WEAK CONVERGENCE OF U STATISTICS

BY

RUPERT G. MILLER, JR.

TECHNICAL REPORT NO. 26
DECEMBER 21, 1970

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-13

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
WEAK CONVERGENCE OF U STATISTICS

BY

RUPERT G. MILLER, JR.

TECHNICAL REPORT NO. 26
DECEMBER 21, 1970

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT USPHS-STI GM 25-13

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
WEAK CONVERGENCE OF U STATISTICS

Rupert G. Miller, Jr.
Stanford University

1. Introduction.

Let $X_1, \ldots, X_n$ be independently, identically distributed random variables. Let $f(x_1, \ldots, x_k)$ be a real-valued symmetric function of its $k$ arguments $x_1, \ldots, x_k$ (i.e., $f(x_1, \ldots, x_k) = f(x_{i_1}, \ldots, x_{i_k})$ where $i_1, \ldots, i_k$ is any permutation of $1, \ldots, k$). A $U$ statistic with $f$ as its kernel is defined by

\begin{equation}
U_n = \frac{1}{\binom{n}{k}} \sum_{C_n} f(X_{i_1}, \ldots, X_{i_k}),
\end{equation}

where the summation is over all $\binom{n}{k}$ combinations of $k$ variables $X_{i_1}, \ldots, X_{i_k}$ selected from $X_1, \ldots, X_n$.

It has been known for some time (Hoeffding [5]) that $U_n$ has a limiting normal distribution. Precisely, if $E(f^2(X_1, \ldots, X_k)) < +\infty$, then as $n \to +\infty$,

\begin{equation}
\frac{\sqrt{n} (U_n - \mu)}{k_0} \overset{d}{\to} N(0, 1),
\end{equation}

where
\[ u = \mathbb{E}(f(X_1, \ldots, X_k)) , \]

\[ \sigma_1^2 = \text{Var}(f_1(X_1)) , \]

\[ f_1(x) = \mathbb{E}(f(X_1, \ldots, X_k) | X_1 = x) , \]

and \( \xrightarrow{d} N(0, 1) \) denotes convergence in distribution to a normal distribution with mean 0, variance 1. Examination of the joint moments of \( U_m \) and \( U_n \), \( m < n \), suggests a stronger type of convergence. For \( m, n \to +\infty \), \( m < n \),

\[ \text{Var}(U_m) = \frac{1}{\binom{m}{k}^2} \sum_{c=1}^{k} \binom{m}{k} \binom{k}{c} \binom{n-k}{k-c} \sigma_c^2 \]

\[ = \frac{k^2 \sigma_1^2}{m} + O\left(\frac{1}{m^2}\right) , \]

\[ (1.4) \]

\[ \text{Cov}(U_m, U_n) = \frac{1}{\binom{m}{k} \binom{n}{k}} \sum_{c=1}^{k} \binom{m}{k} \binom{k}{c} \binom{n-k}{k-c} \sigma_c^2 \]

\[ = \frac{k^2 \sigma_1^2}{n} + O\left(\frac{1}{n^2}\right) , \]

where

\[ \sigma_c^2 = \text{Var}(f_c(X_1, \ldots, X_c)) , \]

\[ f_c(X_1, \ldots, X_c) = \mathbb{E}(f(X_1, \ldots, X_k) | X_1 = x_1, \ldots, X_c) . \]

From (1.4) it follows that the correlation between \( U_m \) and \( U_n \), \( m < n \), is asymptotically behaving like \( \sqrt{m/n} \). This, together with the
asymptotic normality of $U_n$ and $U'_n$, suggests that the sequence
$U_n$, $n = 1, 2, \ldots$, is behaving like a Wiener process.

Define $X_n(t) = 0$ for $0 \leq t \leq (k-1)/N$. For $n = k, \ldots, N$ let

$$X_n^*(t) = \frac{n(U_n - \mu)}{\sqrt{n} \sigma_1},$$

and for $n/N < t < (n+1)/N$ define $X_n(t)$ to be the point on
the straight line segment connecting $X_n(n/N)$ and $X_n((n+1)/N)$.

