STRONG REPRESENTATION OF WEAK CONVERGENCE

Z. D. Bai and W. Q. Liang

Center for Multivariate Analysis
University of Pittsburgh

Center for Multivariate Analysis
University of Pittsburgh

85 9 10 112

Approved for public release, distribution unlimited
STRONG REPRESENTATION OF WEAK CONVERGENCE

Z. D. Bai and W. Q. Liang
Center for Multivariate Analysis
University of Pittsburgh

July 1985
Technical Report No. 85-29

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

Research partially supported by the Air Force Office of Scientific Research (AFOSR) under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.
STRONG REPRESENTATION OF WEAK CONVERGENCE

Z. D. Bai and W. Q. Liang

ABSTRACT

Let $\mu_n$, $n = 1, 2, \ldots$, and $\mu$ be a given sequence of probability measures each of which is defined on a complete separable metric space $S_n$ and $S$, respectively. Also, a sequence of measurable mappings $\phi_n$ from $S_n$ into $S$ is given. In this paper, it is proved that if $\mu_n \circ \phi_n^{-1}$ weakly converge to $\mu$, then there is a probability space $(\Omega, F, P)$, on which we can define a sequence of random elements $X_n$, from $\Omega$ into $S_n$, and a random element $X$, from $\Omega$ into $S$, such that $\mu_n$ is the distribution of $X_n$, $\mu$ is the distribution of $X$ and $\lim_{n \to \infty} \phi_n(X_n) = X$ pointwise.

The result of Skorokhod (1956) is a special case of the result of this paper. Some applications in the area of random matrices, etc., are also given.
1. INTRODUCTION

It is well known that there is a big difference between the concepts of weak and strong convergence of random variables. In the area of limiting theory, it is of interest to study the difference as well as the link between the two concepts of convergence. Recent research work motivates us to investigate them. In Section 2, we shall prove the following theorem.

THEOREM 1: Let $S_n$, $n = 1, 2, \ldots$, and $S$ be complete separable metric spaces, with distance functions $\rho_n$ and $\rho$ respectively, and let $\phi_n$ be a measurable mapping from $S_n$ into $S$. Suppose that $\mu_n$ and $\mu$ are probability measures defined on $S_n$ and $S$, the Borel $\sigma$-fields deduced by the distances $\rho_n$ and $\rho$, respectively, and suppose that $\mu_n \cdot \phi_n^{-1} \Rightarrow \mu$. Then there is a probability space $(\Omega, F, P)$ and a sequence of $S_n$-valued random elements $X_n$, and an $S$-valued random element $X$, defined on $(\Omega, F, P)$, such that

1) $X_n$ has distribution $\mu_n$ and $X$ has distribution $\mu$,

2) $\lim_{n \to \infty} \phi_n(X_n) = X$, pointwise.

In early 1956, Skorokhod proved a special case of Theorem 1, where $S_n = S$, for each $n$, and $\phi_n$ are all identity. It should be pointed out that our Theorem 1 is not a trivial generalization to Skorokhod's theorem. Theorem 1 played a key role in the proof of a theorem in Yin (1984), but Skorokhod's theorem is not applicable there.

Although Skorokhod's paper was published in early fifties, it seems that Skorokhod's theorem had not received much attention unfortunately. For instance, the Helley-Brary theorem can be easily obtained by Skorokhod theorem, but in
many recent probability textbooks, it was still proved by the approach of integration by parts. Even though the proof of Skorokhod's theorem seems a little complicated, we can give a very simple proof to the special case where $S_n = S = \mathbb{R}^d$, the finite dimensional Euclidean space.

The power of Theorem 1 appears in the situation that we often encounter in large sample theory. Suppose that $\phi_n(Y_n) \Rightarrow \phi$ and $F(\cdot, \cdot)$ is a two-variate continuous function. We are concerned with the limiting behavior of the roots of the equation $F(\phi_n(Y_n), X) = 0$. In general, the roots of $F(y, X) = 0$ do not have an obvious expression, but in many cases we can prove that the solution $x = x(y)$ is continuous in $y$. In these cases, by Theorem 1, we only need to investigate the behavior of the solution of $F(\phi, X) = 0$.

Some concrete examples can be found in Bai (1984), Bai and Yin (1984) and Yin (1984).

We generalized Lusin's theorem to the measurable mapping from a complete separable metric space into another one. This result is stated in Theorem 2 and it played a key role in the proof of Theorem 1.
2. A GENERALIZATION OF SKOROKHOD'S THEOREM

We first assume that each \( \phi_n \) is continuous. At the beginning, we construct a series of countable partitions of the space \( S \) as follows:

Let \( B(x,r) \) denote the ball in \( S \), with center \( x \) and radius \( r \). Because \( S \) is separable, there is a countable set \( \{ x_i, i = 1,2,\ldots \} \), which is dense in \( S \). Because there are at most countably many values of \( r \) such that \( u(\partial B(x_i,r)) > 0 \), for some \( i \), where \( \partial B \) denotes the boundary of the set \( B \). Thus for each \( k \), there exists \( r_k \), \( 2^{-(k+1)} < r_k < 2^{-k} \), being such that \( u(\partial B(x_i,r_k)) = 0 \) for any \( i = 1,2,\ldots \). Write \( C(k,1) = B(x_1,r_k) \), \( C(k,i) = B(x_i,r_k) \setminus \bigcup_{j=1}^{i-1} B(x_j,r_k) \), and set

\[
D_{i_1i_2\ldots i_k} = \bigcap_{j=1}^k C(j,i_j) \quad (1)
\]

for any \( i_1,i_2,\ldots,i_k = 1,2,\ldots \). It is obvious that \( \{D_{i_1\ldots i_k}\} \) satisfies the following properties:

1) \( D_{i_1\ldots i_k} \cap D_{j_1\ldots j_k} = \emptyset \) if \( (i_1,\ldots,i_k) \neq (j_1,\ldots,j_k) \),

2) \( D_{i_1\ldots i_{k-1}'i_k} = \bigcup_{i_k=1}^{r_{i_k}} D_{i_1i_2\ldots i_{k-1}i_k} \), \( S = \bigcup_{i_1=1}^{r_{i_1}} D_{i_1} \),

3) \( u(\partial D_{i_1\ldots i_k}) = 0 \),

4) \( d(D_{i_1\ldots i_k}) < 2^{-k} \), where \( d(D) \) denotes the diameter of the set \( D \).

Using the same approach, for each \( n \), we split \( S_n \) into partitions...
\{D(n)_{i_1, i_2, \ldots, i_k}\} having similar properties as \{D_{i_1 i_2, \ldots, i_k}\}. Write

\[ D_{i_1, \ldots, i_k}(j_1, \ldots, j_k) = D(n)_{i_1, \ldots, i_k} \cap \phi_n^{-1} D_{j_1, \ldots, j_k}, \tag{3} \]

and

\[ p_{i_1, \ldots, i_k}(j_1, \ldots, j_k) = u_n(D(n)_{i_1, \ldots, i_k}(j_1, \ldots, j_k)). \tag{4} \]

It is obvious that \( d(D(n)_{i_1, \ldots, i_k}(j_1, \ldots, j_k)) \leq 2^{-k}. \)

Let \( \Omega = [0,1) \), \( F \) be the \( \sigma \)-field of all Borel sets in \( \Omega \), and \( P \) be the Lebesgue measure restricted on \( F \).

Split \( \Omega \) into partitions \( \{I(n)_{i_1, \ldots, i_k}(j_1, \ldots, j_k)\} \) with the following properties:

1. Each \( I(n)_{i_1, \ldots, i_k}(j_1, \ldots, j_k) \) is an interval closed from left and open from right, and has length \( p_{i_1, \ldots, i_k}(j_1, \ldots, j_k) \).

2. For each \( n \) and each \( k \), \( \{I(n)_{i_1, \ldots, i_k}(j_1, \ldots, j_k)\} \) is a partition of \([0,1)\).

3. \( I(n)_{i_1, \ldots, i_{k-1}, (j_1, \ldots, j_{k-1})} = \bigcup_{i_k=1}^{\infty} \bigcup_{j_k=1}^{\infty} I(n)_{i_1, \ldots, i_k}(j_1, \ldots, j_k) \).

