ON THE BEHAVIOR OF RANDOMIZATION TESTS
WITHOUT A GROUP INVARIANCE ASSUMPTION

BY

JOSEPH P. ROMANO

TECHNICAL REPORT NO. 334
SEPTEMBER 1989

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS89-05874

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE BEHAVIOR OF RANDOMIZATION TESTS
WITHOUT A GROUP INVARIANCE ASSUMPTION

BY

JOSEPH P. ROMANO

TECHNICAL REPORT NO. 334
SEPTEMBER 1989

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS89-05874

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
On the Behavior of Randomization Tests
Without a Group Invariance Assumption

Joseph P. Romano
Department of Statistics, Stanford University
Stanford, CA 94305, USA

September, 1989

Abstract. Fisher's randomization construction of hypothesis tests is a powerful tool to yield tests which are nonparametric in nature in that their level is exactly equal to the nominal level in finite samples over a wide range of distributional assumptions. For example, the usual permutation t-test to test equality of means is valid without a normality assumption of the underlying populations. On the other hand, Fisher's randomization construction is not applicable in this example unless the underlying populations differ only in location. In general, the basis for the randomization construction is invariance of the probability distribution of the data under a transformation group. It is the goal of this paper to understand the robustness properties of randomization tests by studying their behavior in situations where the basis for their construction breaks down. In particular, we show that the randomization construction is generally consistent for certain one sample problems, such as for testing a mean or a median, even when the underlying population is not symmetric. In contrast, the randomization construction for two sample problems may yield inconsistent tests, though it depends on the precise nature of the problem. For example, the two sample permutation test based on sample means is generally consistent only if the samples are of the same size. However, when comparing medians, the two sample permutation test is generally inconsistent even if the sample sizes are equal.

Key Words: Consistency, Hypotheses Test, Permutation Test, Randomization Test, Robustness.
1. Introduction

Based on data $X$ taking values in a sample space $X$, it is desired to test the null hypothesis $H_0$ that the underlying probability law $P$ generating $X$ belongs to a certain family $\Omega_0$ of distributions. Let $G$ be a finite group of transformations $g$ of $X$ onto itself. Assume that the null hypothesis implies that the distribution of $X$ is invariant under the transformations in $G$; that is, for every $g$ in $G$, $gX$ and $X$ have the same distribution whenever $X$ has distribution $P$ in $\Omega_0$. For example, consider testing the equality of distributions based on two independent samples $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$. Under the null hypothesis that the samples are generated from the same probability law, the observations can be permuted or assigned at random to either of the two groups, and the distribution of the permuted samples is the same as the distribution of the original samples. In this example, and more generally when the invariance assumption holds, the usual construction of a randomization test applies.

To describe the general construction of the randomization test, we adopt the notation used by Hoeffding (1952). Let $T(X)$ be any test statistic for testing $H_0$. Suppose the group $G$ has $M$ elements. For every $x$ in $X$, let

$$T^{(1)}(x) \leq T^{(2)}(x) \leq \cdots \leq T^{(M)}(x)$$

be the ordered values of $T(gx)$ as $g$ varies in $G$. Given a nominal level $\alpha$, $0 < \alpha < 1$, let $k$ be defined by

$$k = M - [M\alpha],$$

(1.1)

where $[M\alpha]$ denotes the largest integer less than or equal to $M\alpha$. Let $M^+(x)$ and $M^0(x)$ be the number of values $T^{(j)}(x)$ ($j = 1, \ldots, M$) which are greater than $T^{(k)}(x)$ and equal to $T^{(k)}(x)$, respectively. Set

$$a(x) = \frac{M\alpha - M^+(x)}{M^0(x)}.$$  

By construction, $0 \leq a(x) < 1$.

Define the randomization test function $\phi(x)$ to be equal to 1, $a(x)$, or 0 according to whether $T(x) > T^{(k)}(x)$, $T(x) = T^{(k)}(x)$, or $T(x) < T^{(k)}(x)$, respectively. Note that, for every $x$ in $X$,

$$\sum_{g \in G} \phi(gx) = M^+(x) + a(x)M^0(x) = M\alpha.$$
Thus, if the distribution $P$ of $X$ is invariant under all $g$ in $G$, we have

$$M \alpha = E_P[\sum_g \phi(gX)] = \sum_g E_P[\phi(X)] = ME_P[\phi(X)].$$

Therefore, the exact finite sample level of the test $\phi$ is $\alpha$.

The idea of such a construction dates back to Fisher (1935), and then Pitman (1937/8). The beauty of this approach lies in its “robustness” properties. For example, consider testing the hypothesis that a population has mean 0, based on a sample of data $X = (X_1, \ldots, X_n)$. Let $T(X)$ be the usual $t$-statistic defined by

$$T(X) = \sum_i X_i / (\sum_i X_i^2)^{1/2}.$$

By assuming the underlying population is normal, one and two-sided tests may be constructed by using critical values based on the student $t$-distribution with $n$ degrees of freedom. If, on the other hand, it is only assumed that the underlying distribution of the $X_i$ is symmetric, then the method of randomization applies to yield an exact level $\alpha$ test. In particular, let $G$ be the group of transformations $g$ of the form

$$gx = ((-1)^{j_1}x_1, \ldots, (-1)^{j_n}x_n),$$

where $x = (x_1, \ldots, x_n)$ and $j_i = 0$ or 1, $i = 1, \ldots, n$. Evidently, the number $M$ of elements in $G$ is $2^n$. Since $gX$ and $X$ have the same distribution under the hypotheses that the $X_i$ have mean 0 and the $X_i$ have a symmetric distribution, the randomization test $\phi$ has exact level $\alpha$. The usual $t$-test, on the other hand, is asymptotically valid if the underlying population has a finite variance.

Lehmann and Stein (1949) have shown that, in several situations including the previous example, randomization tests enjoy certain finite sample optimality properties. In addition, Hoeffding (1952) has investigated the asymptotic power of randomization tests, and concludes that, for many problems, randomization tests are asymptotically as powerful as standard optimal procedures based on parametric models.

In this paper, the asymptotic behavior of certain randomization tests is studied when the data is not assumed to be generated from a distribution which is invariant under a group $G$. For example, reconsider the problem of testing the null hypothesis that the mean of a distribution on
the real line is 0. It is tempting to apply the randomization test as previously described. However, the construction and argument leading to justifying the validity of the test's level breaks down if the underlying distribution is not symmetric. In practice, the assumption of symmetry may be unrealistic, or at least questionable. Thus, it is desirable to understand the behavior of the randomization test without imposing symmetry. For another example, consider the application of the two-sample permutation t-test to test equality of means. Unless the two underlying populations differ only in location, the randomization construction is not applicable and the test may be invalid.

