described in (i) and (ii) above, where in (ii) the \( g_i \)'s are chosen with replacement from \( G_n \). In general, consistency results and the equivalency of stochastic bootstrap tests and stochastic randomization tests continue to hold, assuming they hold without the use of stochastic approximations. In order to handle the more natural case (in the context here) where the \( g_i \)'s are chosen without replacement from \( G_n \), the following result applies. We give a more general result than what is needed here, and it may be considered a Glivenko-Cantelli theorem for sampling without replacement in the general setting of Vapnik and Cervonenkis.

**Proposition 4.1.** Let \( P_n \) be a sequence of probabilities on a space \( S \). Suppose \( P_n \) represents the distribution for a finite population \( X_n \) of \( N_n \) elements, some of which may be equal. Let \( x_n = (X_{n,1}, \ldots, X_{n,r_n}) \) be a sample of size \( r_n \) chosen at random without replacement from \( X_n \). Let \( \hat{P}_n \) be the empirical measure corresponding to these \( r_n \) values; that is,

\[
\hat{P}_n(V) = r_n^{-1} \sum_{i=1}^{r_n} 1(X_{n,i} \in V).
\]

Let \( V \) be a (countable) Vapnik-Cervonenkis class of (measurable) subsets of \( S \). Then, if \( r_n \to \infty \),

\[
\sup_{V \in \mathcal{V}} \left| \hat{P}_n(V) - P_n(V) \right| \to 0 \quad \text{in} \quad P_n^{r_n}-probability.
\]
Remark 4.1. The case $P_n = P$, independent of $n$, is trivial. Note, however, that no assumptions are made on the sequence $P_n$.

Corollary 4.1. Under the conditions of Proposition 4.1., suppose $S = \mathbb{R}$. Let $F_n$ be the cdf corresponding to the population $X_n$, and let $\hat{F}_n$ be the empirical cdf based on $r_n$ observations chosen without replacement from $F_n$. Then,

$$
sup_t \left| \hat{F}_n(t) - F_n(t) \right| \to 0 \quad \text{in probability.}
$$

Corollary 4.1 can be applied to stochastic null distributions in the following way. As before, if $P_0^{(n)}$ is invariant under $G_n$, then the randomization null distribution $J_n(P_0^{(n)} | G_n x_n)$ is the distribution corresponding to the population $X_n$ of $M_n$ elements $T_n(g x_n), g \in G_n$. A stochastic approximation to $J_n(x, P_0^{(n)} | G_n x_n)$, is $\hat{J}_n(x, P_0^{(n)} | G_n x_n)$ given by the empirical distribution of $r_n$ values, $T_n(g_j x_n), 1 \leq j \leq r_n$, where the $g_j$ are sampled without replacement from $G_n$. Having established (2.8) that the randomization null distribution can be uniformly approximated by a continuous distribution $J(t, Q)$, it follows (by applying Corollary 4.1 conditional on $x_n$) that the stochastic approximation $\hat{J}_n$ has the same property,

$$
sup_t \left| J_n(t, P_0^{(n)} | G_n x_n) - J(t, Q) \right| \to 0 \quad \text{in probability.}
$$

Then, for example, a critical value $r_n$ based on $\hat{J}_n$ tends to $r(\alpha, Q)$ in probability, and consistency of the stochastic randomization test then follows as before.
SECTION 5
PROOFS

Proof of Proposition 3.1. Conditions A-C are verified in Romano (1986). Indeed, \( \tau \) is differentiable in the sense (2.1) with \( f(x, V, P) \) given by

\[
f(x, V, P) = \sum_{j=1}^{d} r_j 1(x \in A_j),
\]