4. If \( i_k < i_k' \), for any \( i_1, \ldots, i_{k-1}, j_1, \ldots, j_k \), \( j_1', \ldots, j_k' \), \( I(n)_{i_1, \ldots, i_k}(j_1, \ldots, j_k) \) is located on the left of \( I(n)_{i_1, \ldots, i_{k-1}, i_k'}(j_1', \ldots, j_k') \).
5) If $j_t < j_{t+1}$, $t < k$, then for any $i_1, \ldots, i_k$, $j_1, \ldots, j_{t-1}, j_{t+1}, \ldots, j_k$, $j_1, \ldots, j_{t-1}, j_{t+1}, \ldots, j_k$ is located on the left of $I_{i_1, \ldots, i_k}^{(n)} (j_1, \ldots, j_k)$. We take a point $x_{i_1, \ldots, i_k}^{(n)} (j_1, \ldots, j_k)$ arbitrarily from $D_{i_1, \ldots, i_k}^{(n)} (j_1, \ldots, j_k)$ if it is not empty and define

$$x_n^{(k)}(\omega) = x_{i_1, \ldots, i_k}^{(n)} (j_1, \ldots, j_k), \text{ if } \omega \in I_{i_1, \ldots, i_k}^{(n)} (j_1, \ldots, j_k). \quad (6)$$

Evidently, for each $n$ and $k$, $x_n^k$ is measurable. Because

$$D_{i_1, \ldots, i_k+1}^{(n)} (j_1, \ldots, j_k+1) \subset D_{i_1, \ldots, i_k}^{(n)} (j_1, \ldots, j_k)$$

and

$$d \left( D_{i_1, \ldots, i_k}^{(n)} (j_1, \ldots, j_k), \right) < 2^{-k},$$

we have

$$d_n \left( x_n^{(k)}(\omega), x_n^{(k+1)}(\omega) \right) < 2^{-k}. \quad (7)$$

Thus for each $n$, $(x_n^{(k)}, k = 1, 2, \ldots)$ forms a Cauchy sequence and there exists a measurable function $X_n$ such that

$$x_n^{(k)} \to X_n, k \to \infty, \forall \omega \in \Omega,$$

because $S_n$ is complete. Therefore, we have defined an $S_n$-valued random element $X_n$ for each $n$. 

Next, we shall prove that \( \nu_n \) is the distribution of \( X_n \). Take any open set \( A_n \subset S_n \), define \( A_{m,n} = \{ x^{(n)} \in A_n : \rho_n(x^{(n)}, x A_n) > \frac{1}{m} \} \), where \( m \) is a positive integer and \( x A_n \) is the boundary of \( A_n \) and 
\[
\rho_n(x^{(n)}, y) = \inf \{ \rho_n(x^{(n)}, y^{(n)}) : y^{(n)} \in B \}.
\]
It is obvious that for each pair \((n,m)\), \( A_{m,n} \) is an open set contained in \( A_n \) and \( A_{m,n} \subset A_{m+1,n} \). Thus we have an expression of \( A_{m,n} \) as follows
\[
A_{m,n} = \sum_{(i_1,j_1) \in N_1(n)} D_1^{(n)}(j_1) + \sum_{(i_1,i_2;j_1,j_2) \in N_2(n)} D_1^{(n)}(j_1,j_2) + \ldots
\]
where \( N_1(n), N_2(n), \ldots \) are suitable index sets and all the right hand side terms are disjoint each other.

For each \( k \) with \( 2^{-k+1} < \frac{1}{2m^2} \), we have
\[
\rho_n(x^{(n)}, x^{(k)}) < \left( \frac{1}{2} \right)^{k-1} \frac{1}{2m^2} < \frac{1}{2m^2}.
\]
Hence
\[
(X_n \in A_{m-1,n}) \subset (X_n \in A_{m,n}) \subset (X_n \in A_{m+1,n}).
\]
Thus
\[
P(X_n \in A_{m-1,n}) \leq P(X_n \in A_{m,n}) \leq P(X_n \in A_{m+1,n}).
\]

On the other hand, we have
\[
P(x^{(k)}_n \in A_{m,n}) = \sum_{(i_1,j_1) \in N_1(n)} P(x^{(k)}_n \in D_1^{(n)}(j_1))
\]
\[+ \sum_{(i_1,i_2;j_1,j_2) \in N_2(n)} P(x^{(k)}_n \in D_1^{(n)}(j_1,j_2)) + \ldots
\]
\[= \sum_{(i_1,j_1) \in N_1(n)} \nu_i^{(n)}(j_1)
\]
\[ + \sum_{(i_1, i_2, j_1, j_2) \in N_{2,m}} |I_{i_1 j_2}(j_1, j_2)| + \ldots \]
\[ = \sum_{(i_1, j_1) \in N_{1,m}} \mu_n(D_{i_1}^{(n)}(j_1)) \]
\[ + \sum_{(i_1, i_2, j_1, j_2) \in N_{2,m}} \mu_n(D_{i_1 i_2}^{(n)}(j_1, j_2)) + \ldots \]
\[ = \mu_n(A_{m,n}). \quad (12) \]

From (11) and (12) it follows that

\[ P(X_n \in A_{m-1,n}) \leq \mu_n(A_{m,n}) \leq P(X_n \in A_{m+1,n}) \leq P(X_n \in A_n). \quad (13) \]

If we let \( m \to \infty \), we obviously have \( A_{m-1,n} \uparrow A_n \), \( A_{m,n} \uparrow A_n \). Hence from (13) we get

\[ P(X_n \in A_n) = \mu_n(A_n) \quad (14) \]

Therefore, \( \mu_n \) is the distribution of \( X_n \), for each \( n \).

Write

\[ p_{i_1, \ldots, i_k} = \mu(D_{i_1, \ldots, i_k}), \]

and split \( \Omega \) into partitions \( \{I_{i_1, \ldots, i_k}\}_1 \) such that

1) for each \( k \), \( \{I_{i_1, \ldots, i_k}\}_1, i_1, \ldots, i_k = 1, 2, \ldots \) is a partition of \( \Omega \).
2) each $I_{i_1,\ldots,i_k}$ is an interval closed from left and open from right, with its length $p_{i_1,\ldots,i_k}$.

3) $I_{i_1,\ldots,i_k} = \bigcup_{i_k=1}^\infty I_{i_1,\ldots,i_k}$.

4) if $i_k < i_k'$, then $I_{i_1,\ldots,i_k}$ is located on the left of $I_{i_1',\ldots,i_k}.

We arbitrarily take a point $x_{i_1,\ldots,i_k}$ from $D_{i_1,\ldots,i_k}$ for each $(i_1,\ldots,i_k)$ and define

$$X(k) = x_{i_1,\ldots,i_k}, \text{ if } \omega \in I_{i_1,\ldots,i_k}.$$ 

Similarly as before, we can prove that there exists an $S$-valued random element $X$ such that

$$\rho(X(k), X) < 2^{-k},$$

and that $\mu$ is the distribution of $X$.

To complete the proof of the special case of Theorem 1, we need only to prove that

$$\phi_n(X_n) \to X, \text{ a.s.}$$

Write

$$I_{i_1,\ldots,i_k} = \bigcup_{j_1=1}^\infty \bigcup_{j_k=1}^\infty I(n)_{j_1,\ldots,j_k}.$$ 

According to the definition of $I(n)_{i_1,\ldots,i_k}(j_1,\ldots,j_k)$, $\{I(n)_{i_1,\ldots,i_k}\}$ has analogous properties as $\{I_{i_1,\ldots,i_k}\}$, and their length satisfies
\[ |I_{i_1 \ldots i_k}| = \sum_{j_1=1}^{\infty} \sum_{j_k=1}^{\infty} u_n(D_{i_1 \ldots j_k}) (j_1, \ldots, j_k) \]

\[ = u_n(\phi_n^{-1}(D_{i_1 \ldots i_k})). \]

Since \( u(D_{i_1 \ldots i_k}) = 0 \) and \( u \circ \phi_n \rightarrow u \), we have

\[ |I_{i_1 \ldots i_k}| = u(D_{i_1 \ldots i_k}), \text{ as } n \rightarrow \infty. \quad (17) \]