We now outline the balance of the paper. In Section 2, some consistency results for some one sample problems are presented. The results for the corresponding two sample problems are presented in Section 3. In short, the consistency or lack of consistency of randomization tests depends heavily on the particular problem and, in particular, the one sample case is quite distinct from the two sample case. For example, it is quite generally the case that the randomization construction is asymptotically valid for testing one sample location functionals, such as the mean or the median, even without a symmetry assumption. On the other hand, the randomization construction may be inconsistent for testing equality of parameters of two populations based on independent samples. Within the context of comparing two samples, however, the behavior of the randomization approach depends on the functional of interest. For example, if the two samples have equal sample sizes, the randomization approach for testing means is generally consistent, but is inconsistent if the ratio of sample sizes does not tend to one. On the other, the randomization approach is generally inconsistent for comparing medians, even when the sample sizes are identical. The proofs of these results are presented in Section 4. The arguments for the one sample problems are straightforward. In the two sample case, a general theorem assuming differentiability of the functional of interest is given. Much use is made of the behavior of the empirical process based on sampling without replacement.

It should be clear that the results generalize quite easily to k-sample problems and analysis of variance problems. Moreover, since rank tests are merely permutation tests once the data has been transformed to ranks, one can deduce results about rank procedures. For example, the median test (see Hogg and Tanis, p.391) is generally inconsistent for testing equality of medians unless it is assumed the two underlying populations are identical. Finally, by the usual inversion of tests to
form confidence intervals, the results have direct implications for confidence intervals as well.
2. One-sample Tests.

In this section, we consider the behavior of randomization tests for some one-sample problems. In anticipation of asymptotic results, the objects considered thus far, such as \(X, G, \phi, M,\) and \(k,\) will now have subscripts \(n\) to indicate dependence on the sample size \(n.\)

2.1. Testing for the mean.

The first problem considered is testing the null hypothesis that the mean \(\mu(F)\) of a distribution \(F\) is equal to \(\mu_0.\) Without loss of generality, we suppose \(\mu_0 = 0.\) Let \(G_n\) be the group of transformations \(g\) of the form (1.2). Given data \((X_1, \cdots, X_n)\) from \(F\) and a test statistic \(T_n = T_n(X_1, \cdots, X_n),\) let \(\hat{R}_n(x, G_n)\) denote the randomization distribution. That is, \(\hat{R}_n(x, G_n)\) is the conditional distribution function of \(T_n(\epsilon_1 X_1, \cdots, \epsilon_n X_n)\) given \((X_1, \cdots, X_n),\) where \(\epsilon_1, \cdots, \epsilon_n\) are independent (and independent of \((X_1, \cdots, X_n))\) such that \(P(\epsilon_1 = 1) = P(\epsilon_i = -1) = 1/2.\) To put it another way, \(\hat{R}_n(x, G_n)\) is the empirical distribution of the \(2^n\) values \(T_n(g(X_1, \cdots, X_n))\) as \(g\) varies in \(G_n.\)

Under the hypothesis that \(F\) is symmetric about 0, the randomization test \(\phi_n\) described in Section 1 has exact level equal to the nominal level \(\alpha.\) A nice feature of this construction is that the validity of the level of the test holds for any choice of test statistic \(T_n.\)

We now specialize to the case where \(T_n(X_1, \cdots, X_n) = n^{1/2} \bar{X}_n,\) where \(\bar{X}_n = \sum_i X_i/n.\) The introduction of \(n^{1/2}\) in the definition of \(T_n\) is convenient for asymptotic purposes and in no way affects the resulting outcome of the test. Let \(J_n(x, F)\) denote the actual (unconditional) distribution function of \(T_n(X_1, \cdots, X_n)\) under \(F.\) Recall \(k = k_n\) defined by (1.1) and let \(r_n = T^{(k_n)}(X_1, \cdots, X_n)\) denote the "critical value" of the randomization test \(\phi_n.\) Also, let \(\Phi(\cdot)\) denote the standard normal distribution function, and let \(z_{1-\alpha} = \Phi^{-1}(1 - \alpha).\) In the following theorem, the underlying population \(F\) is not assumed symmetric.

**Theorem 2.1.** Let \(X_1, \cdots, X_n\) be a sample from a population \(F\) with mean 0 and finite variance \(\sigma^2(F) > 0.\) Then, as \(n \to \infty,\)

\[
\sup_x |\hat{R}_n(x, G_n) - J_n(x, F)| \to 0 \quad (2.1)
\]

with probability one; moreover,

\[
\sup_x |\hat{R}_n(x, G_n) - \Phi(x/\sigma(F))| \to 0 \quad (2.2)
\]
with probability one. Thus,

$$ r_n \rightarrow \sigma(F)z_{1-\alpha} \quad (2.3) $$

with probability one, and

$$ E_F[\phi_n(X_1, \ldots, X_n)] \rightarrow \alpha. $$

The theorem implies that the randomization construction is valid in the sense that the
probability of a Type I error tends to the nominal level, even if the underlying population is not symmetric. Hence, one can claim that the exact finite sample level of the test is equal to the nominal level when the underlying population is symmetric and is asymptotically equal to the nominal level in general. Most methods, such as those relying on asymptotic expansions or even bootstrap re-
sampling, can not make any exactness claim.

An analogous result based on the test statistic \(|n^{1/2}\bar{X}_n|\) can readily be deduced from Theorem 2.1. Also, note that the randomization distribution based on the studentized test statistic

$$ \tilde{T}_n = n^{1/2}\bar{X}_n/\left( \sum_i X_i^2/n \right)^{1/2} $$

is asymptotically standard normal in the same sense as Theorem 2.1. This is also immediate from Theorem 2.1 because the denominator \((\sum_i X_i^2/n)^{1/2}\) is invariant under the transformations \(g \in G_n\), and the fact that \(\sum_i X_i^2/n \rightarrow \sigma^2(F)\) almost surely. In fact, the outcomes of the tests based on \(T_n\) and \(\tilde{T}_n\) are actually the same.