where \( A_j \) is a set in \( \mathbb{V}_j \) so that \( V = \times_{j=1}^{d} A_j \), and \( r_j = r_j(P) = P(\mathbb{V}_j|P_j(A_j)) \). Let \( P_n \) be a sequence in \( \Omega_0 \) satisfying \( d_{\mathbb{V}}(P_n, P_0) \to 0 \) for some \( P_0 \) in \( \Omega_0 \). Then, from Romano (1986), \( J_n(P_n) \) converges weakly to a continuous, strictly increasing limit law \( J(P_0) \), and \( J(P_0) \) is the distribution of the norm of a certain Gaussian process \( L(P_0) \). We now verify Condition E when \( P^{(n)} = Q_n \) and \( Q = \tau Q_n \), where \( Q_n \) is a sequence satisfying \( d_{\mathbb{V}}(Q_n, Q_n') \to 0 \). Let \( S_n(\mathbb{V}_n, \mathbb{V}) \) be given by (2.9). Here, \( \mathbb{V}_n = (X_{n,1}, \ldots, X_{n,n}) \) is a vector of \( n \) i.i.d. variables with distribution \( P_n \), and \( X_{n,i} \) is made up of \( d \) independent components \( (X_{n,i,1}, \ldots, X_{n,i,d}) \). Let \( G_n \) and \( G_n' \) be independent of \( \mathbb{V}_n \) and each other, each uniformly distributed over \( G_n \). We must show \( S_n(G_n \mathbb{V}_n, \cdot) \) and \( S_n(G_n' \mathbb{V}_n, \cdot) \) are asymptotically independent with law \( L(Q) \). From the discussion after Remark 2.4, it suffices to examine the finite dimensional distributions of the process \( \left[ S_n(G_n \mathbb{V}_n, \cdot), S_n(G_n' \mathbb{V}_n, \cdot) \right] \). Using the differentiability of \( \tau \) (or direct verification) shows

\[
S_n(\mathbb{V}_n, V) = n^{-1/2} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[ Z_{n,i,j} - E(Z_{n,i,j}) \right] \right\} + o_{P_n}(1),
\]

where \( Z_{n,i,j} = 1(X_{n,i,j} \in A_j) \) and \( V = \times_{j=1}^{d} A_j \). Thus, Lemma 5.1 is applicable to yield that \( \left[ S_n(G_n \mathbb{V}_n, V), S_n(G_n' \mathbb{V}_n, W) \right] \) converges weakly to a bivariate Gaussian distribution with correlation zero. An argument similar to the proof of Lemma 5.1 shows a finite linear combination of elements \( S_n(G_n \mathbb{V}_n, V_i) \) is independent of a linear combination of elements \( S_n(G_n' \mathbb{V}_n, W_i) \). The result follows.
Lemma 5.1. For $1 \leq i \leq n$, let $(Y_{n,i,1}, \ldots, Y_{n,i,d})$ be $n$ i.i.d. random vectors made up of $d$ independent components. Moreover, suppose $Y_{n,i,j}$ has mean 0 and variance $\sigma_{n,j}^2$, and is bounded in absolute value by 1. Suppose the law of $Y_{n,i,j}$ converges weakly (as $n \to \infty$) to a distribution with variance $\sigma_j^2$ so that $\sigma_{n,j}^2 \to \sigma_j^2$. For $1 \leq j \leq d$, let $G_{n,j}$ and $G_{n,j}'$ be independent random permutations of \{1, 2, ..., n\}. Define

$$T_n = n^{-1/2} \sum_{i=1}^{n} \left[ \prod_{j=1}^{d} Y_{n,G_{n,j}'(i),j} \right]$$

and

$$W_n = n^{-1/2} \sum_{i=1}^{n} \left[ \prod_{j=1}^{d} Y_{n,G_{n,j}(i),j} \right].$$

Then, the law of $(T_n, W_n)$ converges weakly to the law of a Gaussian random variable $(T, W)$, where $T$ and $W$ are i.i.d. with mean 0 and variance $\sigma^2 = \prod_{j=1}^{d} \sigma_j^2$.

Proof of Lemma 5.1. To simplify the notation, consider the case $d=2$. The general case is similar. Let $X_{n,i} = Y_{n,i,1}$ and $Y_{n,i} = Y_{n,i,2}$, and let $H_n = G_{n,2}$ and $G_n = G_{n,1}$. Notice that the conditional distribution of $(T_n, W_n)$ given $(G_n, G_n')$ is independent of the actual values $(G_n, G_n')$. Hence, we may take $G_{n,i}'$ to each be the identity transformation. For any constants $a$ and $b$, we must show that the law of

$$aT_n + bW_n = n^{-1/2} \sum_{i=1}^{n} X_{n,i} (aY_{n,H_n(i)} + bY_{n,G_n(i)})$$

converges weakly to a Gaussian random variable with mean 0 and variance $(a^2 + b^2)\sigma^2$. If the $Y_{n,i}$s, $H_n$, and $G_n$ were fixed constants, a triangular array version of Theorem A10 of Hettmansperger (1984) (easily proved by Lindeberg) shows that a sufficient condition is

$$n^{-1} \sum_{i=1}^{n} \left[ aY_{n,H_n(i)} + bY_{n,G_n(i)} \right]^2 \to (a^2 + b^2)\sigma^2.$$