If \( \omega \) is a point of \( \mathbb{N} \) and is not an endpoint of any interval \( I_{i_1 \ldots i_k} \), \( k = 1, 2, \ldots, i_1, \ldots, i_k = 1, 2, \ldots \), then for each \( k \) there exists a \( k \)-multiple \((a_1, \ldots, a_k)\) such that \( \omega \) is an inner point of \( I_{a_1, \ldots, a_k} \). In view of the definition of \( \{I_{i_1 \ldots i_k}\} \), we know that the left and right endpoints of \( I_{a_1, \ldots, a_k} \) are

\[ a_k = \sum_{i_1=1}^{a_1-1} u(D_{i_1}) + \sum_{i_2=1}^{a_2-1} u(D_{a_1 i_2}) + \ldots + \sum_{i_k=1}^{a_k-1} u(D_{a_1 \ldots a_k-1 i_k}) \]

and

\[ b_k = a_k + u(D_{a_1, \ldots, a_k}). \]

Similarly, the two endpoints of \( I_{a_1, \ldots, a_k}^{(n)} \) are

\[ a_k^{(n)} = \sum_{i_1=1}^{a_1-1} |I_{i_1}| + \sum_{i_2=1}^{a_2-1} |I_{a_1 i_2}| + \ldots + \sum_{i_k=1}^{a_k-1} |I_{a_1 \ldots a_k-1 i_k}| \]

and

\[ b_k^{(n)} = a_k^{(n)} + u(D_{a_1, \ldots, a_k}). \]
Finally, we shall prove that

\[ u \left( x \in S_1, \phi_\epsilon(x) \text{ is discontinuous at } x \right) = 0. \]

If \( x \in \left( \bigcap_{j=1}^{\infty} K_{a_1, \ldots, a_{j-1}} \right) \), for some \( k_0 \) and \( a_1, \ldots, a_{k_0-1} \), according to the definition of \( \phi_\epsilon(x) \)

\[ \phi_\epsilon(z) = y_{a_1, \ldots, a_{k_0-1}}, \text{ for any } z \in K_0 \cap K_{a_1, \ldots, a_{k_0-1}}. \]

Hence \( \phi(x) \) is continuous at \( x \). If \( x \in \bigcup_{k=1}^{\infty} K_k \), then for any \( k \), there exists a \( k \)-ple \( (a_1, \ldots, a_k) \) such that \( x \in K_{a_1, \ldots, a_k}, a_1 \leq N_1, \ldots, a_k \leq N_k \).

Since \( K_{a_1, \ldots, a_k} \) is an open set, we have that \( \rho \left( x, \bigcap_{a_1}^{a_k} \right) = \rho_k > 0 \).

If \( y \in S_1 \) and \( \rho(x, y) < \rho_k \), then \( y \in K_{a_1, \ldots, a_k} \). Hence

\[ \rho \left( \phi_k(x), \phi_\epsilon(x) \right) < \frac{1}{2^{k-1}} \]

and

\[ \rho \left( \phi_k(y), \phi_\epsilon(y) \right) < \frac{1}{2^{k-1}} \]

Therefore,

\[ \rho \left( \phi_\epsilon(x), \phi_\epsilon(y) \right) < 1/2^{k-2}. \]
\[ \frac{\varepsilon}{2} + \sum_{j=1}^{k-1} \frac{N_j}{2^{j+1}(N_1, \ldots, N_j+1)^2} \]

\[ \frac{\varepsilon}{2} + \sum_{j=1}^{k-1} \frac{(N_j+1)\varepsilon}{2^{j+1}(N_1, \ldots, N_j+1)} \]

Thus

\[ \mu(\bigcup_{k=1}^{N} K^k_0) < \frac{\varepsilon}{2} + \sum_{j=1}^{k-1} \frac{(N_j+1)\varepsilon}{2^{j+1}(N_1, \ldots, N_j+1)}. \]  \hspace{1cm} (31)

On the other hand, we have

\[ \mu\left( \bigcup_{i=1}^{N_1} \bigcup_{i_k=1}^{N_k} (K_{i_1}, \ldots, i_k, \Delta E_{i_1}, \ldots, i_k) \right) \]

\[ \leq \sum_{i=1}^{N_1} \sum_{i_k=1}^{N_k} \mu(K_{i_1}, \ldots, i_k, \Delta E_{i_1}, \ldots, i_k) \]

\[ \leq \sum_{i=1}^{N_1} \sum_{i_k=1}^{N_k} \sum_{j=1}^{i_k} \sum_{j=1}^{i_k-1} \mu(G_{i_1}, \ldots, i_k, \epsilon E_{i_1}, \ldots, i_k, \epsilon E_{i_1}, \ldots, i_k) \]

\[ \leq N_1 \ldots N_k \left( \frac{\varepsilon}{2^{k+1}(N_1, \ldots, N_k+1)^2} \right) \leq \frac{\varepsilon}{2^{k+1}} \]  \hspace{1cm} (32)

By (30) (31) (32), we obtain

\[ \mu\left( \sigma(\varepsilon) = \sigma(x) \right) = \lim_{k \to \infty} \mu\left( \sigma(\varepsilon(\sigma(x), : (x)) > \frac{1}{2^{k-2}} \right) \]
\[
\frac{\varepsilon}{2} + \sum_{i_1=1}^{N_1} \sum_{i_{k-1}=1}^{N_{k-1}} u(E_{i_1, \ldots, i_{k-1} \setminus K_{i_1, \ldots, i_{k-1}}}) + \sum_{i_1=1}^{N_1} \sum_{i_{k-2}=1}^{N_{k-2}} u(\phi^{-1}D_{i_1, \ldots, i_{k-2} \setminus K_{i_1, \ldots, i_{k-2}}}) 
\leq \cdots
\]

\[
\frac{\varepsilon}{2} + \sum_{j=1}^{N_j} \sum_{i_1=1}^{N_1} \sum_{i_j=1}^{N_j} u(E_{i_1, \ldots, i_j \setminus K_{i_1, \ldots, i_j}})
\]

\[
\frac{\varepsilon}{2} + \sum_{j=1}^{N_j} \sum_{i_1=1}^{N_1} \sum_{i_j=1}^{N_j} \left[ u(G_{i_1, \ldots, i_j \setminus E_{i_1, \ldots, i_j}}) + u(G_{i_1, \ldots, i_j} \cap \bigcup_{\ell=1}^{i_{j-1}} G_{i_1, \ldots, i_{j-1} \setminus \ell}) \right]
\]

\[
\frac{\varepsilon}{2} + \sum_{j=1}^{N_j} \sum_{i_1=1}^{N_1} \sum_{i_j=1}^{N_j} \left[ \frac{\varepsilon}{2^{j+1}(N_1, \ldots, N_j+1)^2} + u(E_{i_1, \ldots, i_j \setminus G_{i_1, \ldots, i_j} \cup G_{i_1, \ldots, i_{j-1} \setminus \ell}) \right]
\]

\[
\frac{\varepsilon}{2} + \sum_{j=1}^{N_j} \sum_{i_1=1}^{N_1} \sum_{i_j=1}^{N_j} \left[ \frac{\varepsilon}{2^{j+1}(N_1, \ldots, N_j+1)^2} + \sum_{\ell=1}^{i_j} u(E_{i_1, \ldots, i_j \setminus G_{i_1, \ldots, i_{j-1} \setminus \ell}}) \right]
\]
We have

