Statistics and Probability Letters 112 (2016) 20–25

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Simplicial bivariate tests for randomness

Germain Van Bever

The Open University, MK7 6AA, Milton Keynes, United Kingdom

ARTICLE INFO

Article history:
Received 12 January 2016
Received in revised form 18 January 2016
Accepted 18 January 2016
Available online 28 January 2016

Keywords:
Angular symmetry
Bivariate runs
Invariance
Simplex

ABSTRACT

This paper introduces a simplex-based extension of the concept of runs to the bivariate setup which allows to test for randomness under the null hypothesis of angularly symmetric distributions. The statistic's null limiting distribution is derived and Monte Carlo studies evaluate the test performances.

1. Introduction

The notion of randomness is central in serial contexts. Indeed, detecting the presence or absence of structure in chronological observations is crucial and often the first step in the analysis of such data. Runs tests for randomness have been around for decades. These nonparametric procedures can either be based on the length of the longest run or – as in Wald and Wolfowitz (1940) – on the number of runs. The classical univariate description of the latter is the following.

For some fixed $\theta \in \mathbb{R}$, the hypothesis $\mathcal{H}_\theta^{(n)}$ under testing states that the observations $X_1, \ldots, X_n$ are mutually independent random variables satisfying $P(X_t < \theta) = P(X_t \leq \theta) = 1/2$. Denoting I_A the indicator function of the set A, the number of runs in the sequence $X_1, \ldots, X_n$ is defined as

$$R_\theta^{(n)} = 1 + \sum_{t=2}^{n} I_{[U_{t,\theta} \neq U_{t-1,\theta}]}, \tag{1}$$

where $U_{t,\theta} = I_{[X_t < \theta]} - I_{[X_t > \theta]}$. That is, $R_\theta^{(n)}$ counts the number of *runs* (i.e. blocks) of consecutive 1's or -1's in the sequence of signs of the residual $X_t - \theta$.

Denoting E_θ the expectation under $\mathcal{H}_\theta^{(n)}$, classical Wald–Wolfowitz-type tests then reject the null hypothesis of randomness when $|R_\theta^{(n)} - E_\theta(R_\theta^{(n)})|$ is too large. This is sensible since large (resp. small) values of $R_\theta^{(n)}$ indicate negative (resp. positive) serial dependence. Exact tests are derived easily since, under $\mathcal{H}_\theta^{(n)}$, $R_\theta^{(n)} \sim 1 + Bin\left((n-1), 1/2\right)$. Moreover, rewriting (1) using $I_{[U_{t,\theta} \neq U_{t-1,\theta}]} = (1 - U_{t,\theta}U_{t-1,\theta})/2$, it is seen that the quantity

$$r_\theta^{(n)} = \frac{-2\left(R_\theta^{(n)} - E_\theta(R_\theta^{(n)})\right)}{\sqrt{n-1}} := \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} U_{t,\theta}U_{t-1,\theta} \tag{2}$$

E-mail address: germain.van-bever@open.ac.uk.

http://dx.doi.org/10.1016/j.spl.2016.01.013

is asymptotically standard normal. The asymptotic runs test therefore rejects $\mathcal{H}_\theta^{(n)}$ at asymptotic level α as soon as $(r_\theta^{(n)})^2 > \chi^2_{1,1-\alpha}$, where $\chi^2_{k,1-\alpha}$ denotes the upper α quantile of the χ^2_k distribution.

Both tests require the common median to be specified, though. If the median θ is unknown, only the latter test can be implemented, then rejecting the null of randomness for large value of $(r_{\hat\theta}^{(n)})^2$, where $\hat\theta$ is an appropriate estimator of θ.

Many subsequent generalizations of (2) are available in the literature. For example, improving on the fact that Wald–Wolfowitz-type tests cannot detect serial dependence at a lag greater than one, Dufour et al. (1998) introduced *generalized runs tests*, rejecting the null of randomness $\mathcal{H}_\theta^{(n)}$ at asymptotic level α as soon as, for H a fixed positive integer,

$$C_{H,\theta}^{(n)} := \sum_{h=1}^{H}\left(r_{h,\theta}^{(n)}\right)^2 := \sum_{h=1}^{H}\left(\frac{1}{\sqrt{n-h}}\sum_{t=h+1}^{n} U_{t,\theta}U_{t-h,\theta}\right)^2 > \chi^2_{H,1-\alpha}.$$