But the $Y_{n,i}$s, $H_n$, and $G_n$ are not fixed, but by a subsequence argument, it suffices to show
this last convergence holds in probability. Equivalently, it suffices to show

\[ R_n = n^{-1} \sum_{i=1}^{n} Y_{n,H_n(i)} Y_{n,G_n(i)} \to 0 \quad \text{in probability.} \]

But the conditional distribution of \( R_n \) given \( H_n \) is independent of \( H_n \). Therefore, it suffices to show

\[ P_n = n^{-1} \sum_{i=1}^{n} Y_{n,i} Y_{n,G_n(i)} \to 0 \quad \text{in probability.} \]

Let \( I_n \) be the number of \( i \)'s satisfying \( G_n(i) = i \). Note that \( I_n \) has expectation 1. Hence,

\[ \mathbb{E}(P_n) = \mathbb{E}\left( \mathbb{E}(P_n \mid I_n) \right) = n^{-1} \mathbb{E}\left( \sigma_{n,2} I_n \right) = n^{-1} \sigma_{n,2}^2 \to 0. \]

Also,

\[ n^2 \mathbb{E}(P_n^2) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[ Y_{n,i} Y_{n,j} Y_{n,G_n(i)} Y_{n,G_n(j)} \right]. \]

Summing over terms with \( i = j \) yields

\[ n \mathbb{E}\left[ Y_{n,i}^2 Y_{n,G_n(i)}^2 \right] \leq n \mathbb{E}(I_n) = n. \]

If \( i \) is different from \( j \), the only way that \( \mathbb{E}\left[ Y_{n,i} Y_{n,j} Y_{n,G_n(i)} Y_{n,G_n(j)} \right] \) can be nonzero is if \( G_n(i) = i \) or \( j \) and also \( G_n(j) \) is \( i \) or \( j \). Let \( J_n \) be the number of pairs \((i,j)\) with \( i \neq j \) satisfying this requirement. Then, \( n^2 \mathbb{E}(P_n^2) \leq n + \mathbb{E}(J_n) \). An easy calculation shows \( \mathbb{E}(J_n) = 2 \), so that \( \mathbb{E}(P_n^2) \to 0 \). The result follows.

**Proof of Proposition 3.2.** Conditions A–C are verified in Romano (1986). Indeed, the differentiability condition (2.1) on \( \tau \) holds with no error term at all by taking

\[ f(x,V,P) = \mu(V_1):1(x \in S_1 \times V_2), \]

where \( V = V_1 \times V_2 \) and \( \mu \) is the uniform probability measure on \( S_1 \). Let \( P_n \) be a sequence in \( \Omega_0 \) satisfying \( d_V(P_n,P_0) \to 0 \) for some \( P_0 \) in \( \Omega_0 \). Then, from Romano (1986), \( J_n(P_n) \) converges weakly to a continuous, strictly increasing limit law \( J(P_0) \), and \( J(P_0) \) is the distribution
of the norm of a certain Gaussian process \( L(P_0) \). We now verify Condition E when \( P^{(n)} = P_0^n \) and \( Q = P_0 \). Let \( S_n(\mathbf{x}_n, V) \) be given by (2.9). Here, \( \mathbf{x}_n = (X_{n,1}, \ldots, X_{n,n}) \) is a vector of \( n \) i.i.d. variables with distribution \( P_n \). Let \( G_n \) be independent of \( \mathbf{x}_n \) and uniformly distributed over \( G_n \). From Remark 2.4, we must show \( S_n(\mathbf{x}_n, V) \) and \( S_n(G_n, \mathbf{x}_n, \cdot) \) are asymptotically independent with law \( L(Q) \). As in example 1, it suffices to examine the finite dimensional distributions of these processes. Now, if \( V = V_1 \times V_2 \) and \( W = W_1 \times W_2 \), then

\[
\text{Cov} \left[ S_n(\mathbf{x}_n, V), S_n(G_n, \mathbf{x}_n, W) \right] =
\]

\[
\text{Cov} \left[ 1(X_{n,1} \in V) - \mu(V_1)1(X_{n,1} \in S_1 \times V_2), 1(G_{n,1}X_{n,1} \in W) - \mu(W_1)1(G_{n,1}X_{n,1} \in S_1 \times W_2) \right],
\]