\begin{align*}
\mu\left( \bigcup_{j=1}^{k} K_{0}^{j} \right) & \leq \mu(S_{1} \setminus \bigcup_{i_{1}=1}^{N_{1}}, \ldots, \bigcup_{i_{k}=1}^{N_{k}} K_{1}, \ldots, i_{k}) \\
& = \mu(S_{1} \setminus \bigcup_{i_{1}=1}^{N_{1}}, \ldots, \bigcup_{i_{k}=1}^{N_{k}} G_{i_{1}}, \ldots, i_{k}) \\
& \leq \mu(S_{1} \setminus \bigcup_{i_{1}=1}^{N_{1}}, \ldots, \bigcup_{i_{k}=1}^{N_{k}} E_{i_{1}}, \ldots, i_{k}) + \sum_{i_{1}=1}^{N_{1}} \sum_{i_{k}=1}^{N_{k}} \mu(G_{i_{1}}, \ldots, \phi^{i_{k}} \Delta E_{i_{1}}, \ldots, i_{k}) \\
& \leq \mu(S_{1} \setminus \bigcup_{i_{1}=1}^{N_{1}}, \ldots, \bigcup_{i_{k}=1}^{N_{k}} \phi^{-1} D_{i_{1}}, \ldots, i_{k}) \\
& \quad + \sum_{i_{1}=1}^{N_{1}} \sum_{i_{k}=1}^{N_{k}} u(\phi^{-1} D_{i_{1}}, \ldots, \phi^{-1} D_{i_{1} \setminus K_{1}}, \ldots, \phi^{-1} D_{i_{k}-1}) \\
& + N_{1}, \ldots, N_{k} \epsilon / 2^{k+1}(N_{1}, \ldots, N_{k} + 1)^{2} \\
& \leq \frac{\epsilon}{2} - \frac{\epsilon}{2^{k+1}} + \sum_{i_{1}=1}^{N_{1}} \sum_{i_{k}=1}^{N_{k}-1} u(\phi^{-1} D_{i_{1}}, \ldots, \phi^{-1} D_{i_{1} \setminus K_{1}}, \ldots, \phi^{-1} D_{i_{k}-1}) \\
& \quad + \frac{\epsilon}{2^{k+1}(N_{1}, \ldots, N_{k} + 1)} \\
& \leq \frac{\epsilon}{2} + \sum_{i_{1}=1}^{N_{1}} \sum_{i_{k}=1}^{N_{k}-1} u(E_{i_{1}}, \ldots, \phi^{-1} D_{i_{1} \setminus K_{1}}, \ldots, \phi^{-1} D_{i_{k}-1}) \\
& \quad + \sum_{i_{1}=1}^{N_{1}} \sum_{i_{k}=1}^{N_{k}-1} u(\phi^{-1} D_{i_{1}}, \ldots, \phi^{-1} D_{i_{1} \setminus K_{1}}, \phi^{-1} D_{i_{k}-1}) \\
& + \sum_{i_{1}=1}^{N_{1}} \sum_{i_{k}=1}^{N_{k}-1} u(\phi^{-1} D_{i_{1}}, \ldots, \phi^{-1} D_{i_{1} \setminus K_{1}}, \phi^{-1} D_{i_{k}-2}) \\
& \quad + \frac{\epsilon}{2^{k+1}(N_{1}, \ldots, N_{k} + 1)}
\end{align*}
By the definition of $\phi_k$ and $\phi_{k+1}$ we get that

$$\phi_k(x) = y_{\alpha_1}, \ldots, \alpha_k \in D_{\alpha_1}, \ldots, \alpha_k$$

$$\phi_{k+1}(x) = y_{\alpha_1}, \ldots, \alpha_k \beta_{k+1} \in D_{\alpha_1}, \ldots, \alpha_k \beta_{k+1} \subset D_{\alpha_1}, \ldots, \alpha_k. $$

Therefore

$$\rho(\phi_k(x), \phi_{k+1}(x)) < 2^{-k}$$

Thus there must be a limit point, denoted by $\phi_\varepsilon(x)$, of the sequence $\phi_k(x)$. Combining this and (26) we obtain that $\lim_{k \to \infty} \phi_k(x) = \phi_\varepsilon(x)$ pointwise. (27)

Now we shall prove that

$$\mu(\phi_\varepsilon(x) \neq \phi(x)) < \varepsilon. \quad (28)$$

Note that if $x \not\in \bigcup_{k=1}^\infty k_0^k$, we always have

$$\rho(\phi_\varepsilon(x), \phi_k(x)) < 1/2^{k-1}, \text{ for any } k \geq 1. \quad (29)$$

Therefore, for any $k$

$$\mu\left(\rho(\phi_\varepsilon(x), \phi(x)) > 1/2^{k-2}\right) \leq \mu\left(\bigcup_{k=1}^\infty k_0^k\right) +$$

$$+ \mu\left(\rho\left(\phi_k(x), \phi(x) \geq \frac{1}{2^{k-1}}, x \in \bigcup_{i_1=1}^{N_1} \ldots, \bigcup_{i_k=1}^{N_k} K_{i_1}, \ldots, i_k\right)\right). \quad (30)$$
5) \( K_{i_1}, \ldots, i_{k-1}^1 = G_{i_1}, \ldots, i_{k-1}^1 \), \( K_{i_1}, \ldots, i_{k-1}^2 = (G_{i_1}, \ldots, i_{k-1}^2 \setminus G_{i_1}, \ldots, i_{k-1}^1) \)

\[ \ldots, K_{i_1}, \ldots, i_{k-1}^N_k = (G_{i_1, \ldots, i_{k-1}^N_k} \setminus \bigcup_{i_k = 1}^{N_k} G_{i_1, \ldots, i_k}^0) \]

6) \( k_0^k = \bigcup_{i_1 = 1}^{N_1} \ldots \bigcup_{i_{k-1} = 1}^{N_{k-1}} (K_{i_1}, \ldots, i_{k-1} \setminus \bigcup_{i_k = 1}^{N_k} K_{i_1}, \ldots, i_k) \)

7) \( \phi_k(x) = \begin{cases} y_{i_1}, \ldots, i_k & \text{if } x \in K_{i_1}, \ldots, i_k \quad 1 \leq i, \leq N_1, \ldots, 1 \leq i_k \leq N_k \\ \phi_{k-1}(x) & \text{if } x \in K_0 \cup \ldots \cup K_0^k. \end{cases} \)

If \( x \in k_0^k \) for some \( k_0 \), then for any \( k > k_0 \)

\[ \phi_k(x) = \phi_{k_0}(x) \quad (26) \]

because \( k_0 \cup \ldots \cup K_{i_1}, \ldots, i_k \subset S_1 \setminus (K_0^{k-1} \cup \ldots \cup K_0^1). \) If \( x \notin \bigcup_{k=1}^{\infty} k_0^k \),

then for any \( k \), \( x \in \bigcup_{i_1 = 1}^{N_1} \ldots \bigcup_{i_k = 1}^{N_k} K_{i_1}, \ldots, i_k \). Suppose that \( x \in K_{\beta_1}, \ldots, \beta_k \),

and \( x \in K_{\beta_1}, \ldots, \beta_{k+1} \). Since \( K_{\beta_1}, \ldots, \beta_{k+1} \subset G_{\alpha_1}, \ldots, \alpha_{k+1} \subset K_{\beta_1}, \ldots, \beta_k \)

and \( K_{i_1}, \ldots, i_k \)'s are disjoint, it follows that \( \beta_1 = \alpha_1, \ldots, \beta_k = \alpha_k \).
Write
\[ K_{i1} = G_{i1}, \quad K_{i2} = (G_{i2} \setminus G_{i1})^0, \quad \ldots, \quad K_{i1 N_2} = (G_{i1 N_2} \bigcup_{i2=1}^{N_2-1} G_{i1 i2})^0, \]

and
\[ K_0^2 = \bigcup_{i1=1}^{N_2} (K_{i1} \setminus \bigcup_{i2=1}^{N_2} K_{i1 i2}). \]

Define
\[ \phi_2(x) = \begin{cases} y_{i1 i2} & \text{if } x \in K_{i1 i2}, 1 \leq i1 \leq N_1, 1 \leq i2 \leq N_2, \\ \phi_1(x) & \text{if } x \in K_0^1 \cup K_0^2. \end{cases} \]

Then let \( E_{i1 i2 i3} = K_{i1 i2} \bigcap \phi^{-1}D_{i1 i2 i3}, 1 \leq i1 \leq N_1, 1 \leq i2 \leq N_2, \)
\( i \leq i3 \leq N_3. \) Similarly define \( G_{i1 i2 i3}, K_{i1 i2 i3}, K_0^3 \) and \( \phi_3(x). \) By induction we can define \( E_{i1 \ldots i_k}, G_{i1 i2 \ldots i_k}, K_{i1 \ldots i_k}, K_0^k, \phi_k(x) \)
satisfying the following relations:

1) \( E_{i1 \ldots i_k} = K_{i1 \ldots i_k} \bigcap \phi^{-1}D_{i1 \ldots i_k}, 1 \leq i1 \leq N_1, \ldots, 1 \leq i_k \leq N_k. \)

2) \( G_{i1 \ldots i_k} \subseteq K_{i1 \ldots i_k} \bigcap G_{i1 \ldots i_k}, \) \( G_{i1 \ldots i_k} \) 's are open sets.

