NON PARAMETRIC TESTS OF INDEPENDENCE

by

Paul DEHEUVELS

Université Paris VI & E.P.H.E.

Key words: Non parametric methods, tests of independence, distribution free procedures, rank statistics


Summary: If \((X_{n(1)}, \ldots, X_{n(p)})\) is for \(n = 1, 2, \ldots\), an i.i.d. sequence, with \(F_n(x_1, \ldots, x_p)\) as its empirical c.d.f. with margins \(F^{(i)}_n\), \(1 \leq j \leq p\), the empirical dependence function \(D_n\) is the c.d.f. of a probability distribution with uniform margins on \([0,1]^p\), and such that \(F_n(x_1, \ldots, x_p) = D_n(F^{(1)}_n(x_1), \ldots, F^{(p)}_n(x_p))\). We show in this paper that \(D_n(u_1, \ldots, u_p)\) is asymptotically normal for \(p \geq 3\) and show the weak convergence of \(n^{1/2} \{D_n(u_1, \ldots, u_p) - E(D_n(u_1, \ldots, u_p))\}\) toward a limiting gaussian process of which we derive the covariance function in the independence case. These results extend the bivariate case studied in [3] and [5].

Some applications are given to tests of independence, including in particular Kendall's \(\tau\) and Spearman's \(\rho\). We give a tabulation of our test \(T_n(4)\), developed in [3], for \(n = 11 - 30\), extending the tabulation for \(n = 3 - 10\) obtained in [4].

(*) 7 avenue du Château
92340 BOURG-LA-REINE
France
1.- GENERAL RESULTS ON EMPIRICAL DEPENDENCE FUNCTIONS:

Let \( \{X_n(1), \ldots, X_n(p)\} \), \( n = 1, 2, \ldots, p \) be a sequence of i.i.d. \( p \)-variate random vectors, with cumulative distribution function (c.d.f.) \( F(x_1, \ldots, x_p) \), and marginal c.d.f. \( F(1), \ldots, F(p) \). If we assume \( F(1), \ldots, F(p) \) to be continuous, the order and rank statistics associated to the first \( n \) observations of the sequence are uniquely defined w.p. 1 as:

\[
\forall 1 \leq k \leq p, \quad X_{1,n}(k) < X_{2,n}(k) < \ldots < X_{n,n}(k),
\]

\[
\forall 1 \leq i \leq n, \quad X_i(k) = X_{r_{i,n},n}(k).
\]

Let also \( F_n(x_1, \ldots, x_p) \) be the empirical c.d.f. associated to the first \( n \) observations of the sequence, and set likewise \( F_n^{(1)}, \ldots, F_n^{(p)} \) to be its marginal c.d.f.s.

We will use the name of dependence function for any c.d.f. of a probability measure on \([0,1]^P\) with uniform margins. We will define the dependence function of \( F \) by

\[
F(x_1, \ldots, x_p) = D(F(1)(x_1), \ldots, F(p)(x_p)), \quad \forall x_1, \ldots, x_p
\]

and the empirical dependence function of \( F \) by

\[
F_n(x_1, \ldots, x_p) = D_n(F_n^{(1)}(x_1), \ldots, F_n^{(p)}(x_p)), \quad \text{where } x_1, \ldots, x_p \text{ are continuity points of } F_n^{(1)}, \ldots, F_n^{(p)}.
\]

The existence of \( D_n \) was proved in [3], and some of its properties given in [4], [5]. It can be proved that \( D_n \) is uniquely defined on \( I_n = \{(i_1/n, \ldots, i_p/n)\} \), \( 0 \leq i_j \leq n, \ 1 \leq j \leq p \) by

\[
D_n(i_1/n, \ldots, i_p/n) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p} \{ I(i_j - r_{i,n}(j)) \},
\]

where \( I \) stands for the indicator function of \([0, \infty) \), \( I(u) = 1 \) if \( u \geq 0 \), \( 0 \) if \( u < 0 \).

For a probability measure giving \( D_n \) as an admissible dependence function, satisfying (2), we can take, for instance:

\[
\nu_n = \frac{1}{n} \sum_{i=1}^{n} U_n(r_{i,n}(1), \ldots, r_{i,n}(p)), \quad \text{where } U_n(1, \ldots, i_p) \text{ is a uniform probability measure on } \prod_{j=1}^{p} [(i_j - 1)/n, i_j/n]. \quad \text{We can assume, unless otherwise specified that it is the case. It will be also useful to introduce :}
\]

\[
\nu_n^* = \frac{1}{n} \sum_{i=1}^{n} \delta(r_{i,n}(1)/n, \ldots, r_{i,n}(p)/n), \quad \text{where } \delta(x) \text{ stands for the Dirac measure at point } x.
\]