Theorem 1 in Section 2 proves that as $N \to \infty$ the random functions
$X_n(\cdot)$ converge weakly to a Wiener process $X(\cdot)$ with $E(X(t)) = 0$,
$E(X^2(t)) = t$, $0 \leq t \leq 1$, in the space $C[0,1]$ of all continuous functions
defined on $[0,1]$ with the uniform topology $\rho(x(\cdot), y(\cdot))$
$$= \max_{0 \leq t \leq 1} |x(t) - y(t)|.$$ (See Billingsley [5] for terminology and
concepts.) This theorem verifies the intuitive notion gained from the
 correlation structure of the $U_n$.

In Section 3 this theorem is extended to the two sample case
where the kernel $f$ is a function of two sets of variables

$$f(x_1, \ldots, x_{k_1}; y_1, \ldots, y_{k_2})$$

which is symmetric under permutations
of $(x_1, \ldots, x_{k_1})$ and permutations of $(y_1, \ldots, y_{k_2})$. The generalizations
to the case where $X_i$ (and $Y_i$) are vector random variables
are also mentioned. In all the theorems the only condition on $f$ and
the random variables $X_i$ (and $Y_i$) for weak convergence to the Wiener
process is $E(f^2) < +\infty$.

These results represent another extension of Donsker's Theorem
to partial sums of dependent random variables. Other extensions involving
mixing conditions, martingale conditions, etc. have already been
established, and the reader is referred to Billingsley [3] for their
exposition. The lack of any conditions on the \( U \) statistic other
than the finiteness of its variance should make the theorems in this
paper useful. Many statistics pertinent to statistical inference are
\( U \) statistics (e.g., the mean, variance, two-sample Wilcoxon statistic,
rank correlation coefficients), and others are simple functions of
several \( U \) statistics (e.g., \( t \) statistics, one-sample Wilcoxon
statistic, product-moment correlation coefficient). For additional
examples of \( U \) statistics the reader is referred to Hoeffding [5].

Weak convergence of \( U \) statistics to a Wiener process will
aid in the sequential analysis of these statistics. It shows that the
probability of a sequence of \( U \) statistics hitting a barrier (and the
distribution of the hitting time) can be approximated by the behavior
of a Wiener process with respect to an analogous barrier. There is some
evidence (see Miller [8]) that \( N \) does not have to be very large for
this to be a valid approximation. When the analytical or numerical
problems involved in evaluating the sequential behavior of a \( U \)
statistic become insurmountable, the corresponding Wiener calculations
may be less formidable analytically or numerically.

2. Main Theorem.

Lemma 1 below is needed in the proof of Theorem 1 so it is
presented first.
Let $Z_n$, $n = 1, 2, \ldots$, be a sequence of random variables and $\mathcal{F}_n$, $n = 1, 2, \ldots$, a decreasing sequence of o-fields (i.e., $\mathcal{F}_n \supseteq \mathcal{F}_{n+1}$) for which $Z_n$ is measurable with respect to $\mathcal{F}_n$. The sequence $(Z_n, \mathcal{F}_n)$ is a reverse martingale if $E|Z_n| < +\infty$ and

\begin{equation}
Z_{n+1} = E[Z_n \mid \mathcal{F}_{n+1}], \quad \text{a.s.,}
\end{equation}

for $n = 1, 2, \ldots$.

**Lemma 1.** Let $(Z_n, \mathcal{F}_n)$, $n = 1, 2, \ldots$, be a reverse martingale with $E(Z_n) = 0$ and $E(Z_n^2) < +\infty$. Then,

\begin{equation}
P\left( \sup_{1 \leq n < \infty} |Z_n| > \lambda \right) \leq \frac{1}{\lambda^2} E(Z_1^2).
\end{equation}

**Proof.** The proof of the analogous result for a forward martingale is given in many books, for example, Loève [7], Sec. 30.2. For a reverse martingale the direction in the sequence is simply reversed.

**Theorem 1.** Let $X_N(t)$, $0 \leq t \leq 1$, be defined by

\begin{align*}
X_N(t) &= 0, & 0 \leq t \leq (k-1)/N, \\
X_N\left(\frac{n}{N}\right) &= \frac{n(U_n - \mu)}{\sqrt{N} k \sigma} & n = k, \ldots, N, \\
(2.3) \quad X_N(t) &= X_N\left(\frac{n}{N}\right) + N(t - \frac{n}{N}) \left( X_N\left(\frac{n+1}{N}\right) - X_N\left(\frac{n}{N}\right) \right), & \frac{n}{N} < t < \frac{n+1}{N}.
\end{align*}
where \( U_n, k, \mu, \sigma_1 \) are defined in (1.1), (1.3). If \( E(f(X_1, \ldots, X_k)) < +\infty \), then as \( N \to +\infty \), \( X_N(\cdot) \) converges weakly in the uniform topology on \( C[0,1] \) to a Wiener process \( X(\cdot) \) with \( E(X(t)) = 0 \), \( E(X^2(t)) = t, 0 \leq t \leq 1 \).

