A Distribution-Free Test for Comparing the
Associations Between Two Pairs of Variables in a
Multivariate Sample

by

C. E. Davis and Dana Quade
Department of Biostatistics
University of North Carolina School of Public Health

Institute of Statistics Mimeo Series No. 575

March 1968
Errata

Page 7

...; the number of observations discordant with the $i$th observation is the
sum of the cells which are simultaneously above and to the right of the
$(j,k)$ cell or below and to the left.

Page 12

where the summation (c) extends over $1 \leq c_1 < c_2 < \ldots < c_m \leq n$.

Let $s_k^2 = \frac{1}{n-1} \sum_{i=1}^{n} (V_k^{(i)} - U_k)^2$; $k = 1,2$. Sen shows that $z_k = \frac{\sqrt{n}(U_k - \theta_k)}{ms_k}$ ...

Page 15

Since $U_1$ converges in probability to $\theta_1$ and $U_2$ converges in probability to
$\theta_2$, ...

As the asymptotic joint distribution of $\sqrt{n}(U_1 - \theta_1)$ and $\sqrt{n}(U_2 - \theta_2)$ is bivariate
normal ...

where $\rho_{kk} = \int \cdots \int \phi(X_1, \ldots, X_m) dF(X_2) \ldots dF(X_m) \int dF(X_1) - \theta_k^2$; $k = 1,2$; ...

$$r = \frac{s_{12}}{s_1 s_2}$$
A Distribution-Free Test for Comparing the Associations Between Two Pairs of Variables in a Multivariate Sample

by

C. E. Davis and Dana Quade
Department of Biostatistics
University of North Carolina School of Public Health

Summary

Consider a random sample of observations on \( \mathbf{A} = (X, Y, W, Z) \). Let \( N_C^{(1)}, N_D^{(1)}, \) and \( N_T^{(1)} \) be the number of pairs of observations which are concordant, discordant, and tied, respectively, for variables \( X \) and \( Y \); and \( N_C^{(2)}, N_D^{(2)}, \) and \( N_T^{(2)} \) be the corresponding numbers for \( W \) and \( Z \). A measure of the association between \( X \) and \( Y \) is \( t_1 = \frac{N_C^{(1)} - N_D^{(1)}}{N_C^{(1)} + N_D^{(1)} + N_T^{(1)}} \) and similarly, \( t_2 \), for \( W \) and \( Z \). In this paper it is shown how, using \( t_1 \) and \( t_2 \), the association between \( X \) and \( Y \) may be compared with that between \( W \) and \( Z \), if \( n \) is sufficiently large.

1. Introduction

Let \( \mathbf{A}_i = (X_i, Y_i, W_i, Z_i), 1 \leq i \leq n \), be a random sample of observations on a random variable \( \mathbf{A} \), where each component of \( \mathbf{A} \) is at least ordinal. Based on this sample it is desired to test a hypothesis concerning, or place a confidence interval on, the difference between the correlation of variables \( X \) and \( Y \), and that of variables \( W \) and \( Z \), when no assumption is made concerning the form of the underlying distribution. It is the purpose of this paper to indicate how this can be carried out for large samples.

The measure of association used is the unconditional index of order association or Kendall's tau-a, hereafter referred to as the
unconditional index, which we now define. A pair of bivariate observations 
(X1, Y1) and (X2, Y2) is said to be:

  concordant if X1 < X2, Y1 < Y2 or X1 > X2, Y1 > Y2  
  discordant if X1 < X2, Y1 > Y2 or X1 > X2, Y1 < Y2  
  tied       if X1 = X2 or Y1 = Y2 or both.

Let \(p_C\), \(p_D\), and \(p_T\) be the probabilities that a randomly chosen pair of 
observations is concordant, discordant, or tied, respectively. The uncondi-
tional index between variables X and Y is then

\[
\tau(X,Y) = \tau = \frac{p_C - p_D}{p_C + p_D + p_T} = p_C - p_D .
\]

Note that \(\tau(X,Y) = \tau(Y,X) = -\tau(-X,Y)\); that \(\tau^2(X,Y) \leq 1\); and that \(\tau(X,Y) = 0\) 
if X and Y are independent. Thus \(\tau\) is a standardized measure much like a 
correlation coefficient.