On the one hand, the validity of Theorem 2.1 follows readily from a Central Limit Theorem argument. On the other hand, it is important to note that the basis for the randomization argument completely breaks down when the underlying population is asymmetric. To be specific, suppose \(F\) is the distribution assigning equal mass to -3, -1 and 4. Then, \(F\) has mean 0 but is not symmetric. The basis for randomization is that, given the absolute value of an observation is \(z\), it is equally likely to be \(z\) or \(-z\). Here, however, given \(|X_i| = 1\), then \(X_i = -1\) with probability one.

Remark 2.1. Theorem 2.1 can readily be applied to yield consistency properties of the random-
ization test. Indeed, suppose \(G\) is an arbitrary alternative with mean \(\mu(G) \neq 0\) and finite variance \(\sigma^2(G)\). Then, the behavior of \(\hat{R}_n\) under \(G\) only depends on \(G\) through the distribution of \(|Y|\) when \(Y\) has distribution \(G\) because the construction of \(R_n\) depends on the sample \(X_1, \ldots, X_n\).
only through \(|X_1|, \ldots, |X_n|\). For simplicity, assume \(G\) is continuous and let \(F\) be the symmetric distribution obtained by

\[
F(x) = \frac{[G(x) + 1 - G(-x)]}{2}.
\]

Since \(|Y|\) has the same distribution when \(Y\) has distribution \(F\) or \(G\), the behavior of \(\hat{R}_n\) under \(G\) is the same as under \(F\). But since \(F\) is symmetric about 0, Theorem 2.1 is directly applicable, so that the critical value based on the randomization distribution of \(n^{1/2} \bar{X}_n\) satisfies

\[
r_n \to \sigma(F) z_{1-\alpha}
\]

with probability one. On the other hand, the actual value of the test statistic \(n^{1/2} \bar{X}_n\) tends to \(\infty\) with probability one if \(\mu(G) > 0\). Therefore, the power of the test against \(G\) must tend to one. One can also obtain consistency of the two-sided test based on \(n^{1/2}|\bar{X}_n|\) against a general alternative \(G\) with \(\mu(G) \neq 0\). In a similar fashion, one can obtain the asymptotic power of the test against (local) alternatives \(G_n\) which may depend on \(n\).

2.2. Testing for the median.

Consider the case where \(T_n(X_1, \ldots, X_n) = n^{1/2} \text{med}(X_1, \ldots, X_n)\), where \(\text{med}(X_1, \ldots, X_n)\) is a sample median of \(X_1, \ldots, X_n\). Under the hypothesis of symmetry about 0, the same group \(G_n\) applies as in section 2.1. Except for the definition of \(T_n\), retain the same notation as in Section 2.1 above. Thus, \(\hat{R}_n(x, G_n)\) is the randomization based on the sample median. The asymptotic behavior of \(\hat{R}_n(x, G_n)\) without imposing symmetry is given in Theorem 2.2.

**Theorem 2.2.** Let \(X_1, \ldots, X_n\) be a sample from a population \(F\) with median 0 and having a positive continuous density \(f\) in some neighborhood of 0. Further, assume the differentiability hypothesis

\[
F(x) = \frac{1}{2} + xf(0) + o(x)
\]

as \(x \to 0\). Then, as \(n \to \infty\),

\[
\sup_x |\hat{R}_n(x, G_n) - J_n(x, F)| \to 0 \tag{2.4}
\]

with probability one; moreover,

\[
\sup_x |\hat{R}_n(x, G_n) - \Phi(x/\tau(F))| \to 0 \tag{2.5}
\]
with probability one, where \( \tau^2(F) = [4f^2(0)]^{-1} \). Thus,

\[
  r_n \to \tau(F) z_{1-\alpha}
\]

with probability one, and

\[
  E_F[\phi_n(X_1, \cdots, X_n)] \to \alpha.
\]

Following the argument in Remark 2.1, one can obtain consistency of the randomization test based on the sample median.

2.3. General One-Sample Functionals

Based on a sample \( X_1, \cdots, X_n \) from a distribution \( F \), consider testing a general hypothesis about \( F \). As in subsections 2.1 and 2.2, consider the group \( G_n \) of sign changes of the observations. In order to understand the behavior of the randomization distribution \( \hat{R}_n(\cdot, G_n) \) of the test statistic, let

\[
  F_s(x) = \frac{[F(x) + 1 - F(-x)]}{2}
\]

be the “symmetrized” version of \( F \). By the argument presented in Remark 2.1, the behavior of \( \hat{R}_n \) under \( F \) is identical to that under \( F_s \).

Now, suppose \( \theta = \theta(F) \) is a statistical functional, and the null hypotheses specifies \( \theta = 0 \). Let \( T_n = n^{1/2} \theta(\hat{F}_n) \). Typically, \( T_n \) is asymptotically normal with mean 0 and variance \( \sigma^2(F) \). Based on verification of the conditions in Hoeffding’s (1959) Theorem 3.2, it is typically the case that the randomization distribution converges with probability one to the same limit law as the unconditional sampling distribution, as long as the hypothesis testing problem remains invariant under \( G_n \). Since \( F_s \) is symmetric about 0, the behavior of \( R_n \) under \( F_s \) should then be asymptotically Gaussian with mean 0 and variance \( \sigma^2(F_s) \). It follows that the behavior of \( \hat{R}_n \) under \( F \) is that \( \hat{R}_n \) converges in distribution to a Gaussian distribution with mean 0 and variance \( \sigma^2(F) \) with probability one. Hence, consistency of the randomization distribution follows when \( \sigma^2(F) = \sigma^2(F_s) \). For general location functionals, this is typically the case. For example, if \( \theta(F) \) is an M-estimator defined as the solution \( \theta \) of

\[
  \int \psi(x - \theta)dF(x) = 0,
\]

9
for some \( \psi \), then

\[
\sigma^2(F) = \int \psi^2(x) dF(x) / [ \int \psi(x) dF(x) ]^2.
\]

Since \( \psi \) is typically an odd function, it follows easily that \( \sigma^2(F) = \sigma^2(F_s) \). Therefore, general results analogous to Theorems 2.1 and 2.2 are expected to hold for general location functionals.
3. Two-sample Tests.