**Proof.** Let \( P_N \) denote the probability measure on \( C[0,1] \) governing \( X_N(\cdot) \), and let \( P \) denote the Wiener measure on the same space. To prove that \( P_N \) converges weakly to \( P \) it is sufficient (see Billingsley [5], Theorem 8.1) to verify the convergence of the finite dimensional distributions and the tightness of \( \{P_N\} \).

Let \( 0 \leq t_1 \leq t_2 \leq \cdots \leq t_K \leq 1 \) be \( K \) fixed times. It has been shown (see Hoeffding [5], pp. 307-308) that

\[
(2.4) \quad U_n - \mu = \frac{k}{n} \sum_{i=1}^{n} (f_i(X_i) - \mu) + R_n,
\]

where \( f_i(x) \) is defined in (1.3) and \( E(R_n) = 0, E(R_n^2) = O(1/n^2) \) provided \( E(f^2) < +\infty \). Define \( S_n = \sum_{i=1}^{n} (f_i(X_i) - \mu) \). From (2.3) and (2.4)

\[
(2.5) \quad X_N(t_j) = \frac{\lfloor Nt_j \rfloor (U_{\lfloor Nt_j \rfloor} - \mu)}{\sqrt{N} \ \sigma_1} + N(t - \frac{\lfloor Nt_j \rfloor}{N}) Q_{\lfloor Nt_j \rfloor},
\]

\[
= \frac{S_{\lfloor Nt_j \rfloor}}{\sqrt{N} \ \sigma_1} + \frac{\lfloor Nt_j \rfloor R_{\lfloor Nt_j \rfloor}}{\sqrt{N} \ k\sigma_1} + N(t - \frac{\lfloor Nt_j \rfloor}{N}) Q_{\lfloor Nt_j \rfloor},
\]

where \( \lfloor Nt_j \rfloor \) is the greatest integer less than or equal to \( N(t_j) \) and \( Q_{\lfloor Nt_j \rfloor} = X_N(\lfloor Nt_j \rfloor + 1)/N) - X_N(\lfloor Nt_j \rfloor)/N) \). Due to Chebychev's inequality
and \( E(R_n^2) = O(n^{-2}) \), each term \( \frac{R_{[N_t_j]}^2}{\sqrt{N} \sigma_1} \), \( j = 1, \ldots, K \), converges in probability to zero. The factor \( N(t - \frac{[N_t_j]}{N}) \) is always less than 1, and

\[
Q_{[N_t_j]} = \frac{1}{\sqrt{N} \sigma_1} \left\{ (k-1) \left( U_{[N_t_j]} - \mu \right) - k \left[ \frac{1}{C_{[N_t_j]}} \sum_{k-1} f(x_{i_1}, \ldots, x_{i_k}) - \mu \right] \right\},
\]

where \( C_{[N_t_j]} \) denotes summation over all combinations \( X_{i_1}, \ldots, X_{i_k} \) from \( X_1, \ldots, X_{[N_t_j]+1} \) in which \( X_{[N_t_j]+1} \) is included as one of the members. Since both quantities inside the braces in (2.6) have zero expectations and bounded second moments when \( E(f^2) < +\infty \),

\( Q_{[N_t_j]} \xrightarrow{p} 0 \) for \( j = 1, \ldots, K \) by Chebyshev's inequality.

Since the remainder terms converge in probability to zero, \( (X_N(t_1), \ldots, X_N(t_K)) \) has the same limiting joint distribution as

\[
\left( \frac{S_{[N_t_1]}}{\sqrt{N} \sigma_1}, \ldots, \frac{S_{[N_t_K]}}{\sqrt{N} \sigma_1} \right).
\]

Since \( f_i(X_i), i = 1, \ldots, N \) is a sequence of independently, identically distributed random variables with mean \( \mu \), variance \( \sigma_i^2 \), the limiting distribution for (2.7) is a multivariate normal with zero means, variances \( t_1, \ldots, t_K \), and covariance \( t_{ij} \) between the \( i \)th and \( j \)th elements \( (i < j) \). This establishes the convergence of the finite dimensional distributions.