If \(\tau_1\) and \(\tau_2\) are the unconditional indices for \((X,Y)\) and \((W,Z)\), 
respectively, we shall show how to test hypotheses of the form \(H: \tau_1 - \tau_2 = 5\).

The procedure is outlined in Section 2; Section 3 consists of an 
example; a discussion of the method is in Section 4; and theoretical con-
considerations are presented in the Appendix.

2. A Test for the Difference in Associations

For \(1 \leq i \leq n\), let \(Q_{1i}\) be the number of observations which are 
concordant with the \(i\)th observation, less the number which are discordant, 
for variables \((X,Y)\); and let \(Q_{2i}\) be the corresponding difference for \((W,Z)\). 
The estimates of \(\tau_1\) and \(\tau_2\) are then

\[
t_1 = \frac{1}{n(n-1)} \sum_{i=1}^{n} Q_{1i}
\]
and
\[ t_2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} Q_{2i}, \]
respectively. We note that if \( N_{Tk}, k = 1, 2 \) is the number of tied observations, then
\[ |t_k| \leq 1 - \frac{N_{Tk}}{\binom{n}{2}}. \]

For \( k = 1, 2 \), define
\[ s_k^2 = \frac{4}{n^2(n-1)^3} \left[ n \sum_{i=1}^{n} Q_{k1}^2 - \left( \sum_{i=1}^{n} Q_{ki} \right)^2 \right]. \]

Then the random variables
\[ z_k = \frac{t_k - \tau_k}{s_k}, \quad k = 1, 2 \]
are asymptotically standard normal random variables. Hence, for large \( n \), we can construct confidence intervals for \( \tau_k \) or test hypotheses such as
\[ H: \tau_k = \tau_k^*, \quad \text{where } \tau_k^* \text{ is some constant}. \]

Now let \( Q_i = Q_{1i} - Q_{2i} \) and
\[ s^2 = \frac{4}{n^2(n-1)^3} \left[ n \sum_{i=1}^{n} Q_{1i}^2 - \left( \sum_{i=1}^{n} Q_{i1} \right)^2 \right]. \]

Then for large \( n \),
\[ z = \frac{(t_1 - t_2) - (\tau_1 - \tau_2)}{s} \]
is also a standard normal random variable. This enables us to test
\[ H: \tau_1 - \tau_2 = \delta, \quad \text{or to construct a confidence interval for } (\tau_1 - \tau_2). \]

An estimate of the correlation between \( t_1 \) and \( t_2 \) is
\[ r = \frac{s_{12}}{s_1 s_2}, \quad \text{where} \]
\[ s_{12} = \frac{4}{n^2(n-1)^3} \left[ n \sum_{i=1}^{n} Q_{1i} Q_{2i} - \left( \sum_{i=1}^{n} Q_{i1} \right) \left( \sum_{i=1}^{n} Q_{2i} \right) \right]. \]
Note that $s^2 = s_1^2 + s_2^2 - 2s_{12}$.  

That these results are valid can be seen by noting that $\tau_1$ and $\tau_2$ are $U$-statistics and appealing to the results of the appendix.¹ To this end, identify the general random variable $X$ used there with $A$, $\tau_1$ and $\tau_2$ with $\theta_1$ and $\theta_2$ respectively, and note that $m = 2$. Define

$$\varphi_1(\hat{A}_i, \hat{A}_j) = \begin{cases} 1 & \text{if } (X_i, Y_i) \text{ is concordant with } (X_j, Y_j) \\ 0 & \text{if } (X_i, Y_i) \text{ is tied with } (X_j, Y_j) \\ -1 & \text{if } (X_i, Y_i) \text{ is discordant with } (X_j, Y_j) \end{cases},$$

and

$$\varphi_2(\hat{A}_i, \hat{A}_j) = \begin{cases} 1 & \text{if } (W_i, Z_i) \text{ is concordant with } (W_j, Z_j) \\ 0 & \text{if } (W_i, Z_i) \text{ is tied with } (W_j, Z_j) \\ -1 & \text{if } (W_i, Z_i) \text{ is discordant with } (W_j, Z_j) \end{cases}.$$ 

$\varphi_k(\hat{A}_i, \hat{A}_j)$, $k = 1, 2$ is then a symmetric, unbiased estimator of $\tau_k$. Hence the $U$-statistic for estimating $\tau_k$ is

$$U_k = \frac{2}{n(n-1)} \sum_{i<j} \varphi_k(\hat{A}_i, \hat{A}_j) = \frac{1}{n(n-1)} \sum_{i=1}^{n} Q_{ki} = \tau_k ; \quad k = 1, 2.$$ 