3) \( \mu(G_{i1 \ldots i_k} \Delta E_{i1 \ldots i_k}) < \varepsilon/2^{k+1}(N_1, \ldots, N_{k+1})^2 \)

4) \( \mu(\partial G_{i1 \ldots i_k}) = 0 \)
Let \( E_{i_1} = \phi^{-1}D_{i_1} \). For each \( i_1 \), there is an open set \( G_{i_1} \) such that

\[
\mu(\partial G_{i_1}) = 0
\]

and

\[
\mu(G_{i_1} \setminus E_{i_1}) < \epsilon/(N_1+1)^2.
\]

Write

\[
K_1 = G_1, \quad K_2 = (G_2 \setminus G_1)^0, \ldots, \quad K_N = (G \setminus \bigcup_{i_1=1}^{N-1} G_{i_1})^0
\]

and \( K_0^1 = S_1 \setminus \bigcup_{i=1}^{N_1} K_i \), where \( A^0 \) denotes the interior of the set \( A \). Define

\[
\phi_1(x) = \begin{cases} 
  y_{i_1} & \text{if } x \in K_{i_1}, \ i_1 = 1, \ldots, N_1, \\
  y^0 & \text{if } x \in K_0^1.
\end{cases}
\]

Secondly, let \( E_{i_1 i_2} = K_{i_1} \cap \phi^{-1}(D_{i_1 i_2}) \). Then there exist open sets

\( G_{i_1 i_2}, \ i_1 \leq N_1, \ i_2 \leq N_2 \), such that

1) \( G_{i_1 i_2} \subseteq K_{i_1}, \ i_1 = 1, 2, \ldots, N_1, \ i_2 = 1, 2, \ldots, N_2 \),

2) \( \mu(G_{i_1 i_2} \setminus E_{i_1 i_2}) < \epsilon/(N_1 N_2 + 1)^2 \),

3) \( \mu(\partial G_{i_1 i_2}) = 0 \).
As before, we can make a slight modification on $X_n$ and $X$ so that (25) holds pointwise. Theorem 1 is proved.

Now we turn to prove Theorem 2. Suppose that $S_1$ and $S_2$ are two complete separable metric spaces, $\mu$ is a finite measure defined on $S_1$, $\phi: S_1 \to S_2$ is a measurable mapping.

Using the same approach, we split $S_2$ into a sequence of partitions \(\{D_{i_1,\ldots,i_k} : i_1,\ldots,i_k = 1,2,\ldots\}\), $k = 1,2,\ldots$ such that
\[
D_{i_1,\ldots,i_{k-1}} = \bigcup_{i_k=1}^{\infty} D_{i_1,\ldots,i_{k}}, \quad k = 2,3,\ldots
\]
\[
S_2 = \bigcup_{i_1=1}^{\infty} D_{i_1},
\]
and $d(D_{i_1,\ldots,i_k}) < 1/2^k$. For any fixed $\varepsilon > 0$, we can select a sequence of positive integers $N_1, N_2, \ldots$, such that
\[
\sum_{i_1=1}^{N_1} \sum_{i_k=1}^{N_k} \mu(\phi^{-1}D_{i_1,\ldots,i_k}) > \mu(S_1) - (\varepsilon/4 + \varepsilon/8 + \ldots + \varepsilon/2^{k+1}) = \mu(S_1) - \varepsilon/2 + \varepsilon/2^{k+1},
\]
for any $k = 1,2,\ldots$. Without any loss of generality, we can assume that each $D_{i_1,\ldots,i_k}$ is nonempty, $i_1 = 1,\ldots,N_1,\ldots,i_k = 1,\ldots,N_k$. Arbitrarily take $y_{i_1,\ldots,i_k} \in D_{i_1,\ldots,i_k}$, $i_1 \leq N_1,\ldots,i_k \leq N_k$, and $y^0 \in S_2$. 

\[
\phi_n(x_n) \to X, \text{ a.s., } n \to \infty. \tag{25}
\]
On the other hand, it is obvious that
\[ \phi_n^{-1} B \Delta \phi_n^{-1} B \subseteq D_n \]
where \( A \Delta B \) denotes \( (A \setminus B) \cup (B \setminus A) \). Thus we have
\[
|\mu_n\phi_n^{-1} B - \mu_n\phi_n^{-1} B| \leq \mu_n(\phi_n^{-1} B \Delta \phi_n^{-1} B) \leq \mu_n(D_n) \leq \frac{1}{2^n} \to 0,
\]
as \( n \to \infty \).

Therefore (24) follows from the above estimate and the fact that \( \mu_n\phi_n^{-1} \rightarrow \mu \).

According to the special case that we just proved, we can find a probability space \((\Omega, F, P)\) on which there is a sequence of random elements \( X_n \) and \( X \) such that

1) \( \mu_n \) is the distribution of \( X_n \), \( \mu \) is the distribution of \( X \),
2) \( \phi_n(X_n) \rightarrow X \) pointwise.

Since
\[
\sum_{n=1}^{\infty} P(\omega: \phi_n(X_n(\omega)) \neq \phi_n(X(\omega)))
= \sum_{n=1}^{\infty} \mu_n(x \in S_n \phi_n(x) \neq \phi_n(x))
\leq \sum_{n=1}^{\infty} \frac{1}{2^n} = 1 < \infty,
\]
by Borel-Cantelli lemma we know that
Example: Let $S_1 = [0,1]$ with Euclidean norm and Lebesgue measure, and let $S_2 = \{0,1\}$ with $\rho(0,1) = 1$. Define

$$\phi = I_{[0,\frac{1}{2}]}(x), \quad x \in S_1$$

where $I_A$ denotes the indicator of the set $A$. For any $\varepsilon < \frac{1}{2}$, we cannot find a continuous mapping $\phi_\varepsilon: S_1 \to S_2$ satisfying (22).

A little more complex example yields from the above example with the measure replaced by $\mu$:

$$\mu(B) = \frac{1}{2} L(B) + \sum_{r_n \in B} \frac{1}{2^{n+1}}, \quad B \in B([0,1])$$

where $L(B)$ is the Lebesgue measure of the set $B$ and $Q = \{r_n, \ n = 1,2,...\}$ is the set of all rational numbers in $[0,1]$.

Before we prove Theorem 2, we first use Theorem 2 to complete the proof of Theorem 1. For each $n$, according to Theorem 2, there exists a measurable mapping $\bar{\phi}_n$ such that

1) $\mu_n(\phi_n \neq \bar{\phi}_n) < \frac{1}{2^n}$,

2) $\mu_n(x \in S_n; \bar{\phi}_n$ is discontinuous at $x) = 0$.

We shall first prove that

$$\mu_{n+1}^{-1} \downarrow \mu_n$$

(24)

Let $B$ be a Borel subset of $S$ and $B_n = \phi_n^{-1}B$, $\bar{B}_n = \bar{\phi}_n^{-1}B$. Denote by $D_n = \{x \in S_n, \phi_n(x) \neq \bar{\phi}_n(x)\}$. By the definition of $\bar{\phi}_n$, we have $\mu_n(D_n) < \frac{1}{2^n}$. 
Since \( P(N) = 0 \), \( \tilde{X}_n \) and \( X_n \) (correspondingly \( \tilde{X} \) and \( X \)) have the same distribution. Thus Theorem 1 holds when \( \phi_n \) are all continuous.

Note that the continuity of \( \phi_n \) is only used in deriving that

\[
\rho(\phi_n(X_n), \phi_n(x^{(m)})) \to 0, \text{ as } m \to \infty,
\]

(see (20) and (21)). We can relax the continuity restriction as the following

\[
u_n(\{x \in S_n; \phi_n \text{ is discontinuous at } x\}) = 0,
\]

for each \( n \). Therefore, to complete the proof of Theorem 1, we only need the following generalized Lusin's Theorem.

**THEOREM 2:** Let \( S_1 \) and \( S_2 \) be two complete separable metric spaces, \( \mu \) be a finite measure defined on \( S_1 \) and let \( \phi \) be a measurable mapping from \( S_1 \) into \( S_2 \). Then for any \( \varepsilon > 0 \), there exists a measurable mapping \( \phi_\varepsilon : S_1 \to S_2 \), satisfying

1) \( \mu(\phi \neq \phi_\varepsilon) < \varepsilon \),

2) \( \mu(x \in S_1; \phi_\varepsilon \text{ is discontinuous at } x) = 0 \).