To prove the tightness of the \( P_N \) it is necessary and sufficient (see Billingsley [3], Theorem 8.2) to prove:
(i) for each $\gamma > 0$, there exists a $c$ such that

\[(2.8) \quad P(|X_N(0)| > c) < \gamma,\]

and

(ii) for each $\epsilon > 0$, $\gamma > 0$, there exists a $\delta$, $0 < \delta < 1$, and an integer $N_0$ such that

\[(2.9) \quad P\left\{ \sup_{|s-t| < \delta} |X_N(s) - X_N(t)| > \epsilon \right\} < \gamma, \quad N \geq N_0.\]

Condition (i) is trivially satisfied since $X_N(0) = 0$. Because of the polygonal character of $X_N(\cdot)$ it is sufficient to prove that for each $\epsilon > 0$, $\gamma > 0$, there exists a $\delta$, $0 < \delta < 1$, and an integer $N_0$ such that

\[(2.10) \quad P\left\{ \max_{|m-n| < N0} |X_N^M - X_N^N| > \epsilon \right\} < \gamma, \quad N \geq N_0,\]

for $N \geq N_0$, in order to satisfy (2.9).

Since $X_N(n/N) = (kS_n + nR_n)/\sqrt{N} k\sigma_1$,

\[(2.11) \quad P\left\{ \max_{|m-n| < N0} |X_N^M - X_N^N| > \epsilon \right\} \leq P\left\{ \max_{|m-n| < N0} \left( \frac{S_m - S_n}{\sqrt{N} \sigma_1} + \frac{mR_m}{\sqrt{N} k\sigma_1} + \frac{nR_n}{\sqrt{N} k\sigma_1} \right) > \epsilon \right\} \leq P\left\{ \max_{|m-n| < N0} \left( \frac{S_m - S_n}{\sqrt{N} \sigma_1} + \frac{nR_n}{\sqrt{N} k\sigma_1} \right) > \epsilon \right\} \leq P\left\{ \max_{k \leq n \leq N} \frac{S_m - S_n}{\sqrt{N} \sigma_1} > \epsilon \right\}.\]
A graphical picture quickly shows that for any two random variables \( X, Y \) the relation \( P(X + Y > \varepsilon) \leq P(X > \varepsilon/2) + P(Y > \varepsilon/2) \) is valid so that the last expression in (2.11) is less than or equal to

\[
(2.12) \quad P \left\{ \max_{|m-n| < N \delta} \left| \frac{S_m - S_n}{\gamma N \sigma_1} \right| > \frac{\varepsilon}{2} \right\} + P \left\{ \max_{k \leq n \leq N} \left| \frac{nR_n}{\sqrt{N} \kappa_{\sigma_1}} \right| > \frac{\varepsilon}{4} \right\}.
\]

The first probability in (2.12) can be made less than \( \gamma/2 \) for \( \delta \leq \delta_1, n \geq N_1 \) by the known tightness of the probability measures for a sum of iid random variables (see Billingsley [3], Ch. 2). To prove that the second probability can be made less than \( \gamma/2 \) for \( n \geq N_2 \), define \( N_i = k + \left[ \frac{i \cdot N - k}{\ell_N} \right], i = 0, 1, \ldots, \ell_N \), where \( \ell_N \) will be selected later. (\([x]\) again denotes the greatest integer less than or equal to \( x \).) Then,

\[
(2.13) \quad P \left\{ \max_{k \leq n \leq N} \left| \frac{nR_n}{\sqrt{N} \kappa_{\sigma_1}} \right| > \frac{\varepsilon}{4} \right\} = \sum_{i=0}^{\ell_N-1} P \left\{ N_i \leq n \leq N_{i+1} \left| \max_{n_i \leq n \leq n_{i+1}} \left| \frac{nR_n}{\sqrt{N} \kappa_{\sigma_1}} \right| > \frac{\varepsilon}{4} \right\} \]

\[
\leq \sum_{i=0}^{\ell_N-1} P \left\{ N_i \leq n \leq N_{i+1} \left| \frac{R_n}{\sqrt{N} \kappa_{\sigma_1}} \right| > \frac{\varepsilon}{4} \right\} \]

\[
\leq \sum_{i=0}^{\ell_N-1} P \left\{ n_i \leq n < +\infty \left| R_n \right| > \frac{k\sigma_{\varepsilon}}{\ln N_i} \right\}.
\]
Let \( \mathcal{F}_n \) be the \( \sigma \)-field generated by the random variables 
\( X_{n+1}, X_{n+2}, \ldots \) and the order statistics \( X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)} \).