The components of $U_k$ are

$$V_k^{(i)} = \frac{1}{n-1} \sum_{j=1 \atop j \neq i}^{n} \varphi_k(\hat{A}_i, \hat{A}_j) = \frac{1}{n-1} Q_{ki} ; \quad 1 \leq i \leq n ; \quad k = 1, 2.$$ 

¹Those readers who are not mathematically inclined may skip the remainder of this section and the Appendix.
Moreover, $v^{(i)} = v_1^{(i)} - v_2^{(i)}$ and a little algebra then verifies the formulae for $s_k^2$, $s^2$, and $s_{12}^2$. Hence the results outlined above are seen to be special cases of the results in the Appendix.

3. Example

As an example we use a portion of the data collected in a dental survey of North Carolina, described in detail by Fulton et al. [1]. The observations consisted of noting whether a tooth was healthy (i.e. non-carious), carious, or missing. We shall use the data for the second molar, for white males aged 15-19. As past experience has indicated that for this age group most missing teeth are missing due to previous caries, we shall consider a missing tooth to be a carious tooth.

Denoting a non-carious tooth by 0 and a carious tooth by X, there are 16 possible outcomes for an individual (see Table 1); e.g. (outcome #9) the second molar on the upper left quarter of the mouth may be carious while the remaining 3 second molars are healthy.

| Table 1 |
| Possible Outcomes for an Individual |
|---|---|---|---|---|
| 1. | 0|0 | 5. | 0|X | 9. | X|0 | 13. | X|X |
| 2. | 0|0 | 6. | 0|X | 10. | X|0 | 14. | X|X |
| 3. | 0|X | 7. | 0|X | 11. | X|0 | 15. | X|X |

The frequency distribution for the second molar is given in Table 2. There are two questions of interest:
(i) Is there an association in caries rates between the second molars on the left and those on the right of the mouth? the maxillary and mandibular halves?

(ii) If the answer to both parts of (i) is yes, is there a difference in the degree of association?

Table 2

<table>
<thead>
<tr>
<th>Observation</th>
<th>Frequency</th>
<th>$Q_{1i}$</th>
<th>$Q_{2i}$</th>
<th>$Q_i$</th>
<th>$fQ_{1i}$</th>
<th>$fQ_{2i}$</th>
<th>$fQ_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>146</td>
<td>119</td>
<td>27</td>
<td>3504</td>
<td>2856</td>
<td>648</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>80</td>
<td>75</td>
<td>5</td>
<td>320</td>
<td>300</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>72</td>
<td>75</td>
<td>-3</td>
<td>360</td>
<td>375</td>
<td>-15</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>89</td>
<td>-44</td>
<td>133</td>
<td>2314</td>
<td>-1144</td>
<td>3458</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>80</td>
<td>50</td>
<td>30</td>
<td>240</td>
<td>150</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>-66</td>
<td>62</td>
<td>-128</td>
<td>-264</td>
<td>248</td>
<td>-512</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>89</td>
<td>62</td>
<td>27</td>
<td>445</td>
<td>310</td>
<td>135</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>17</td>
<td>12</td>
<td>5</td>
<td>221</td>
<td>156</td>
<td>65</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>72</td>
<td>50</td>
<td>22</td>
<td>432</td>
<td>300</td>
<td>132</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>89</td>
<td>62</td>
<td>27</td>
<td>445</td>
<td>310</td>
<td>135</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>74</td>
<td>62</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>-3</td>
<td>63</td>
<td>84</td>
<td>-21</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>89</td>
<td>-69</td>
<td>158</td>
<td>445</td>
<td>-345</td>
<td>790</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
<td>17</td>
<td>-13</td>
<td>30</td>
<td>153</td>
<td>-117</td>
<td>270</td>
</tr>
<tr>
<td>15</td>
<td>7</td>
<td>9</td>
<td>-13</td>
<td>22</td>
<td>63</td>
<td>-91</td>
<td>154</td>
</tr>
<tr>
<td>16</td>
<td>69</td>
<td>83</td>
<td>56</td>
<td>27</td>
<td>5727</td>
<td>3864</td>
<td>1863</td>
</tr>
</tbody>
</table>