**Remark:** The main difference between Theorem 2 and the ordinary Lusin's Theorem is the condition (23). But in the general case, we cannot require that \( \phi_\varepsilon \) is continuous. This can be seen from the following example:
From this and \( \phi_n(X_n^{(k)}(\omega)) \in D_{\alpha_1 ... \alpha_k} \), we get

\[
\rho(\phi_n(X_n^{(m)}(\omega)), \phi_n(X_n^{(k)}(\omega))) < 2^{-k}. \tag{20}
\]

From (19) and (20), we get

\[
\rho(\phi_n(X_n^{(k)}(\omega)), X(\omega)) < 3 \cdot 2^{-k}. \tag{21}
\]

Since \( \phi_n \) is continuous and \( X_n^{(m)}(\omega) \to X_n(\omega), m \to \infty \), it follows that

\[
\rho(\phi_n(X_n^{(k)}(\omega)), X(\omega)) \leq 3 \cdot 2^{-k}.
\]

This proves (16) because the set of all endpoints of \( I_{i_1, \ldots, i_k} \) \( k = 1, 2, \ldots, i, \ldots, i_k = 1, 2, \ldots \), is countable, hence its Lebesgue measure is zero.

Let \( N \subset \Omega \) be the set on which \( \phi_n(X_n(\omega)) \) do not converge to \( X(\omega) \) and let \( \phi_n(X_n^{(k)}(\omega_0)) \to X(\omega_0) \). Define a new sequence of random elements as follows:

\[
\tilde{X}_n(\omega) = \begin{cases} 
X_n(\omega) & \text{if } \omega \in \Omega \setminus N \\
X_n^{(k)}(\omega_0) & \text{if } \omega \in N
\end{cases}
\]

and

\[
\tilde{X}(\omega) = \begin{cases} 
X(\omega) & \text{if } \omega \in \Omega \setminus N \\
X(\omega_0) & \text{if } \omega \in N
\end{cases}
\]

Then we have

\[
\phi_n(\tilde{X}_n(\omega)) \to \tilde{X}(\omega) \text{ pointwise.}
\]
and

\[ b_k(n) = a_k(n) + |I(n)|. \]

From (17) we have that

\[ a_k(n) \rightarrow a_k \quad (n \rightarrow \infty) \]

and

\[ b_k(n) \rightarrow b_k \quad (n \rightarrow \infty) \]

Therefore, when \( n \) is large enough, \( \omega \in I(n)_{\alpha_1, \ldots, \alpha_k} \). Hence

\[ \phi_n(x_n^{(k)}(\omega)) \in D_{\alpha_1, \ldots, \alpha_k}. \]

Note that \( x_n^{(k)}(\omega) \in D_{\alpha_1, \ldots, \alpha_k} \), we get

\[ \rho(\phi_n(x_n^{(k)}(\omega)), x_n^{(k)}(\omega)) < 2^{-k}. \]  

(18)

From (15) and (18) it follows that

\[ \rho(\phi_n(x_n^{(k)}(\omega)), x(\omega)) < 2^{-k+1}. \]  

(19)

For fixed \( n \) and \( k \), and for any \( m > k \), there exist \( \alpha_{k+1}, \alpha_{k+2}, \ldots, \alpha_m \). Such that \( \omega \in I(n)_{\alpha_1, \ldots, \alpha_k, \alpha_{k+1}, \ldots, \alpha_m} \). Thus we have

\[ x_n^{(m)}(\omega) \in \phi_n^{-1}(D_{\alpha_1, \ldots, \alpha_m}) \cap \phi_n^{-1}(D_{\alpha_1, \ldots, \alpha_k}) \]

or

\[ \phi_n(x_n^{(m)}(\omega)) \in D_{\alpha_1, \ldots, \alpha_k}. \]
Since $k$ is arbitrary, we have proved that $\phi_\epsilon(x)$ is continuous at $x$.

Thus

$$u(x \in S, \phi_\epsilon(x) \text{ is discontinuous at } x) \leq \sum_{k=1}^{\infty} \sum_{i_1=1}^{N_1} \ldots \sum_{i_k=1}^{N_k} u(ak_{i_1}, \ldots, i_k) = 0.$$  

This completes the proof of Theorem 2.

3. A SIMPLE PROOF OF THEOREM 1 FOR THE FINITE DIMENSION CASE.

3.1 ONE DIMENSION CASE

Suppose that $F_n$ and $F$ are one-dimensional distributions satisfying that $F_n \rightarrow F$ as $n \rightarrow \infty$. Let $\Omega = (0,1)$. $F = \mathcal{B}(0,1)$ and $P$ be the Lebesgue measure restricted on $\Omega$. Define

$$X_n(\omega) = \text{Sup}\{x: F_n(x) < \omega\}, \omega \in \Omega = (0,1),$$

and

$$X(\omega) = \text{Sup}\{x: F(x) < \omega\}.$$  

According to this definition, it is evident that $X_n(\omega) \leq x, \omega \in \Omega$, $x \in \mathbb{R}$ is equivalent to the fact that $F_n(x) \geq \omega$. This ensures that $F_n$ is the distribution of $X_n$. Similarly, $F$ is the distribution of $X$.

For any $\omega \in (0,1)$, take arbitrarily $x_0 < X(\omega)$ and $x_0$ is a continuous point of $F(x)$. Then $F(x_0) < \omega$. Since $x_0$ is a continuous point of $F(x)$ and $F_n \rightarrow F$, we have $F_n(x_0) \rightarrow F(x_0)$. Thus when $n$ is large enough we have $F_n(x_0) < \omega$, so that $X_n(\omega) \geq x_0$. Hence $X(\omega) \leq \lim_{n \rightarrow \infty} X_n(\omega)$ for any $\omega \in (0,1)$. 


Let \( \omega \epsilon (0,1) \) be such that for any \( \epsilon > 0, F(\omega) + \epsilon > \omega \). Take \( \epsilon > 0 \) such that \( X(\omega) + \epsilon \) is a continuous point of \( F(x) \). Since \( F_n(X(\omega) + \epsilon) > F(X(\omega) + \epsilon) > \omega \), when \( n \) is large enough we have \( F_n(X(\omega) + \epsilon) > \omega \).

Thus, according to the definition of \( X_n(\omega) \), \( X_n(\omega) \leq X(\omega) + \epsilon \). This shows that \( \lim_{n \to \infty} X_n(\omega) \leq X(\omega) \). Hence, \( \lim_{n \to \infty} X_n(\omega) = X(\omega) \). If for some \( \omega \epsilon (0,1) \), there exists a constant \( \epsilon_0 > 0 \) such that \( \omega = F(X(\omega) + \epsilon_0) \), then for any \( 0 < \epsilon < \epsilon_0, F(X(\omega) + \epsilon) \geq F(X(\omega) + \epsilon) \geq \omega = F(X(\omega) + \epsilon_0) \).

Hence \( F(X(\omega) + \epsilon) = F(X(\omega) + \epsilon_0) = \omega \). Thus there exists a rational number \( \gamma = \gamma(\omega) \epsilon (X(\omega), X(\omega) + \epsilon_0) \), corresponding to \( \omega \). If there are two points \( \omega_1 < \omega_2 \) which correspond to \( \gamma_1, \gamma_2 \) respectively, we shall prove that \( \gamma_1 < \gamma_2 \). In fact, if \( \omega_i = F(X(\omega_i) + \epsilon_i), \epsilon_i > 0, i = 1, 2 \), then \( X(\omega_1) + \epsilon_1 < X(\omega_2) \). Otherwise, \( X_1(\omega_1) < X(\omega_2) < X(\omega_1) + \epsilon_1 < X(\omega_2) + \epsilon_2 \) would imply that \( \omega_1 = F(X(\omega_1) + \epsilon_1) = F(X(\omega_2) + \epsilon_2) = \omega_2 \), contradicting the assumption that \( \omega_1 < \omega_2 \). Thus there are at most countably many \( \omega \) such that \( X_n(\omega) \to X(\omega) \). Hence

\[
X_n(\omega) \to X(\omega) \quad \text{a.s.,} \quad n \to \infty.
\]

As before, we can change the definition of \( X_n(\omega) \) and of \( X(\omega) \) at those \( \omega \)'s at which \( X_n(\omega) \to X(\omega) \), so that

\[
X_n(\omega) \to X(\omega) \quad \text{pointwise,} \quad n \to \infty.
\]
3.2 TWO DIMENSION CASE