In this section, we discuss the behavior of some permutation tests associated with two samples. To begin, assume $X_1, \cdots, X_n$ is a sample of $n$ independent observations from a distribution $F_X$ and $Y_1, \cdots, Y_m$ is a sample of $m$ observations from $F_Y$. Here, $X = (X_1, \cdots, X_n, Y_1, \cdots, Y_m)$. Under the hypothesis $F_X = F_Y$, we may permute the coordinates of $X$ to form a new sample having the same joint distribution as the original sample $X$. More specifically, let $N = n + m$, and for $x = (x_1, \cdots, x_N) \in \mathbb{R}^N$, let $gx \in \mathbb{R}^N$ be defined by $(x_{\pi(1)}, \cdots, x_{\pi(N)})$, where $(\pi(1), \cdots, \pi(N))$ is a permutation of $(1, 2, \cdots, N)$. Let $G_N$ be the collection of all such $g$ so that $M = N!$. Under the hypothesis $F_X = F_Y$, $gX$ and $X$ have the same distribution for any $g$ in $G_N$. Throughout this section, the group $G_N$ is as just described.

3.1 Testing equality of means.

Suppose that $F_X$ and $F_Y$ are two distributions on the real line and it is desired to test the null hypothesis that $\mu(F_X) = \mu(F_Y)$, where $\mu(F)$ is the mean of $F$. Under the further hypothesis that $F_X = F_Y$, exact randomization tests may be constructed by applying the general construction discussed in Section 1 to the group $G_N$. However, we do not wish to make this assumption.

Consider the test statistic

$$T_{n,m}(X) = T_{n,m}(X_1, \cdots, X_n, Y_1, \cdots, Y_m) = n^{1/2}(\bar{X}_n - \bar{Y}_m),$$

(3.1)

where $\bar{X}_n = \sum X_i/n$ and $\bar{Y}_m = \sum Y_j/m$. As before, let $\hat{R}_N(\cdot, G_N)$ denote the randomization distribution; that is, $\hat{R}_n(\cdot, G_N)$ is the distribution of $T_{n,m}(gX)$ when $X = (X_1, \cdots, X_n, Y_1, \cdots, Y_m)$ is fixed and $g$ is picked at random from $G_N$. The behavior of the randomization distribution is given in the following theorem. In the theorem, $\sigma^2(F)$ denotes the variance of $F$.

**Theorem 3.1.** Suppose $F_X$ and $F_Y$ have common mean $\mu$ and finite variances $\sigma^2(F_X)$ and $\sigma^2(F_Y)$, respectively. Assume $m/N \to \lambda$ as $n \to \infty$, where $\lambda \in (0, 1)$. Let $\sigma_p^2$ denote the variance of $\lambda F_Y + (1 - \lambda) F_X$, so

$$\sigma_p^2 = \sigma^2(\lambda F_Y + (1 - \lambda) F_X) = \lambda \sigma^2(F_Y) + (1 - \lambda) \sigma^2(F_X).$$

Then,

$$\sup_x |\hat{R}_n(x, G_N) - \Phi(x \sigma_p / \lambda^{1/2})| \to 0$$

11
with probability one, and so the critical value \( r_n \) of the randomization distribution satisfies

\[
  r_n \to \sigma_{1-a}/\lambda^{1/2}
\]

with probability one.

Under the conditions of Theorem 3.1, the actual (unconditional) sampling distribution of the test statistic (3.1) is asymptotically Gaussian with mean 0 and variance

\[
\sigma^2(F_X) + (1 - \lambda)\sigma^2(F_Y)/\lambda.
\]

According to the theorem, however, the randomization distribution is asymptotically Gaussian with mean 0 and variance

\[
\sigma^2(F_Y) + (1 - \lambda)\sigma^2(F_X)/\lambda.
\]

These Gaussian distributions are identical if and only if \( \sigma^2(F_X) = \sigma^2(F_Y) \) or \( \lambda = 1/2 \). When (and only when) these distributions are identical, consistency of the randomization follows in the sense that the probability of a Type I error is asymptotically equal to the nominal level. Hence, unlike the one-sample test based on the sample mean, the two sample test for differences of means based on differences in sample means is asymptotically valid only if the distributions have the same variance or the sample sizes are the same, or at least approximately the same. To summarize, we have the following.

**Corollary 3.1.** Under the conditions of Theorem 3.1, the randomization test is consistent if and only if \( \lambda = 1 \) or \( \sigma^2(F_X) = \sigma^2(F_Y) \).

### 3.2. Testing equality of medians.

Given \( X \) and the same group \( G_N \), now consider testing the hypothesis that the distributions \( F_X \) and \( F_Y \) have identical medians. Under the further hypotheses \( F_X = F_Y \), exact permutation tests may be constructed for this problem, but we wish to understand the behavior of such tests without this further assumption.

Let \( \theta(F) \) denote the median of \( F \); formally, \( \theta(F) = \inf\{x : F(x) \geq 1/2\} \). Consider the test statistic

\[
T_n(X) = n^{1/2}[\theta(\hat{F}_n) - \theta(\hat{G}_m)],
\]  

(3.2)
where $\hat{F}_n$ is the empirical distribution of $X_1, \ldots, X_n$ and $\hat{G}_m$ is the empirical distribution of $Y_1, \ldots, Y_m$. Again, let $\hat{R}_N(\cdot, G_N)$ denote the randomization distribution of (3.2).

**Theorem 3.2.** Suppose $F_X$ and $F_Y$ have common median $\theta$. Let $f_X$ and $f_Y$ denote the densities of $F_X$ and $F_Y$, respectively, assumed to exist and be positive and continuous in a neighborhood of $\theta$. Assume $m/N \to \lambda$ as $n \to \infty$, where $\lambda \in (0, 1)$. Let

$$\tau^2 = (\lambda f_Y(\theta) + (1 - \lambda)f_X(\theta))^{-2}/4.$$  \hfill (3.3)

Then,

$$\sup_x |\hat{R}_n(x, G_N) - \Phi(x\tau/\lambda^{1/2})| \to 0$$

with probability one, and so the critical value $r_n$ satisfies

$$r_n \to \tau z_{1-\alpha}/\lambda^{1/2}$$

with probability one.

Under the hypotheses of Theorem 3.2, the actual (unconditional) null sampling distribution of (3.2) is asymptotically Gaussian with mean 0 and variance

$$\frac{1}{4f_X^2(\theta)} + \frac{1 - \lambda}{\lambda} \cdot \frac{1}{4f_Y^2(\theta)}.$$  \hfill (3.4)

Hence, consistency occurs when these asymptotic variances (3.4) and $\tau^2/\lambda$ are identical. Here, we see a marked difference between the two sample test based on means and the two sample test based on medians. For suppose $n = m$, so $\lambda = 1/2$. Then, $2\tau^2$ and (3.4) are identical if and only if $f_X(\theta) = f_Y(\theta)$. In the nonparametric context of not assuming $F_X = F_Y$, this assumption is strong indeed, and it is concluded that the permutation test is generally inconsistent. For various choices of $f_X$ and $f_Y$, the asymptotic variances (3.4) and $2\tau^2$ can be markedly different, so that the probability of a Type I error can tend asymptotically to a value that is far from the nominal level.
3.3. General Two Sample Problem.