Total 192 14,468 7256 7212

Let $\tau_1$ and $\tau_2$ be the population unconditional indices for the left-right and maxillary-mandibular comparisons, respectively. A computational aid, which also indicates an important application of the method, is to put the data into two contingency tables as in Table 3.
Table 3
Number of Carious Teeth

(a) 
Left Side

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>24</td>
<td>11</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>41</td>
<td>14</td>
<td>62</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>22</td>
<td>69</td>
<td>95</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>74</td>
<td>83</td>
<td>192</td>
</tr>
</tbody>
</table>

(b) 
Maxillary

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>24</td>
<td>9</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>14</td>
<td>16</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>20</td>
<td>69</td>
<td>115</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>43</td>
<td>90</td>
<td>192</td>
</tr>
</tbody>
</table>

From these $Q_{1i}$ and $Q_{2i}$ can easily be determined. Suppose the $i$th observation is in the $j$th row and the $k$th column, i.e. the $(j,k)$ cell. The number of observations which are concordant with the $i$th observation is the sum of all cells which are simultaneously above and to the left of the $(j,k)$ cell or below and to the right; the number of observations discordant with the $i$th observation is the sum of the cells which are simultaneously below and to the right of the $(j,k)$ cell or above and to the left. For example, consider an observation in the second row and second column of the first contingency table. There are 41 such observations. The number of observations concordant with each of these is $69 + 24 = 93$, while the number discordant is $0 + 4 = 4$. Hence $Q_{1i}$ for each of these 41 observations is $93 - 4 = 89$. $Q_{2i}$ can be computed in a similar manner after noting in which cell of the second contingency table the $i$th observation lies.
If the data cannot be put into contingency tables, no such easy method of counting the number of concordant and discordant pairs is available.

From the computations indicated in Table 2 we have

\[ t_1 = \frac{14,468}{36,672} = 0.39452 \]

\[ s_1 = 0.032266 \]

\[ t_2 = \frac{7,256}{36,672} = 0.19786 \]

and

\[ s_2 = 0.038225 \]

Thus

\[ z_1 = \frac{0.39452}{0.032266} = 12.23 \]

and

\[ z_2 = \frac{0.19786}{0.038225} = 5.17 \]

Hence we reject both hypotheses \( H: \tau_k = 0, k = 1, 2 \); and conclude that there is an association in caries rates for teeth on the left and right side of the mouth and similarly for the maxillary and mandibular halves. Further computation yields

\[ t_1 - t_2 = 0.19666 \]

and

\[ s = 0.037570 \]

Thus for the test of \( H: \tau_1 = \tau_2 \), we have

\[ z = \frac{0.19666}{0.03575} = 5.23 \]

and we conclude that there is a difference between \( \tau_1 \) and \( \tau_2 \). The estimate
of the correlation between $t_1$ and $t_2$ is

$$r = \frac{0.000545}{(0.032266)(0.038225)} = 0.44201 .$$

4. Discussion

As was suggested in the example of Section 3, this method is applicable when a set of multivariate observations is arranged in two or more contingency tables, as is often the case with sample survey data. Since we can estimate the correlation between $t_1$ and $t_2$, we have an indication of the degree of dependency of the two tests of association.

Even if only one contingency table is being analyzed the unconditional index will be preferable to the usual $\chi^2$ test, if an estimate of the degree of association is desired. Various coefficients of contingency, which are functions of $\chi^2$, have been used for this in the past. However, these functions of $\chi^2$ are difficult, if not impossible, to interpret. It should be kept in mind that the test of no order association and the test of independence based on $\chi^2$ are not tests of the same hypothesis. If the variables are independent, the population unconditional index will be zero, but the converse is not true since variables which have no order association may still have some other kind of association.

As the small sample distribution of $(t_1 - t_2)$ is not known, a question to which we have no answer is how large $n$ should be in order that the asymptotic results be applicable. Further research will be required before this question can be answered.