The remainder \( R_n = (U_n - \mu) - (kS_n/n) \) is the difference of two averages: the first over all combinations of \( k \) variables selected from \( X_1, \ldots, X_n \) and the second over all combinations of one variable from \( X_1, \ldots, X_n \).

This is equivalent to taking the conditional expectations of \( f(X_1, \ldots, X_k) \) and \( f_1(X_1) \) given \( X_{(1)} \leq \cdots \leq X_{(n)} \). Thus, \( R_n = E[f(X_1, \ldots, X_k) - kf_1(X_1) | \mathcal{F}_n] \).

Since \( \mathcal{F}_n \supset \mathcal{F}_{n+1} \),

\[
E[R_n | \mathcal{F}_{n+1}] = E[E[f(X_1, \ldots, X_k) - kf_1(X_1) | \mathcal{F}_n] | \mathcal{F}_{n+1})
\]

\[
(2.14)
= E[f(X_1, \ldots, X_k) - kf_1(X_1) | \mathcal{F}_{n+1}] = R_{n+1},
\]

so \( (R_n, \mathcal{F}_n), n = k, k+1, \ldots \) is a reverse martingale. (Cf., Berk [2], p. 56, Arveson [1], p. 2078, and Geertsema [4], p. 1024).

By Lemma 1 the last sum in (2.13) is bounded above by

\[
(2.15) \quad \frac{16}{k^2 \epsilon^2} \sum_{i=0}^{\frac{k}{2}} \left( \frac{n_i + 1}{n_i} \right)^2 \frac{E(R_i^2)}{n_i}.
\]

Since \( E(R_i^2) = O(1/n_i^4) \), there exists a constant \( C \) such that

\[
(2.15) \quad \text{is less than or equal to}
\]

\[
(2.16) \quad \frac{C}{N} \sum_{i=0}^{\frac{k}{2}} \left( \frac{n_i + 1}{n_i} \right)^2 = \frac{C}{Nk^2} \left( k + \left( \frac{n_i - k}{N} \right)^2 \right) + \frac{C}{N} \sum_{i=1}^{\frac{k}{2}} \left( \frac{n_i + 1}{n_i} \right)^2.
\]

10
The ratio \( \frac{n_i^{N+1}}{n_i^N} \) for \( i = 1, \ldots, \ell_{N-1} \), remains bounded as \( N \to +\infty \), because

\[
(2.17) \quad \frac{n_i^{N+1}}{n_i^N} = \frac{k + [(i+1) \cdot \frac{N-k}{\ell_N}]}{k + [i \cdot \frac{N-k}{\ell_N}]} \leq \frac{k + (i+1) \cdot \frac{N-k}{\ell_N}}{k-1 + i \cdot \frac{N-k}{\ell_N}} \leq \max\{\frac{k}{k-1}, \frac{i+1}{i}\} \leq 2.
\]

(The case \( k = 1 \) can be disregarded since \( R_m \equiv 0 \) when \( k = 1 \).) Thus, (2.16) is bounded by

\[
(2.18) \quad \frac{C}{Nk^2} \left( k + \frac{N-k}{\ell_N} \right)^2 + \frac{\ell_N \cdot (\ell_N - 1)}{N}.
\]

If \( \ell_N \) is chosen so that (i) \( \ell_N^2 / N \to +\infty \) and (ii) \( \ell_N / N \to 0 \), then both terms in (2.18) converge to zero as \( N \to +\infty \). (For example, \( \ell_N = \sqrt{N} \)). Therefore, there exists an integer \( N_2 \) such that for \( N \geq N_2 \) the expression in (2.18), and consequently the second term in (2.12), is less than \( \gamma / 2 \). Selection of \( \delta = \delta_{11} \) and \( N_0 = \max(N_1, N_2) \) establishes the tightness of \( \{P_N\} \).