Let \( F^{(n)}(\cdot, \cdot) \) and \( F(\cdot, \cdot) \) be two-dimensional distributions such that \( F^{(n)} \xrightarrow{w} F, n \to \infty \). Let \( F^{(n)}(\cdot) \) and \( F^{(n)}(\cdot | x) \) denote the marginal distribution of the first component and the conditional distribution of the second component when given the first component to be \( x \), corresponding to \( F(n) \). Define random vectors \((X_n, Y_n)\) on \( \Omega = (0,1) \times (0,1) \) as follows:

\[
\begin{cases}
X_n(\omega_1, \omega_2) \overset{\Delta}{=} X_n(\omega_1) \overset{\Delta}{=} \sup \left\{ x : F^{(n)}_X(x) < \omega_1 \right\}, & \text{if } \omega_1 \in (0,1) \\
Y_n(\omega_1, \omega_2) = \sup \left\{ y : F^{(n)}_Y(y | X_n(\omega_1)) < \omega_2 \right\}, & \text{if } \omega_i \in (0,1), i = 1,2.
\end{cases}
\]

We can similarly define random vector \((X, Y)\). As before we can show that

\( \{X_n(\omega_1) \leq x, Y_n(\omega_1, \omega_2) \leq y\} \) is equivalent to

\( \{F^{(n)}_X(x) \geq \omega_1, F^{(n)}_Y(y | X_n(\omega_1)) \geq \omega_2\} \), and that

\[
P(X_n \leq x, Y_n \leq y) = \int_0^\infty F^{(n)}_X(x) F^{(n)}_Y(y | X_n(\omega_1)) \, d\omega_1
\]

\[
= \int_{-\infty}^x F^{(n)}_Y(y | t) F^{(n)}_X(dt) = F_n(x, y).
\]

Similarly, \( F(\cdot, \cdot) \) is the distribution of \((X, Y)\). Using the conclusion in 3.1, we have

\( X_n(\omega_1) \longrightarrow X(\omega_1) \quad \text{a.s.} \)
with respect to the one-dimensional Lebesgue measure restricted on \((0,1)\).

By Fubini's Theorem, we know that

\[ X_n(\omega_1, \omega_2) = X_n(\omega_1) \rightarrow X(\omega_1) = X(\omega_1, \omega_2) \]

with respect to the two-dimensional Lebesgue measure restricted on \((0,1) \times (0,1)\).

Again using the conclusion about one dimension case, for any fixed \(\omega_1 \in (0,1)\), we get that

\[ Y_n(\omega_1, \omega_2) \rightarrow Y(\omega_1, \omega_2) \text{ a.s.} \]

with respect to the one-dimensional Lebesgue measure restricted on \((0,1)\).

Again using Fubini's Theorem, we obtain

\[ Y_n(\omega_1, \omega_2) \rightarrow Y(\omega_1, \omega_2) \text{ a.s.} \]

with respect to the two-dimensional Lebesgue measure restricted on \((0,1) \times (0,1)\).

For \(d\)-dimension case, the proof is the same as in two-dimension case.
4. APPLICATIONS OF THEOREM 1

4.1 HELLEY-BRAY THEOREM ([2] and [4]).

If $F_n \xrightarrow{w} F$ and $g(x)$ is a continuous bounded function, then

$$\int g(x)F_n(dx) \rightarrow \int g(x)F(dx)$$

PROOF. Construct $X_n \xrightarrow{P} F_n$, $X \xrightarrow{P} F$ and $X_n \xrightarrow{P} X$, according to Theorem 1. Then by the dominated convergence theorem we have

$$\int g(x)F_n(dx) = Eg(X_n) \rightarrow Eg(X) = \int g(x)F(dx)$$

4.2 (See [4])

If $F_n \xrightarrow{w} F$, then $f_n(t) \rightarrow f(t)$ uniformly on any bounded interval, where $f_n$ and $f$ are the characteristic functions of $F_n$ and $F$, respectively.

PROOF.

Let $T > 0$ be any fixed number. Then

$$|f_n(t) - f(t)| = |E(e^{itX_n} - e^{itX})|$$

$$\leq E|e^{itX_n} - e^{itX}|$$

$$\leq 2P(|X_n - X| \geq \epsilon/T) + \epsilon \rightarrow 0, \forall |t| \leq T.$$
4.3 (See [4]).

If \( F_n \xrightarrow{w} F \) and \( r > 0 \), then

\[
\int |x|^r F(dx) \leq \lim_{n \to \infty} \int |x|^r F_n(dx).
\]

PROOF.

Let \( X_n \rightarrow F_n \), \( X \sim F \) and \( X_n \rightarrow X \). Then what to be proved is equivalent to

\[
E|X|^r \leq \lim_{n \to \infty} E|X_n|^r.
\]

The latter is just a special case of Fatou Lemma.

4.4

If \( \{X_n\} \) converges in distribution to \( F \), \( Y_n \) to \( \mathcal{E}_a \), the degenerate distribution concentrating its mass at \( a \), and \( Z_n \) converges to \( \mathcal{E}_b \), \( b > 0 \), then \( \{(X_n + Y_n)Z_n\} \) converges in distribution to \( G(x) = F(x-a) \) (See [4], Th.4.4.6 and the corollary after it).

PROOF.

Since \( \{(X_n, Y_n, Z_n)\} \) converges in distribution to \( F(x) \mathcal{E}_a(y) \mathcal{E}_b(z) \).

By Theorem 1, we can construct \( (Z_n, Y_n, Z_n) \rightarrow (Z, a, b) \sim F(x) \mathcal{E}_a(y) \mathcal{E}_b(z) \).

Thus \( (Z_n + Y_n)Z_n \rightarrow (z+a)b - F(x^+ - a) \). Q.E.D.
Though the original proofs of the above four results are not very complicated, the proofs given here are relatively easier. In the following examples, the proofs will be involved with much difficulty if you do not use Theorem 1.

4.5 (See [1], [7] and [8]).

Suppose that \( \{W_{ij}, 1 \leq i \leq p, 1 \leq j \leq k-1\} \) in \( d \), \( \{W_{ij}, 1 \leq 1 \leq P, 1 \leq j \leq k-1\} \) and that \( \{U_{ij}^{(m)}, 1 \leq i \leq j \leq P\} \) in \( d \), \( \{U_{ij}, 1 \leq 1 \leq j \leq P\} \).

Consider the detrimental equation

\[
\det \left( \frac{1}{m} W_{m}W_{m}^{T} - \frac{1}{\sqrt{m}} C_{m} + D - \frac{1}{m} U_{m} \right) = 0,
\]

where \( W_{m} = \|W_{ij}\| : px(k-1) \), \( U_{m} = \|U_{ij}^{(m)}\| : pxp \), with \( U_{ij}^{(m)} = U_{ij}^{(m)} \),

\[
D = \begin{vmatrix}
(\lambda_{1} - \phi)I_{u_{1}} & \ldots & (\lambda_{v} - \phi)I_{u_{v}} \\
\vdots & \ddots & \vdots \\
(\lambda_{1} - \phi)I_{u_{v}} & \ldots & (\lambda_{v} - \phi)I_{u_{v}} \\
-\phi I_{p-r}
\end{vmatrix}
\]

\[
C_{m} = \begin{vmatrix}
C_{i_{1}}^{(m)}, \ldots, C_{i_{v}}^{(m)}, E_{1}^{(m)} \\
\vdots & \ddots & \vdots \\
C_{v_{1}}^{(m)}, \ldots, C_{v_{v}}^{(m)}, E_{v}^{(m)} \\
E_{v_{1}}^{(m)}, \ldots, E_{v}^{(m)}, 0
\end{vmatrix}
\]
\[ c_{gh}^{(m)} = \left| \sqrt{\lambda_g} w_{ij}^{(m)} + \sqrt{\lambda_h} w_{ij}^{(m)} \right|, \quad i = a_h - 1, \ldots, a_h, \quad j = a_{g-1} + 1, \ldots, a_g, \quad 1 \leq h \leq g \leq v \]

\[ e_h^{(m)} = \left| \sqrt{\lambda_h} w_{ij}^{(m)} \right|, \quad i = r + 1, \ldots, p, \quad j = a_h + 1, \ldots, a_h \]

and

\[ a_0 = 0, \quad a_h = a_h - 1 + u_h, \quad h = 1, 2, \ldots, v, \quad a_v = r + p - \lambda_1 > \ldots > \lambda_v > 0. \]