where \( G_{n,1} \) is an independent uniform element from \( G_1 \). The independence of the events \( \{ X_{n,1} \in V_1 \times R \} \) and \( \{ G_{n,1}X_{n,1} \in W_1 \times R \} \) and the equivalency of the events \( \{ G_{n,1}X_{n,1} \in S_1 \times W_2 \} \) and \( \{ X_{n,1} \in S_1 \times W_2 \} \) shows that the above covariance is zero. The only thing that remains to show is that any finite linear combination of elements \( S_n(\mathbf{x}_n, V_i) \) with a finite linear combination of elements \( S_n(G_n, \mathbf{x}_n, W_i) \) is asymptotically Gaussian. But, \( G_n \mathbf{x}_n = \mathbf{y}_n \) is a vector \( (Y_{n,1}, \ldots, Y_{n,n}) \) of i.i.d. variables with \( X_{n,i} \) independent of \( Y_{n,i} \) if \( i \) is different from \( j \). Hence, the Lindeberg C.L.T. is directly applicable, yielding the result.

In the case \( Q_n \) is not in \( \Omega_0 \) with \( d_\mathcal{V}(Q_n, Q_\infty) \to 0 \), we need to verify Condition E when \( P^{(n)} = Q_n^n \) and \( Q = \tau Q_\infty \). The above method extends to this case. Alternatively, let \( P_n = \tau Q_n \) and \( P_0 = Q \) so that \( d_\mathcal{V}(P_n, P_0) \to 0 \). Observe that the behavior of the critical value from the randomization test (under \( Q_n^n \)) as obtained from \( J_n(\cdot, P_0^n | G_n, \mathbf{x}_n) \) only depends on the distribution of \( G_n, \mathbf{x}_n \). But \( G_n \mathbf{x}_n \) has the same distribution whether \( \mathbf{x}_n \) has distribution \( Q_n \) or \( P_n \). Thus, the above analysis is applicable, and Condition E is verified.

**Proof of Proposition 3.3.** Conditions A-C are verified in Romano (1986). The differentiability condition (2.1) for \( \tau \) holds with no error term at all by taking
\[ f(x, V, P) = \frac{1}{2} \left[ 1(x \in V) - 1(x \in V') \right]. \]

Let \( P_n \) be a sequence in \( \Omega_0 \) satisfying \( d_V(P_n, P_0) \to 0 \). Let \( S_n(x_n, \cdot) \) be defined by (2.9), where \( x_n = (X_{n,1}, \ldots, X_{n,n}) \) has distribution \( P_n^n \). Let \( G_n \) be an independent uniform transformation from \( G_n \). We must show \( S_n(x_n, \cdot) \) and \( S_n(G_n x_n, \cdot) \) are asymptotically independent.

Using the fact that \( G_n = (G_{n,1}, \ldots, G_{n,n}) \) is made up of \( n \) i.i.d. components shows that

\[
\text{Cov} \left[ S_n(x_n, V), S_n(G_n x_n, W) \right] = 1/2 \cdot \text{Cov} \left[ 1(X_{n,1} \in V) - 1(X_{n,1} \in V'), 1(G_{n,1} X_{n,1} \in W) - 1(G_{n,1} X_{n,1} \in W') \right].
\]

Since \( P_n \) is in \( \Omega_0 \), this must also be the same as the covariance between \( S_n(x_n, V') \) and \( S_n(G_n x_n, W) \), and hence is zero. Moreover, the fact that \( G_n \) is made up of i.i.d. components allows us to apply the C.L.T. so that

\[
\sum_{i=1}^{k} \left[ a_i S_n(x_n, V_i) + b_i S_n(G_n x_n, W_i) \right]
\]

is asymptotically Gaussian in law, and Condition E is verified. Finally, as in the proof of Proposition 3.2., if \( Q_n \) is a sequence not in \( \Omega_0 \) satisfying \( d_V(Q_n, Q_\infty) \to 0 \), apply the above to \( P_n = \tau Q_n \) and \( P_0 = \tau Q_\infty \).