3. **Generalizations.**

Let \( f(x_1, \ldots, x_{k_1}; y_1, \ldots, y_{k_2}) \) be a real-valued function which is symmetric in \( x_1, \ldots, x_{k_1} \) and symmetric in \( y_1, \ldots, y_{k_2} \). A \( U \) statistic with kernel \( f \) for the two samples \( X_1, \ldots, X_{n_1} \) and \( Y_1, \ldots, Y_{n_2} \) is defined by

\[
(3.1) \quad U_{n_1, n_2} = \frac{1}{\binom{n_1}{k_1} \binom{n_2}{k_2} \cdot \binom{n_1}{k_1} \binom{n_2}{k_2}} \sum_{\sigma} f(X_{\sigma_1}, \ldots, X_{\sigma_{k_1}}; Y_{\sigma_1}, \ldots, Y_{\sigma_{k_2}})
\]

11
where the summation is over all \( \binom{n_1}{k_1} \) combinations of \( k_1 \) \( X \) variables and all \( \binom{n_2}{k_2} \) combinations of \( k_2 \) \( Y \) variables.

The sample sizes will be defined by \( n_1 = \lambda_n n \) and \( n_2 = (1-\lambda_n)n \). It will be assumed \( \lambda_n n \) and \( (1-\lambda_n)n \) are non-decreasing functions of \( n \) and \( \lambda_n \) tends to a constant \( \lambda \) as \( n \to +\infty \).

Define

\[
\mu = E(f(X_1, \ldots, X_{k_1}; Y_1, \ldots, Y_{k_2})) ,
\]

\[
f_{10}(x) = E(f(X_1, \ldots, X_{k_1}; Y_1, \ldots, Y_{k_2})|X_1 = x) ,
\]

\[
(3.2) \quad f_{01}(y) = E(f(X_1, \ldots, X_{k_1}; Y_1, \ldots, Y_{k_2})|Y_1 = y) ,
\]

\[
\sigma_{10}^2 = \text{Var}(f_{10}(X_1)) ,
\]

\[
\sigma_{01}^2 = \text{Var}(f_{01}(Y_1)) .
\]

**Theorem 2.** Let \( X_N(t), 0 \leq t \leq 1, \) be defined by

\[
X_N(t) = 0 , \quad 0 \leq t \leq \frac{k_1+k_2-1}{N} ,
\]

\[
X_{N'}^N = \frac{\binom{n}{k_1} \binom{n}{k_2} (1-\lambda_n)n - \mu}{\binom{N}{k_1} \binom{k_1}{1} \sigma_{10}^2 + \frac{k_2}{1-\lambda} \binom{N}{k_2} \binom{k_2}{2} \sigma_{01}^2}^{1/2} , \quad n = k_1+k_2, \ldots, N ,
\]

\[
(3.3) \quad X_N(t) = X_N^N + N(t - \frac{n}{N}) (X_N^N(n+1) - X_N^N(n)) , \quad \frac{n}{N} < t < \frac{n+1}{N} ,
\]

12
where \( U_{n_1, n_2}, k_1, k_2, \mu, \sigma_{10}^2, \sigma_{01}^2 \) are defined in (3.1), (3.2). If \( E(f^2) < +\infty, \lambda_n, (1-\lambda_n)n \) are non-decreasing functions of \( n \), and \( \lambda_n \rightarrow \lambda, 0 < \lambda < 1 \), as \( n \rightarrow +\infty \), then \( X_n(\cdot) \) converges weakly in the uniform topology on \( C[0,1] \) to a Wiener process \( X(\cdot) \) with \( E(X(t)) = 0, E(X^2(t)) = t \) as \( N \rightarrow +\infty \).

**Proof.** The proof is omitted because it is a straightforward duplication of the proof of Theorem 1. Relation (2.4) is replaced by

\[
(3.4) \quad U_{\lambda_n n, (1-\lambda_n)n} - \mu = \frac{k_1}{\lambda_n} \sum_{i=1}^{\lambda_n n} (f_{10}(X_i) - \mu) + \frac{k_2}{(1-\lambda_n)n} \sum_{i=1}^{(1-\lambda_n)n} (f_{01}(Y_j) - \mu) + R_n.
\]

The asymptotic multivariate normality for the sum of the first two terms on the right in (3.4) is immediate. The remainder term \( R_n \) converges in probability to zero, and a reverse martingale argument again shows that \( R_n \) does not disturb the tightness of the measures.

Theorem 2 extends a theorem of Lehmann [6] from asymptotic normality to weak convergence. Both Theorems 1 and 2 could be generalized to the case where the \( X_i \) (and \( Y_j \)) are random vectors.
References


