SPEARMAN'S *RHO* AND KENDALL'S *TAU* FOR MULTIVARIATE DATA SETS

HAMANI EL MAACHE AND YVES LEPAGE
Université de Montréal

A class of U-statistics matrices is introduced to obtain the distribution of the matrices of the Spearman and Kendall correlation coefficients between the components of a random vector. These results are used to construct nonparametric tests of independence between two sets of variables based on three measures of multivariate relationship. The tests are illustrated by an example and a simulation study is performed to compare the tests based on Kendall's matrix with those based on Spearman's matrix.

1. Introduction

Let $F(x) = F(x^{[1]}, x^{[2]})$ be the continuous c.d.f. (cumulative distribution function) of a random vector $X = (X^{[1]}, X^{[2]})'$, where $x = (x^{(1)}, \ldots, x^{(m)})' \in \mathbb{R}^m$, $m \geq 2$, $x^{[1]} \in \mathbb{R}^p$, $x^{[2]} \in \mathbb{R}^q$ $(p + q = m)$ and $F^{[k]}(x^{[k]})$ $(k = 1, 2)$ denote the marginal c.d.f. of $X^{[k]}$. The objective of this paper is to detect deviation from the null hypothesis of independence that is, to test H_0: $F(x) = F^{[1]}(x^{[1]})F^{[2]}(x^{[2]})$ against appropriate classes of alternatives $\mathrm{H}_{1:n}$. A nonparametric approach to this problem was explored by Puri, Sen and Gokhale (1970) who defined a class of association parameters based on componentwise ranking. The statistic they proposed uses the elements of the matrix $D_n = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$, where

$$(1.1) \qquad D_n^{(i,j)} = \frac{1}{n} \sum_{\alpha=1}^{n} J\left(\frac{R_\alpha^{(i)}}{n}\right) J\left(\frac{R_\alpha^{(j)}}{n}\right), \quad i, j = 1, \ldots, m.$$

Here, $R_\alpha^{(i)}$ is the rank of $X_\alpha^{(i)}$, that denote the ith coordinate of the vector X_{α},; the symbol α will run over the sample (from X) with $\alpha = 1, \ldots, n$ and J represents an arbitrary standardized score function. Puri, Sen and Gokhale (1970) established the joint asymptotic multivariate normality of the vector formed by the elements of D_n.

When the score function is $J(u) = J_0(u) = \sqrt{12}(u - \frac{1}{2})$, then

$$(1.2) \quad D_n^{(i,j)} = \frac{12}{n(n^2-1)} \sum_{\alpha=1}^{n} \left(R_\alpha^{(i)} - \frac{n+1}{2}\right)\left(R_\alpha^{(j)} - \frac{n+1}{2}\right), \qquad i, j = 1, \ldots, m,$$

AMS subject classifications: Primary 62H20; Secondary 62H15..

Keywords and phrases: U-statistics, Spearman *rho* correlation coefficient, Kendall *tau* correlation coefficient, measures of multivariate association, rank correlation, tests of independence..