**Proof of Proposition 3.4.** Let \( P_n \) be any sequence of probabilities on \( S \) satisfying \( d_V(P_n, P_0) \to 0 \) as \( n \to \infty \). From Romano (1986), \( J_n(P_n, \ldots, P_n) \) converges weakly to a fixed continuous limit law \( J(P_0, \ldots, P_0) \). Define a process \( S_n(x_n, \cdot): V \to \mathbb{R}^k \) by \( S_n(x_n, V) \) to be the vector with ith component

\[
S_{n,i}(x_n, V) = \frac{1}{n} \cdot c_{n,i} \left[ \hat{P}_n(V) - \hat{P}_{n,i}(V) \right],
\]

where \( x_n \) is an observation in \( S^n \), \( \hat{P}_n \) is the empirical of all \( n \) observations, and \( \hat{P}_{n,i} \) is the empirical of those \( n_i \) observations corresponding to the ith sample. Regard \( S_n \) as a random variable on \( L_\infty^k(V) \), the metric space of bounded \( \mathbb{R}^k \)-valued functions on \( V \) with metric \( \rho \) given
by \( \rho(D, S^*) = \max_{1 \leq j \leq k} \sup_{V \in \mathcal{V}} |S_j(V) - S_j^*(V)| \). Let \( L_n(P^{(n)}) \) be the distribution of \( S_n(x_n, \cdot) \) when \( x_n \) has distribution \( P^{(n)} \). From Romano (1986), \( L_n(P^{(n)}) \) converges in \( L_\infty^k(V) \) to a mean 0 Gaussian limit law \( L(P_0) \) (whose paths lie in a separable subset of \( L_\infty^k(V) \)). By the continuous mapping theorem, \( J_n(P_n, \ldots, P_k) \) converges weakly to a limit law \( J(P_0, \ldots, P_0) \) if \( d_V(P_n, P_0) \to 0 \). Let \( \hat{P}_n \) be the empirical measure of \( n \) observations from \( P_0 \). It follows by the generalized Glivenko-Cantelli theorem that \( d_V(\hat{P}_n, P_0) \to 0 \) a.s. as \( n \to \infty \). Thus,

\[
\sup_{t} |J_n(t, \hat{P}_n, \ldots, \hat{P}_n), J(P_0, \ldots, P_0)| \to 0 \quad \text{a.s.}
\]

Moreover, if \( b_n(\alpha, P_1, \ldots, P_k) \) represents an upper \( \alpha \) quantile of \( J_n(P_1, \ldots, P_k) \), then \( b_n(\alpha, \hat{P}_n, \ldots, \hat{P}_n) \) tends to a finite limit \( b(\alpha, P_0) \) in probability. Consistency of the test now follows easily, for suppose the actual distributions of the \( k \) populations are \( P_1, \ldots, P_k \). Clearly, the observed test statistic \( T_n \) tends to \( \infty \) in probability. On the other hand, the bootstrap critical value tends to a finite limit \( b(\alpha, P_0) \) in probability, where \( P_0 = \sum \lambda_i P_i \). This last statement follows from the fact that, in this case, if \( \hat{P}_n \) is the empirical of \( n \) observations under \( (P_1, \ldots, P_k) \), then by Glivenko-Cantelli, \( d_V(\hat{P}_n, P_0) \to 0 \) a.s.

To see that the randomization null distribution behaves in the same way as the bootstrap distribution, we must verify Condition D. Now, assume the conditions on the distribution \( Q^{(n)} \) of \( x_n \) as given in the statement of the proposition. Let \( G_n \) and \( G_n' \) be independent of \( x_n \) and i.i.d. uniform random transformations from \( G_n \). By analogy with Condition E, we must show the processes \( S_n(G_n x_n, \cdot) \) and \( S_n(G_n' x_n, \cdot) \) are asymptotically independent. As before, it suffices to examine the finite dimensional distributions of these processes. But, the linearity of these processes easily implies a limiting Gaussian distribution for linear combinations of elements \( S_n(G_n x_n, V_i) \) and \( S_n(G_n' x_n, W_i) \). Moreover, a covariance calculation (easily obtained by conditioning on \( G_n \) and \( G_n' \)) shows the covariance between \( S_n(G_n x_n, V) \) and \( S_n(G_n' x_n, W) \) to be of order \( O(n^{-1}) \), and the result follows.
Proof of Proposition 4.1. The proof is analogous to the proof of Theorem 1, p.828 in [17], which is the result in the well-known case of sampling with replacement. We assume, without loss of generality, that \( r_n = n \). Let \( X_1, \ldots, X_n \) be a sample of size \( n \) drawn with replacement from \( X_n \). Let \( Y_1, \ldots, Y_n \) be independent of the \( X_i \)'s and also drawn with replacement from \( X_n \). Let \( \hat{P}_n \) denote the empirical distribution of the \( X_i \)'s and let \( \hat{Q}_n \) denote the empirical distribution of the \( Y_i \)'s. By equation (10) of [17] (which only assumes the independence of \( \hat{P}_n \) and \( \hat{Q}_n \)), for \( \lambda > 0 \) and \( 0 < \theta < 1 \),