The computations involved in carrying out the tests for association given here can be tedious for even moderately large $n$, however the methods are readily adaptable to an electronic computer.
Since the only restriction on the components of $A$ is that all are at least ordinal, we may take $W = X$, in which case the procedure given here allows us to compare the association of $X$ and $Y$ with that of $X$ and $Z$.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Dr. H. A. Tyroler for making the data of the example available to them.
5. References


Appendix: The Asymptotic Distribution of the Difference Between Two U-Statistics

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with cumulative distribution function $F(X)$. Consider two estimable parameters $\theta_1$ and $\theta_2$ each of degree $m$ ($\leq n$) and let $\varphi_1 (X_1, \ldots, X_m)$ and $\varphi_2 (X_1, \ldots, X_m)$ be their symmetric unbiased estimators. The U-statistics for estimating $\theta_1$ and $\theta_2$ are then

$$U_1 = \left( \frac{n}{m} \right)^{\frac{1}{m}} \sum_{(c)} \varphi_1 (X_{c_1}, X_{c_2}, \ldots, X_{c_m})$$

and

$$U_2 = \left( \frac{n}{m} \right)^{\frac{1}{m}} \sum_{(c)} \varphi_2 (X_{c_1}, X_{c_2}, \ldots, X_{c_m})$$

where the summation $(c)$ extends over $1 \leq c_1 < c_2 < \ldots < c_m \leq n$.

Define the components of the U-statistics as

$$V_{k}^{(i)} = \left( \frac{n-1}{m-1} \right)^{\frac{1}{m-1}} \sum_{(ci)} \varphi_k (X_{i}, X_{c_1}, \ldots, X_{c_{m-1}}) ; 1 \leq i \leq n, k = 1, 2,$$

where the summation $(ci)$ extends over $1 \leq c_1 < c_2 < \ldots < c_{m-1} \leq n$, except that $c_j = i$ must not occur for any $j$. Note that for $k = 1, 2$

$$U_k = \frac{1}{n} \sum_{i=1}^{n} V_k^{(i)}.$$

Let $s_k^2 = \frac{1}{n-1} \sum_{i=1}^{n} (V_k^{(i)} - V_k)^2 ; k = 1, 2$. Sen [2] shows that $z_k = \frac{\sqrt{n} (V_k - \theta_k)}{ms_k}$ converges in distribution to the standard normal distribution.

Since $\theta_1$ and $\theta_2$ are each estimable of degree $m$, it follows that $\theta = \theta_1 - \theta_2$ is estimable of degree less than or equal to $m$. Assume that the degree of $\theta$ is exactly $m$. A symmetric unbiased estimator of $\theta$ is
\[ \phi(X_1, \ldots, X_m) = \phi_1(X_1, \ldots, X_m) - \phi_2(X_1, \ldots, X_m) \]

and the U-statistic for estimating \( \theta \) is

\[ U = \binom{n}{m}^{-1} \sum_{(c)} \phi(X_{c_1}, \ldots, X_{c_m}) \]

\[ = U_1 - U_2 \]

Define

\[ v^{(i)} = \binom{n-1}{m-1}^{-1} \sum_{(c)} \phi(X_{c_1}, X_{c_1}, \ldots, X_{c_{m-1}}); \quad 1 \leq i \leq n \]

\[ = V_1^{(i)} - V_2^{(i)} \]

and

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (v^{(i)} - U)^2 \]

It then follows immediately from Sen's result that

\[ z = \frac{\sqrt{n}(U-\theta)}{ms} = \frac{\sqrt{n}[(U_1-U_2) - (\theta_1 - \theta_2)]}{ms} \]

is asymptotically a standard normal random variable.

We now prove the following result: If the integral

\[ \rho_{12} = \int \cdots \int \phi_1(X_1, \ldots, X_m) dF(X_2) \cdots dF(X_m) \]

\[ \cdot \int \cdots \int \phi_2(X_1, \ldots, X_m) dF(X_2) \cdots dF(X_m) dF(X_1) - \theta_1 \theta_2 \]

is convergent, then \( s_{12} = \frac{1}{n-1} \sum_{i=1}^{n} (V_1^{(i)} - U_1)(V_2^{(i)} - U_2) \) converges in probability to \( \rho_{12} \).
Proof:

Note that \( s_{12} \) can be written as:

\[
   s_{12} = \frac{1}{n-1} \sum_{i=1}^{n} (V_1^{(i)} - \theta_1)(V_2^{(i)} - \theta_2) - \frac{n}{n-1} (U_1 - \theta_1)(U_2 - \theta_2).
\]