Now consider the more general situation of testing $\theta(F_X) = \theta(F_Y)$, where $\theta(\cdot)$ is some functional on some subset $F$ of distributions on the line. (The generalization to distributions on arbitrary spaces is easily handled by the method of proof used here.) The goal of this section is to establish the asymptotic behavior of the randomization distribution based on the test statistic

$$T_n(X) = n^{1/2}[\theta(\hat{F}_n) - \theta(\hat{G}_m)]. \quad (3.5)$$

It will be assumed that the functional $\theta(\cdot)$ is appropriately differentiable, so that Theorems 3.1 and 3.2 will be immediate consequences of this general approach. This means we will need a rather weak form of differentiability so that implications for the median can be made.

To get started, let $D[-\infty, \infty]$ denote the metric space of right continuous functions on $[-\infty, \infty]$ having left limits, with metric generated by the uniform norm $\| \cdot \|$; that is, for $x \in D[-\infty, \infty]$, 

$$\|x\| = \sup \{ x(t) : t \in [-\infty, \infty] \}. \quad \text{Specifically, we will suppose } \theta(\cdot) \text{ is compactly differentiable at } F \text{ tangentially to the subspace } H_F \text{ consisting of functions in } D[-\infty, \infty] \text{ which are continuous at all points except possibly where } F \text{ has atoms. That is, assume that for some (influence) function } g = g_F, \text{ any sequence of distributions } F_n \text{ in } F \text{ satisfying } \|F_n - F\| \to 0, \text{ any sequence } D_n \text{ in } D[-\infty, \infty] \text{ satisfying } \|D_n - D_0\| \to 0 \text{ for some } D_0 \text{ in } H_F, \text{ and any sequence of real numbers } \epsilon_n \to 0 \text{ so that } G_n = F_n + \epsilon_n D_n \in F, \text{ we have}$$

$$\theta(G_n) - \theta(F_n) - \int g_F d(G_n - F_n) = O(\epsilon_n)$$

as $n \to \infty$. The influence curve $g_F$ is normalized so $\int g_F dF = 0$. Let $\sigma^2(F) = \int g_F^2 dF$. This form of differentiability, as initially developed by Gill (1988), is rather weak in that the differential approximation is only required to hold for $D_0$ in $H_F$ and not necessarily for all $D_0$ in $D[-\infty, \infty]$. Hence, it is a weaker form of compact or Hadamard differentiability. More abstract versions are presented in Sheehy and Wellner (1988). The reader unfamiliar with this notion of differentiability may assume that the differentiability assumption ensures a linear approximation to the statistic of interest.

Under the differentiability assumption, it follows that the law of $T_n(X)$ converges weakly to the Gaussian distribution with mean 0 and variance

$$\sigma^2(F_X) + \frac{1-\lambda}{\lambda} \sigma^2(F_Y), \quad (3.6)$$
where $\lambda$ is such that $m/(m + n) \to \lambda$. The main result of this section is to establish that the randomization distribution $\hat{R}_n(\cdot, G_N)$ behaves asymptotically Gaussian with mean 0 and variance

$$
\sigma^2(\lambda F_Y + (1 - \lambda)F_X) + \frac{(1 - \lambda)}{\lambda} \sigma^2(\lambda F_Y + (1 - \lambda)F_X).
$$

(3.7)

In words, the randomization distribution behaves asymptotically like the unconditional sampling distribution of $T_n$, when all the observations are i.i.d. $\lambda F_Y + (1 - \lambda)F_X$. Specifically, we have the following.

**Theorem 3.3.** Assume $\theta(\cdot)$ is compactly differentiable as described. Let $m/(m + n) \to \lambda$, where $\lambda \in (0, 1)$. As before, let $\hat{R}_n(\cdot, G_N)$ denote the randomization distribution of (3.5). Then, under the null hypothesis $\theta(F_X) = \theta(F_Y)$, we have

$$
sup_x |\hat{R}_n(x, G_N) - \Phi(x \tau / \lambda^{1/2})| \to 0
$$

with probability one, where $\tau^2$ is equal to $\lambda$ times the expression (3.7). Consequently, the critical value $r_n$ satisfies

$$
r_n \to \tau z_{1 - \alpha} / \lambda^{1/2}
$$

with probability one.

The consistency of the randomization test does not hold in general by Theorem 3.3. The theorem merely reduces the problem to comparing (3.6) and (3.7), and consistency follows when these asymptotic variances are the same. As seen in subsections (3.1) and (3.2), the form of $\sigma^2(\cdot)$ may lead to consistency or inconsistency, depending on the functional $\theta(\cdot)$, and possibly on $\lambda$. For example, one may easily conclude from Theorem 3.3 that the permutation test based on differences in trimmed means is generally inconsistent, in the equal sample size case, unless the trimming proportion is zero.

Theorem 3.3 can easily be generalized to study the behavior of the randomization test under test alternatives.

It is perhaps worthwhile to compare the randomization approach to the bootstrap approach of Efron; see Efron and Tibshirani (1986), for example. In the bootstrap case, observations are resampled with replacement from the combined sample to form pseudosamples, from which an
approximation to the sampling distribution of $T_n$ is constructed. In general, the bootstrap distribution is also approximately normal with mean 0 and variance given by (3.7). In contrast, the randomization distribution is constructed by resampling the data without replacement, calling the first $n$ observations $X_i's$ and the remaining $m$ observations $Y_j's$. Both methods are consistent under the more restrictive hypothesis $F_X = F_Y$, but not in general. The asymptotics for the randomization approach is complicated by the fact that the resampled observations are not conditionally independent (given the data), as they are in the bootstrap resamples. Asymptotics for the bootstrap follows easily as in Beran (1984). The bootstrap and randomization methods are compared in other contexts in Romano (1989).
4. Proofs.

In order to prove Theorem 2.1, we need two lemmas.