Let \( \phi_1^{(m)} \geq \ldots \geq \phi_p^{(m)} \geq 0 \) be roots of this determinantal equation and let \( \phi_i^{(m)} = m(m_i^{(m)} - \lambda_i) \), \( i = a_h - 1, \ldots, a_h, \quad h = 1, 2, \ldots, v, \) and \( \phi_i^{(m)} = m \phi_i^{(m)} \), \( i = r + 1, \ldots, p. \) Then the joint distribution of \( (Z_1^{(m)}, \ldots, Z_p^{(m)}) \) tends in distribution to that of \( Z_1, \ldots, Z_p \), where \( Z_{a_h-1} \geq \ldots \geq Z_{a_h} \) are the roots of

\[ \text{det}(C_{hh} + \lambda_h U_h - Z \lambda_i^{(m)}) = 0, \quad h = 1, 2, \ldots, v, \]

and \( Z_{r+1} \geq \ldots \geq Z_p \) are the roots of

\[ \text{det} \left( \left| d_{ij} \right| - Z I_{p-r} \right) = 0, \]

where

\[ c_{kk} = \left| \lambda_h (W_{ij} + W_{ji}) \right|, \quad i, j = a_h - 1, \ldots, a_h, \]

\[ u_{kk} = \left| U_{ij} \right|, \quad j = a_h - 1, \ldots, a_h, \quad U_{ij} = U_{ji} \]

\[ a_{ij} = \sum_{\ell=r+1}^{k-1} W_{i\ell} W_{j\ell}, \quad i, j = r + 1, \ldots, p. \]
PROOF.

According to Theorem 1, without loss of generality, we can assume that \( W_{ij}^{(m)} \to W_{ij} \) and \( U_{ij}^{(m)} \to U_{ij} \) pointwise. The explicit proof refers to [8] and is omitted here.

4.6 (See [2], [5], [6])

Suppose that \((X_{n1}, X_{n2}, \ldots, X_{nk}) \xrightarrow{w} (X_1, \ldots, X_k)\) for any \(k\), and that for each \(k\),

\[
\lim_{K \to \infty} \lim_{n \to \infty} \sup_{1 \leq k \leq K} |X_{nk}| = 0 \quad \text{in probability,} \quad \sum_{k=1}^{\infty} X_{nk} \quad \text{and} \quad \sum_{k=1}^{\infty} X_k \quad \text{a.s. converges.}
\]

Also suppose that \(g_k(t)\) is uniformly bounded in \(k\) and \(t\). Then the sequence of stochastic processes \(\sum_{k=1}^{\infty} X_{nk} g_k(t)\) weakly converges to the stochastic process \(\sum_{k=1}^{\infty} X_k g_k(t)\).

PROOF.

Set \(S = (x_1, x_2, \ldots,): \sum_{k=1}^{\infty} |x_k| < \infty\). Define

\[
\left<(x_1, x_2, \ldots), (y_1, y_2, \ldots)\right> = \sum_{k=1}^{\infty} |x_k - y_k|.
\]

Then it is easy to see that \(S\) is a complete separable metric space and \((X_{n1}, X_{n2}, \ldots), (X_1, X_2, \ldots, )\) are random elements on \(S\) with property

\[
(X_{n1}, \ldots, X_{nk}, \ldots, ) \xrightarrow{w} (X_1, \ldots, X_k, \ldots, )
\]

According to Theorem 1, we can assume that this convergence is true pointwise. It is not difficult to show that

\[
|\sum_{k=1}^{\infty} X_{nk} g_k(t) - \sum_{k=1}^{\infty} X_k g_k(t)| \leq M \sum_{k=1}^{\infty} |X_{nk} - X_k| \to 0 \quad \text{a.s.}
\]

and the proof is complete.
For details of this example, the reader is referred to Bai and Yin (1984). The proof of Theorem 5.2 given there can be greatly simplified by using Theorem 1. In all the above examples, we can use Skorokhod's Theorem. In the following we shall give an example to show that Skorokhod's Theorem is unapplicable.

4.7

Suppose that \( X_p = (X_{ij}) : pxn \) and \( T_p = (t_{ij}^{(p)}) \) satisfy

1) \( (X_{ij}, i, j=1,2,...) \) are i.i.d. random variables with mean zero and variance \( \sigma^2 > 0. \)

2) For each \( p \), \( T_p \) is a non-negative definite random matrix.

3) \( X_p \) is independent of \( T_p \).

4) \( \frac{1}{p} \text{ trace } t_p \xrightarrow{\text{in } p} H_k \) as \( p \to \infty \), for each \( k \).

5) \( \frac{p}{n} \xrightarrow{\text{ Pr.}} y \in (0, \infty) \), \( p \to \infty \).

Then for any \( k > 1 \)

\[
\frac{1}{p} \text{ trace } \left( \frac{1}{n} X_p X_p' T_p \right)^k \xrightarrow{\text{Pr.}} E_k
\]

where \( E_k \) is a constant depending only upon \( \sigma \), \( y \) and \( H_1, ..., H_k \), (See [9]).

**PROOF.**

Take \( S_p = R^{np} + \frac{1}{2} P(P+1) \), the Euclidean Space

\( S_p = \{ \frac{1}{p} \text{ trace } T_p^i, i = 1,2,...,k \} : S_p \xrightarrow{\text{ Pr.}} [0, \infty)^k \) and \( \nu_p \) the measure on \( S_p \), derived by \( (X_p, T_p) \).
By the assumptions we have

\[ \eta_p \xrightarrow{w} E(H_1, \ldots, H_k). \]

Thus, we can assume, by Theorem 1, that for fixed \( k \),

\[ \left\{ \frac{1}{p} \text{trace} T^i_p, \ i = 1, 2, \ldots, k \right\} \longrightarrow \{H_1, \ldots, H_k\} \text{ pointwise.} \]

After truncation and centralization on \( \{X_{ij}, i, j=1,2,\ldots\} \),

we can prove that

\[ E\left[ \frac{1}{p} \text{trace} \left\{ \frac{1}{p} \tilde{X}_p \tilde{X}_p^i T_p \right\}^k | T_p \right] \longrightarrow E_k, \ \ p \rightarrow \infty. \]

and

\[ \sum_{p=1}^{\infty} E\left[ \left( \frac{1}{p} \text{trace} \left\{ \frac{1}{p} \tilde{X}_p \tilde{X}_p^i T_p \right\}^k - H_k \right)^2 | T_p \right] < \infty \]

where \( \tilde{X}_p = ||X_{ij}^{(p)}||, pxn \) and \( X_{ij}^{(p)} \) is the random variable obtained from \( X_{ij} \) by truncation and centralization. Thus

\[ \frac{1}{p} \text{trace} \left\{ \frac{1}{p} X_p X_p^i T_p \right\}^k \longrightarrow E_k, \ a.s. \]

and consequently

\[ \frac{1}{p} \text{trace} \left\{ \frac{1}{p} X_p X_p^i T_p \right\}^k \longrightarrow E_k, \ a.s. \]

Since we have used Theorem 1, the above expression only implies that the convergence to be proved is true in probability. The details of the proof can refer to [9].

Note that in this example Skorohod's Theorem is unapplicable, because \( S_p \), defined here, is not the same.
REFERENCES


Let \( \nu_n, n = 1, 2, \ldots \), and \( \nu \) be a given sequence of probability measures each of which is defined on a complete separable metric space \( S_n \) and \( S \) respectively. Also, a sequence of measurable mappings \( \phi_n \) from \( S_n \) into \( S \) is given. In this paper, it is proved that if \( \nu_n \circ \phi_n^{-1} \) weakly converge to \( \nu \), then there is a probability space \((\Omega, F, P)\), on which we can define a sequence...
of random elements $X_n$, from $\mathcal{N}$ into $S_n$, and a random element $X$, from $\mathcal{N}$ into $S$, such that $\nu_n$ is the distribution of $X_n$, $\nu$ is the distribution of $X$ and $\lim_{n \to \infty} \nu_n(X_n) = X$ pointwise.

The result of Skorokhod (1956) is a special case of the result of this paper. Some applications in the area of random matrices, etc., are also given.