Define

\[
   \phi_k^{(i)} = \int \cdots \int \phi_k(X_1, X_{c_1}, \ldots, X_{c_{m-1}}) dF(X_{c_1}) \cdots dF(X_{c_{m-1}}); \quad 1 \leq i \leq n, \quad k = 1, 2
\]

where \( c_j = i \) must not occur and consider

\[
   \frac{1}{n-1} \sum_{i=1}^{n} (V_1^{(i)} - \theta_1)(V_2^{(i)} - \theta_2) = \frac{1}{n-1} \left[ \sum_{i=1}^{n} (V_1^{(i)} - \phi_1^{(i)})(V_2^{(i)} - \phi_2^{(i)}) + \sum_{i=1}^{n} (V_1^{(i)} - \theta_1)(\phi_2^{(i)} - \theta_2) \right]
\]

\[
   + \sum_{i=1}^{n} (\phi_1^{(i)} - \theta_1)(V_2^{(i)} - \phi_2^{(i)}) + \sum_{i=1}^{n} (\phi_1^{(i)} - \theta_1)(\phi_2^{(i)} - \theta_2) \right].
\]

Now by the Cauchy-Schwartz inequality

\[
   \left| \frac{1}{n-1} \sum_{i=1}^{n} (V_1^{(i)} - \phi_1^{(i)})(V_2^{(i)} - \phi_2^{(i)}) \right| \leq \left[ \frac{1}{n-1} \sum_{i=1}^{n} (V_1^{(i)} - \phi_1^{(i)})^2 \right]^{\frac{1}{2}} \left[ \frac{1}{n-1} \sum_{i=1}^{n} (V_2^{(i)} - \phi_2^{(i)})^2 \right]^{\frac{1}{2}}.
\]

Sen shows that \( \frac{1}{n-1} \sum_{i=1}^{n} (V_k^{(i)} - \phi_k^{(i)})^2 \) converges in probability to zero; \( k = 1, 2 \).

Hence it follows that

\[
   \frac{1}{n-1} \sum_{i=1}^{n} (V_1^{(i)} - \phi_1^{(i)})(V_2^{(i)} - \phi_2^{(i)})
\]

converges in probability to zero. A similar argument proves that both

\[
   \frac{1}{n-1} \sum_{i=1}^{n} (V_1^{(i)} - \theta_1)(\phi_2^{(i)} - \theta_2)
\]

and
\[
\frac{1}{n-1} \sum_{i=1}^{n} (\phi_1^{(i)} - \theta_1)(\phi_2^{(i)} - \theta_2)
\]

converge in probability to zero. By definition, \(\phi_k^{(i)}\) and \(\phi_k^{(j)}\), \(i \neq j\); \(k = 1, 2\); are independent and identically distributed, thus by Khintchine's weak law of large numbers it follows that

\[
\frac{1}{n-1} \sum_{i=1}^{n} (\phi_1^{(i)} - \theta_1)(\phi_2^{(i)} - \theta_2)
\]

converges in probability to \(E[(\phi_1^{(i)} - \theta_1)(\phi_2^{(i)} - \theta_2)] = \rho_{12}\). Thus we see that

\[
\frac{1}{n-1} \sum_{i=1}^{n} (V_1^{(i)} - \theta_1)(V_2^{(i)} - \theta_2)
\]

converges in probability to \(\rho_{12}\). Since \(U_1\) converges in probability to \(\theta_1\) and \(U_2\) converges in probability \(\theta_2\), \(\frac{n}{n-1}(U_1 - \theta_1)(U_2 - \theta_2)\) converges in probability to zero. Hence \(s_{12}\) converges in probability to \(\rho_{12}\).

As the asymptotic joint distribution of \(\sqrt{n}(U_1 - \theta_1)\) and \(\sqrt{n}(U_2 - \theta_2)\) in bivariate normal with variance matrix

\[
\begin{bmatrix}
\rho_{11} & \rho_{12} \\
\rho_{12} & \rho_{22}
\end{bmatrix}
\]

where \(\rho_{kk} = \int [\cdots \int \phi_k(x_1, \ldots, x_m) dF(x_2) \cdots dF(x_m) dF(x_1) - \theta_k^2]; k = 1, 2\); it follows that if \(n\) is sufficiently large a valid estimate of the correlation between \(U_1\) and \(U_2\) is

\[
\gamma = \frac{s_{12}}{s_1 s_2}
\]