**Lemma 4.1.** Let $y_1, y_2, \cdots$ be a sequence of numbers satisfying $\sum_{i=1}^{n} y_i^2 / n \to \sigma^2$, for some $\sigma^2 > 0$. Let $\epsilon_1, \epsilon_2, \cdots$ be independent and identically distributed such that

$$P(\epsilon_i = 1) = P(\epsilon_i = -1) = 1/2.$$

Let

$$S_n = n^{-1/2} \sum_{i=1}^{n} y_i \epsilon_i.$$

Then, $S_n$ converges in distribution to the normal distribution with mean 0 and variance $\sigma^2$.

**Proof of Lemma 4.1.** By Slutsky's Theorem, it suffices to show $S_n/(\sum_{i=1}^{n} y_i^2 / n)^{1/2}$ converges in distribution to the standard normal distribution. To prove this, the Lindeberg condition must be verified. This requires showing: for each fixed $\delta > 0$,

$$\lim_{n \to \infty} \sum_{i=1}^{n} E[Z_{n,i}^2, 1(|Z_{n,i}| > \delta)] = 0,$$

where

$$Z_{n,i} = y_i \epsilon_i / (\sum_{i=1}^{n} y_i^2)^{1/2}.$$

Let $m_n = \text{Max}_{1 \leq i \leq n}|y_i|$. Then,

$$\sum_{i=1}^{n} E[Z_{n,i}^2, 1(|Z_{n,i}| > \delta)] \leq \left(\sum_{i=1}^{n} y_i^2\right)^{-1} \sum_{i=1}^{n} y_i^2 E[\epsilon_i^2, 1(|\epsilon_i| > \frac{\left(\sum_{i=1}^{n} y_i^2\right)^{1/2} \delta}{m_n} \}]

= E[\epsilon_1^2 \left(\epsilon_1^2 > \frac{\delta^2 \sum_{i=1}^{n} y_i^2}{m_n^2}\right)] \leq 1(1 > \delta^2 \sum_{i=1}^{n} y_i^2 / m_n^2)$.$$

Hence, it suffices to show $\sum_{i=1}^{n} y_i^2 / m_n^2 \to \infty$, which follows if it can be shown that $m_n^2 / n \to 0$ as $n \to \infty$. But, this latter fact follows from Lemma 4.2 below.

**Lemma 4.2.** Suppose $y_1, y_2, \cdots$ is a sequence of numbers such that $\sum_{i=1}^{n} y_i^2 / n \to \sigma^2 > 0$. Let $m_n = \text{Max}_{1 \leq i \leq n}|y_i|$. Then, $m_n^2 / n \to 0$ as $n \to \infty$.

**Proof of Lemma 4.2.** Without loss of generality, assume $\sigma = 1$. Suppose the result were false. Then, there exists $\epsilon > 0$ such that, for every $N$, there exists a $k > N$ satisfying $m_k^2 / k > \epsilon$ and
If this latter condition failed, the sequence \( y_n^2 = y_n^2 \); if \( m_k^2 = y_k^2 \) would be bounded and the lemma

Now, given this \( \epsilon \), choose any positive \( \delta < \epsilon / 3 \). Since \( \sum_{i=1}^{\infty} \frac{y_i^2}{n} \to 1 \), there exists \( M = M(\delta) \) such that for all \( m \geq M \),

\[
1 - \delta \leq \sum_{i=1}^{m} \frac{y_i^2}{m} \leq 1 + \delta.
\]

Now, given this \( \epsilon, \delta, \) and \( M \), take \( N \) large enough so that \( N > M \) and \( (N - 1)/N > 1 - \delta \). Then, there exists \( k > N \) such that \( m_k^2 / k > \epsilon \) and \( m_k^2 = y_k^2 \). Hence, for this value of \( k \),

\[
\sum_{i=1}^{k} \frac{y_i^2}{k} = \frac{k-1}{k} \sum_{i+1}^{k-1} \frac{y_i^2}{(k-1)} + m_k^2 / k \geq \frac{k-1}{k} (1 - \delta) + m_k^2 / k \geq (1 - \delta) + \epsilon.
\]

Since \( \delta < \epsilon / 3 \),

\[
(1 - \delta)^2 + \epsilon = 1 - 3\delta + \delta^2 + \epsilon + \delta \geq 1 + \delta + \delta^2 > 1 + \delta,
\]

a contradiction as \( k > M \).

**Proof of Theorem 2.1.** Treating the \( X_i \) as fixed, \( \hat{R}_n(x, G_n) \) is the distribution function of \( n^{-1/2} \sum_{i=1}^{n} X_i \epsilon_i \), where the \( \epsilon_i \) are independent and identically distributed, each \( \epsilon_i \) taking on the values 1 and -1 with probability 1/2 each. By the strong law of large numbers, \( \sum_{i=1}^{n} X_i^2 / n \to \sigma^2(F) \) with probability one. Hence, by Lemma 4.1, on the set where \( \sum_{i=1}^{n} X_i^2 / n \to \sigma^2(F) \), \( \hat{R}_n(x, G_n) \to \Phi(x / \sigma(F)) \). Hence, for any fixed \( x \), \( \hat{R}_n(x, G_n) \to \Phi(x / \sigma(F)) \) with probability one. Let \( Q \) be any countable dense subset of the real line. Then,

\[
\sup_{x \in \mathbb{Q}} |\hat{R}_n(x, G_n) - \Phi(x / \sigma(F))| \to 0
\]

with probability one. By Corollary 1, page 259 of Chow and Teicher (1978), \( \hat{R}_n(\cdot, G_n) \) converges weakly to \( \Phi(\cdot / \sigma(F)) \), for almost all sample sequences \( X_1, X_1, \ldots \). Because this limiting distribution is continuous, it follows by Lemma 3, page 260 of Chow and Teicher (1978), that the convergence is uniform in \( x \), so that (2.2) holds. By (2.2) and the fact that \( J_n(x, F) \to \Phi(x / \sigma(F)) \) uniformly in \( x \), (2.1) follows. To prove (2.3), note that whenever \( G_n(\cdot) \) is a sequence of distribution functions converging uniformly to a continuous, strictly increasing distribution function \( G(\cdot) \), then \( G_n^{-1}(1 - \alpha) \to G^{-1}(1 - \alpha) \); here \( G_n^{-1}(1 - \alpha) \) can be any number between \( \sup\{x : G_n(x) \leq 1 - \alpha\} \) and \( \inf\{x : G_n(x) \geq 1 - \alpha\} \). Since \( r_n \) is a \( 1 - \alpha \) quantile of \( \hat{R}_n(\cdot, G_n) \), the result follows.
**Lemma 4.3.** Suppose $x_1, x_2, \ldots$ is a fixed sequence of numbers. Let $R_n(x)$ be the distribution function of

