Distance covariance independence test and it's application for Markov chains

Y. N. Yesbay[1]

[1]Kazakhstan branch of Lomonosov Moscow State University, Nur-Sultan, Kazakhstan, yesbay185@gmail.com

In the 2009 article by Székely and Rizzo [1] a new approach for testing of multivariate independence was suggested. It is based on distance covariance $\nu(X, Y)$, a measure of dependence between random vectors, that equals zero if and only if X and Y are independent.

Given a sample of independent identically distributed random vectors from the joint distribution $\mathbf{P}_{XY}$

$(\mathbf{X}, \mathbf{Y}) = \{(X_1, Y_1), (X_2, Y_2), \dots (X_n, Y_n)\}$,

where $X \in \mathbf{R}^k$ and $Y \in \mathbf{R}^l$, we want to test the null hypothesis $H_0 : \mathbf{P}_{XY} = \mathbf{P}_X \mathbf{P}_Y$.

For the sample distance correlation, denoted as $\nu_n(\mathbf{X}, \mathbf{Y})$, and a certain statistic $T = T_n(\mathbf{X}, \mathbf{Y})$ defined in [1], page 1245, there is following convergence in distribution:

$$\frac{n\nu_n^2}{T} \xrightarrow{D} Q = \sum_{k=1}^{\infty} \lambda_k Z_k^2, \tag{1}$$

where λ_k are non-negative constants depending on the distribution of $\mathbf{P}_{XY}$, and Z_k are independent standard normal random variables.

For all $0 < \alpha < 0,215$ the following inequality holds:

$$P(Q > y_{1-\alpha}) \leq \alpha, \tag{2}$$

where $y_{1-\alpha}$ is the $(1-\alpha)$ quantile of $\chi^2(1)$ distribution.

Thus, a test rejecting H_0 when $\frac{n\nu_n^2}{T} \geq y_{1-\alpha}$, has an asymptotic significance level at most α.

Assume that the random sequence $(X_1, Y_1), (X_2, Y_2), \dots$ is an irreducible recurrent Markov chain with a finite state space.

We adapted the described above test for Markov chains by proving, that the convergence (1) and inequality (2) hold under these assumptions.

We also proved the convergence (1) in the basic case, where (X_i, Y_i) and (X_j, Y_j) are independent for $i \neq j$, as neither proof nor reference to a proof of it are provided in Székely and Rizzo [1]. Moreover we showed, that the non-negative constants λ_k of the quadratic form Q are the eigenvalues of an integral transform T_K, defined as

$$[T_K x](t, s) = \int K(t, s, u, v) x(u, v) du dv,$$

where $K(t, s, u, v)$ is the covariance function of a certain complex-valued zero mean Gaussian random process U indexed with two real numbers.

$$K(t, s, u, v) = EU(t, s)U(u, v)$$

References

1. Székely, G.J. and Rizzo, M.L.(2009). Brownian distance covariance. *Ann. Appl. Stat.* 3(4): 1236-1265.

2. Van der Vaart, A.W. and Wellner, J.A.(1996). Weak Convergence and Empirical Processes. Springer Science+Business Media, New York.