\[
Pr \left[ \sup_{V \in \mathcal{V}} \left| \hat{P}_n(V) - P_n(V) \right| > \lambda \right] \leq Pr \left[ \sup_{V \in \mathcal{V}} \left| \hat{P}_n(V) - \hat{Q}_n(V) \right| > (1-\theta)\lambda \right] = E \left[ Pr \left( \sup_{V \in \mathcal{V}} \left| \hat{P}_n(V) - \hat{Q}_n(V) \right| > (1-\theta)\lambda | F_n \right) \right],
\]

provided \( 4\theta^2 \lambda^2 n > 1 \). Here, \( F_n \) is the \( \sigma \)-field representing the information of which elements in the population \( X_n \) were drawn as an \( X_i \) or a \( Y_j \), but not whether a specific observation was an \( X_i \) instead of a \( Y_j \). Of course, the above conditional probability is clearly defined as \( F_n \) is induced by a finite partition. Suppose the elements in the population \( X_n \) are labeled \( \omega_1, \ldots, \omega_{n_\omega} \). Some of these may be the same outcome, but we can assume they are labeled so that we can distinguish all elements. Clearly, if the totality of all observations \( X_i \) and \( Y_j \) contains an \( \omega_k \) twice, then both the \( X_i \)'s and the \( Y_j \)'s must contain an \( \omega_k \). Let \( 2D_n \) be the number of duplicates observed in the joint sample, so that \( 2n - 2D_n \) is the number of distinct observations. For a fixed set \( V \), we bound

\[
Pr \left[ \left| \hat{P}_n(V) - \hat{Q}_n(V) \right| > (1-\theta)\lambda | F_n \right].
\]

For a fixed sequence of outcomes \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_n \), let \( R_n = n - D_n \). Since

\[
\hat{P}_n(V) - \hat{Q}_n(V) = \frac{1}{n} \left[ \frac{1}{n} \sum_{i=1}^{n} I(X_i \in V) - 1/2 \left[ \frac{1}{n} \sum_{i=1}^{n} I(X_i \in V) + \frac{1}{n} \sum_{i=1}^{n} I(Y_i \in V) \right] \right],
\]

those \( \omega_k \)'s appearing as both an \( X_i \) and a \( Y_j \) do not affect this difference. For \( i = 1, \ldots, R_n \), let \( T_i \) be an observation drawn at random without replacement from the \( 2R_n \) non-duplicate
observations. Then, conditional on a fixed element of the partition generated by $F_n$, $\hat{F}_n(V) - \hat{Q}_n(V)$ has the same distribution as

$$\frac{1}{n} \left[ \sum_{i=1}^{R_n} 1(T_i \in V) - \frac{1}{2} \sum_{i=1}^{2R_n} 1(T_i \in V) \right].$$

Of course, the sum $S = \sum_{i=1}^{2R_n} 1(T_i \in V)$ is (conditionally) fixed and not random. The distribution of this quantity is that of $1/n (W - EW)$, where $W$ has a hypergeometric distribution, because it is the distribution of the number of $R_n$ balls drawn marked 1 from a box of $2R_n$ balls with $S$ of them marked 1. Thus, by Corollary A.13.1 and A.4.9 of [17], and the fact that $R_n \leq n$, (5.2) is bounded above by $2\exp\left\{ -2n (1-\theta)^2 \lambda^2 \right\}$, and so (5.1) is bounded by this expression multiplied by $m_V(2n)$, where $m_V(\cdot)$ is the growth function for $V$. But, $m_V(2n) \leq 3/2 (2n)^s$ if $2n \geq s + 2$, where $s$ is a constant depending only on $V$ and not on $n$. Hence, (5.1) tends to zero for fixed $\lambda$ by choosing $\theta = 1/2$. 
REFERENCES