$$T_n = n^{1/2} \text{med}(\sigma_1 x_1, \ldots, \sigma_n x_n),$$

where the $\sigma_i$ are independent and identically distributed so that $P(\sigma_i = 1) = P(\sigma_i = -1) = 1/2$. For fixed $x$, let $A_n = A_n(x)$ be the number of $x_i, 1 \leq i \leq n$, such that $|x_i| > n^{-1/2} x$. Assume

$$A_n/n = 1 - n^{-1/2} 2xC + o(n^{-1/2})$$

as $n \to \infty$, where $C$ is some positive constant. Then,

$$R_n(x) \to \Phi(2xC).$$

**Proof of Lemma 4.3.** Fix $x > 0$ satisfying (4.1). Let $Z_i = \sigma_i x_i$. Assume for simplicity that $n = 2N + 1$ tends to infinity through odd values; the case $n = 2N$ is similar. In order to calculate the probability that $T_n \leq x$, note that if $x_i$ is one of the $n - A_n$ observations satisfying $|x_i| \leq n^{-1/2} x$, then regardless of $\sigma_i$, $Z_i \leq n^{-1/2} x$. Any of the remaining $A_n$ observations are less than or equal to $n^{-1/2} x$ with probability $1/2$. Since the event $T_n \leq x$ is equivalent to at least $N + 1$ of the $Z_i$ observations less than or equal to $n^{-1/2} x$, the probability that $T_n \leq x$ becomes

$$P\{\text{at least } N + 1 - (n - A_n) \text{ of the } Z_i, i \in S_n(x) \leq x\},$$

where $S_n(x)$ is the set of indices $i, 1 \leq i \leq n$, satisfying $|x_i| > n^{-1/2} x$. Let $B_n = B_n(x)$ be the number of $Z_i, i \in S_n(x)$ that are less than or equal to $n^{-1/2} x$; $B_n$ has a binomial distribution based on $A_n$ trials and success probability $1/2$. Then, (4.2) becomes

$$P\{B_n \geq N + 1 - (n - A_n)\}.$$  \hspace{1cm} (4.3)

By the normal approximation to the binomial, rigorously verified by the Lindeberg Central Limit Theorem, the difference between (4.3) and

$$1 - \Phi\left(\frac{N + 1 - (n - A_n) - A_n/2}{A_n^{1/2}/2}\right) = \Phi\left(\frac{2N - A_n}{A_n^{1/2}}\right)$$

(4.4)

tends to zero. By assumption (4.1), $(2N - A_n)/A_n^{1/2} \to 2xC$ and the result follows. The case $x \leq 0$ is similar.
Proof of Theorem 2.2. \( \hat{R}_n(\cdot, G_n) \) is the (random) distribution of \( n^{1/2} \text{med}(\sigma_1 X_1, \ldots, \sigma_n X_n) \) with the \( X_i \) treated as fixed. Let \( A_n = A_n(x) \) be the number of \( X_i, 1 \leq i \leq n \), such that \( |X_i| > n^{-1/2} x \). Then, \( A_n \) is binomially distributed based on \( n \) trials and success probability

\[
p_n = 1 - [F(n^{-1/2} x) - F(-n^{-1/2})].
\]

(Note that \( F \) may not be continuous outside a neighborhood of 0, in which case \( p_n \) needs slight modification; however, \( p_n \) as defined holds for all large \( n \) anyway.) In fact,

\[
A_n = n - n[\hat{F}_n(n^{-1/2} x) - \hat{F}_n(-n^{-1/2} x)],
\]

where \( \hat{F}_n \) is the empirical distribution of \( X_1, \ldots, X_n \). We wish to show

\[
A_n/n = 1 - n^{-1/2} 2x f(0) + o(n^{-1/2})
\]

(4.5)

for almost all sample sequences \( X_1, X_2, \ldots \), so that condition (4.1) holds with probability one. By Theorem 14.3.1 of Shorack and Wellner (1986), due to Csorgo and Revesz,

\[
\limsup_{n \to \infty} |F(-n^{-1/2} x) - F(n^{-1/2} x) - \frac{A_n}{n} + 1| \leq K n^{-3/4} [\log(n)]^{1/2}
\]

(4.6)

with probability one, where \( K \) is some constant. Also, by the differentiability hypothesis,

\[
F(-n^{-1/2} x) - F(n^{-1/2} x) = -2n^{-1/2} x f(0) + o(n^{-1/2}).
\]

(4.7)

Combining (4.6) and (4.7) yields

\[
\frac{A_n}{n} - 1 = [-2n^{-1/2} x f(0) + o(n^{-1/2})] + O(n^{-3/4} [\log(n)]^{1/2})
\]

for almost all sample sequences. Thus, Lemma 4.1 is applicable to all such sample sequences. The conclusion is

\[
\hat{R}_n(x, G_n) \to \Phi(2f(0)x)
\]

with probability one. From the pointwise result (fixed \( x \)), one can deduce

\[
\sup_x |\hat{R}_n(x, G_n) - \Phi(2f(0)x)| \to 0
\]

20
with probability one, by exploiting the fact the limiting normal distribution is continuous as done in the proof of Theorem 2.1. The remainder of the proof is also analogous to the proof of Theorem 2.1.

To prove the results in Section 3, the following lemma is needed. Some weak convergence notation and terminology is needed. As in Section 3, let \( D[-\infty, \infty] \) denote the metric space of right continuous functions on \([-\infty, \infty]\) having left limits, with metric generated by the uniform norm \( \| \cdot \| \); that is, for \( x \in D[-\infty, \infty] \), \( \| x \| = \sup \{ x(t) : t \in [-\infty, \infty] \} \). Endow \( D[-\infty, \infty] \) with the projection \( \sigma \)-field. Convergence in \( D[-\infty, \infty] \) is understood in the sense of Pollard (1984), Chapter IV; that is, random elements \( Z_n \) taking values in \( D[-\infty, \infty] \) converges in distribution to a random element \( Z \) if \( Ef(Z_n) \to Ef(Z) \) for all real-valued, continuous, measurable functions \( f \) having domain \( D[-\infty, \infty] \).