In parallel, a natural interest for the definition of multivariate runs tests grew as multivariate time series became a part of the daily practice. Considering $\mathbf{X}_1, \ldots, \mathbf{X}_n$, a series of d-variate observations, Paindaveine (2009) provides two extensions of (2) involving the *standardized spatial signs*

$$\mathbf{U}_{t,\boldsymbol\theta}(\mathbf{V}) = \frac{\mathbf{V}^{-1/2}\left(\mathbf{X}_t - \boldsymbol\theta\right)}{\|\mathbf{V}^{-1/2}\left(\mathbf{X}_t - \boldsymbol\theta\right)\|}, \tag{3}$$

where $\mathbf{V}$ is an appropriate shape matrix (see, for example, Taskinen et al., 2005). Both the elliptical Marden runs

$$r_{\boldsymbol\theta}^{(n)e}(\mathbf{V}) = \sum_{t=2}^{n} \mathbf{U}_{t,\boldsymbol\theta}(\mathbf{V})'\mathbf{U}_{t-1,\boldsymbol\theta}(\mathbf{V}) \tag{4}$$

and the (matrix-valued) full-rank runs (that are also of an elliptical nature; we adopt here the denominations from Paindaveine, 2009)

$$r_{\boldsymbol\theta}^{(n)f}(\mathbf{V}) = \sum_{t=2}^{n} \mathbf{U}_{t,\boldsymbol\theta}(\mathbf{V})\mathbf{U}_{t-1,\boldsymbol\theta}(\mathbf{V})' \tag{5}$$

– the former itself generalizing the Marden (1999) spherical runs (obtained by imposing $\mathbf{V} = \mathbf{I}_d$, the $d \times d$ identity matrix) – provide *affine-invariant* statistics to test for multivariate randomness.

The null hypothesis considered in these multivariate extensions is that of *elliptical directions* (see Randles, 1989, 2000). Although this is indeed a multivariate version of $\mathcal{H}_\theta^{(n)}$, this still puts restrictions on the signs $\mathbf{U}_{t,\theta}(\mathbf{V})$ as, for example, their support need to be the whole hypersphere S^{d-1}. Moreover, even in the $\boldsymbol\theta$-specified case, such tests typically require the estimation of the shape parameter $\mathbf{V}$.

In this paper, we introduce an affine-invariant bivariate runs test that (i) is valid – in the sense that it meets the level constraint – for any angularly symmetric distribution of the random vectors $\mathbf{X}_t$ and (ii) does not require the estimation of any extra parameter. This is achieved by generalizing (1) based on the following remark: In $R_\theta^{(n)}$, a new run is added when residuals $(X_t - \theta)$ and $(X_{t-1} - \theta)$ have different signs, that is, when the origin is contained in the interval (or, equivalently in the univariate case, simplex) with endpoints $U_{t,\theta}$ and $U_{t-1,\theta}$. In the bivariate setup, denoting from now on $\mathbf{U}_{t,\boldsymbol\theta} := \mathbf{U}_{t,\boldsymbol\theta}(\mathbf{I}_2)$, this suggests defining a runs statistic as the number of simplices with vertices $\mathbf{U}_{t,\boldsymbol\theta}$, $\mathbf{U}_{t-1,\boldsymbol\theta}$ and $\mathbf{U}_{t-2,\boldsymbol\theta}$ that contain the origin. As we show below, the resulting statistic allows to test for bivariate, angularly symmetric, randomness and enjoys many properties of its univariate counterpart. For a related approach, see Dyckerhoff et al. (2015).

The rest of the paper is organized as follows. Section 2 describes the proposed test statistic. The null hypothesis of randomness is defined in Section 2.1. The simplicial bivariate runs (and the resulting test) are then defined in Section 2.2, while their invariance properties are discussed in Section 2.3. In Section 3, the null asymptotic distribution of the proposed test is derived. Section 4 is dedicated to Monte Carlo experiments. Final comments are made in Section 5, while the Appendix collects the proofs.

2. Bivariate tests

This section describes the null hypothesis used in this paper and the extension of the univariate classical runs test statistic to the bivariate setup.

2.1. The null hypothesis

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^2$ be a sequence of bivariate random vectors. The null hypothesis we consider in this paper is a multivariate extension of $\mathcal{H}_\theta^{(n)}$ described in the introduction. More precisely, for any 2-variate $\boldsymbol\theta \in \mathbb{R}^2$, we denote by $\mathcal{H}_{\boldsymbol\theta,ang}^{(n)}$ the hypothesis under which (i) $P(\mathbf{X}_t = \boldsymbol\theta) = 0$, for all $t = 1, \ldots, n$, and (ii) the vectors $\mathbf{U}_{t,\boldsymbol\theta}$ are independent and identically distributed with a common continuous and centrally symmetric distribution on the unit circle S^1.