**Lemma 4.4.** Let \( z_1, \cdots, z_N \) be a sequence of numbers satisfying

\[
\sup_t \{|F_N(t) - F_0(t)| \to 0
\]

for some distribution function \( F_0 \), where

\[
F_N(t) = N^{-1} \sum_{i=1}^N 1(z_i \leq t).
\]

Let \( z_1^*, \cdots, z_n^* \) be a sample of size \( n \) \((n < N)\) chosen without replacement from \( z_1, \cdots, z_N \). Let \( F_n^* \) be the empirical distribution of \( z_1^*, \cdots, z_n^* \), and let \( G_m^* \) be the empirical distribution of the remaining \( m = N - n \) of the \( z_i \)'s. Define

\[
Z_n(t) = n^{1/2} [F_n^*(t) - G_m^*(t)].
\]

Regard \( Z_n \) as a random element of \( D[-\infty, \infty] \). Assume \( n \) and \( N = N(n) \) tends to \( \infty \) so that \( m/n \to \lambda \), where \( \lambda \in (0, 1) \).

(i). Then, \( Z_n \) converges in distribution to a Gaussian process \( Z \) having mean 0 and covariance function

\[
E[Z(s)Z(t)] = F_0(s)[1 - F_0(t)]/\lambda,
\]

for \( s \leq t \).
(ii). Assume $\theta(\cdot)$ is a statistical functional which is compactly differentiable at $F_0$ tangentially to $H_{F_0}$, with influence function $g_{F_0}$. The influence function is normalized so $\int g_{F_0} dF_0 = 0$, and let $\sigma^2(F_0) = \int g_{F_0}^2 dF_0$. Then, $n^{1/2}[\theta(F_n^*) - \theta(G_m^*)]$ converges in distribution to the normal distribution with mean 0 and variance $\sigma^2(F_0)/\lambda$.

**Proof.** To prove (i), let $\tilde{Z}_n(t) = n^{1/2}[\tilde{F}_n(t) - F_N(t)]$. Regard $\tilde{Z}_n$ as a random element of $D[-\infty, \infty]$. By Theorem 1 of Bickel and Krieger (1988), $\tilde{Z}_n$ converges in distribution to a mean 0 Gaussian process $\tilde{Z}$ having covariance function

$$E[\tilde{Z}(s)\tilde{Z}(t)] = \lambda F_0(s)[1 - F_0(t)]$$

for $s \leq t$. But, it is trivial to see that $Z_n(t) = Nm^{-1} \tilde{Z}_n(t)$. Since $Nm^{-1} \to \lambda^{-1}$, the result follows.

To prove (ii), by Theorem 3 of Gill (1988) and part (i), it follows that $n^{1/2}[\theta(F_n^*) - \theta(G_m^*)]$ converges in distribution to the distribution of $\int g_{F_0} dZ$, where $Z$ is the limiting process in (i). In order to explicitly identify the distribution of $\int g_{F_0} dZ$, note that the distribution of $Z$ in $D[-\infty, \infty]$ is identical to the limiting process of

$$W_n(t) = n^{1/2}[\tilde{F}_n(t) - \tilde{G}_m(t)],$$

where $\tilde{F}_n$ is the empirical distribution of a sample of $n$ independent observations chosen from $F_0$ and $\tilde{G}_m$ is the empirical distribution of a sample of $m$ independent observations chosen from $F_0$. It follows that the distribution of $\int g_{F_0} dZ$ is identical to the limiting distribution of $n^{1/2}[\theta(\tilde{F}_n) - \theta(\tilde{G}_m)]$, which by differentiability and the Central Limit Theorem is normal with mean 0 and variance $\sigma^2(F_0)/\lambda$.

Lemma 4.1 will be applied in the following setup. Let $X_1, \ldots, X_n$ be a random sample from $F_X$ and $Y_1, \ldots, Y_m$ an independent random sample from $F_Y$, where $F_X$ and $F_Y$ are distributions on the line. The behavior of (3.5) can be deduced by an appropriate linearization of $\theta(\cdot)$ once the behavior of the two sample process is understood, where the two sample process is defined by

$$S_n(t) = n^{1/2}[\tilde{F}_n(t) - \tilde{G}_m(t)], \quad (4.8)$$

where $\tilde{F}_n$ is the empirical distribution of $X_1, \ldots, X_n$ and $\tilde{G}_m$ is the empirical distribution of $Y_1, \ldots, Y_m$. To understand the behavior of the permutation distribution of $T_n$, we first understand the permutation distribution of $S_n$, regarded as a random element of $D[-\infty, \infty]$. To describe
this permutation distribution, conditional on $X = (X_1, \cdots, X_n, Y_1, \cdots, Y_m)$, let $Z_1^*, \ldots, Z_n^*$ be a sample of size $n$ chosen without replacement from the $N = n + m$ values $X_1, \cdots, X_n, Y_1, \cdots, Y_m$. Let $F_n^*$ be the empirical distribution of these $n$ values and let $G_m^*$ be the empirical distribution of the remaining $m = N - n$ values comprising $X$. Then, the permutation distribution of $S_n$ is the distribution (conditional on $X$) of

$$S_n^*(t) = n^{1/2}[F_n^*(t) - G_m^*(t)]. \quad (4.9)$$

Retaining this terminology, we have the following.

**Proposition 4.1.** Assume $m/N \to \lambda$, where $\lambda \in (0, 1)$. Let $F_0$ be the mixture distribution, defined by $F_0 = \lambda F_Y + (1 - \lambda)F_X$. For almost all sample sequences $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$, $S_n^*(\cdot)$ converges in distribution (conditional on $X$) to a Gaussian process having $D[\mathbb{R}]$ sample paths, mean 0, and covariance function

$$E[S(t_1)S(t_2)] = F_0(t_1)[1 - F_0(t_2)]/\lambda,$$

where $t_1 \leq t_2$.

**Proof of Proposition 4.1 and Theorem 3.3.** The results will follow immediately from Lemma 4.1 if it can be shown that, for almost all sample sequences $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$, the empirical distribution $H_n$ of the combined sample of size $N$ converges uniformly to $F_0$ with probability one. But,

$$H_n = \frac{n}{N}\hat{F}_n + \frac{m}{N}\hat{G}_m.$$

By the Glivenko Cantelli Theorem, $\hat{F}_n$ converges uniformly to $F_X$ with probability one, and $\hat{G}_m$ converges uniformly to $F_Y$ with probability one. Hence, the results follow.
References


