Runs tests

Abstract: Given a sequence of two or more types of symbols, a run is defined as a succession of–one or more–identical symbols which are followed and preceded by a different symbol (or by no symbol at all). Runs tests are based on the length of the longest run or on the total number of runs. They figure among the oldest nonparametric procedures, and can be used in various setups. We review the different definitions of runs, their development over the last one hundred years, describe the domains of application of runs tests, and discuss their extension to the multivariate case.

Keywords: Distribution-freeness; Multi-sample problems; Quality control; Robustness; Runs; Tests of randomness; Tests of symmetry; Time series

A famous teaching experiment

Given a sequence of two or more types of symbols, a run is defined as a maximal subsequence of identical symbols. For instance, the sequence HHTTTHTTTTTHH of heads (H) and tails (T) contains five runs and the shortest and longest runs have length one and five, respectively. A well-known teaching experiment of T. Varga nicely illustrates the usefulness of such runs; see, e.g., [1].

The experiment may be described as follows. A class of school children is divided into two sections by the teacher.

(i) Each child in the first section is given a coin and is asked to write down on a slip of paper the head and tail sequence obtained from tossing the coin two hundred times.

(ii) Each child from the second section is rather asked to write down on a slip of paper a "random" head and tail sequence of length two hundred (children from the second section may not use coins).

Collecting these slips of paper, the teacher then considers them one at a time and tries to classify them as originating from the first section or the second one.

One possible strategy is to classify as random any sequence presenting a run whose length is large enough—is strictly larger than four, say. This is justified by the fact that, when asked to produce artificially such a random sequence, children are usually afraid of writing down long runs, while long runs actually appear quite often (the probability that the length of the longest run is strictly larger than four is above 95%; see Figure 1 for an illustration). This explains why this classification rule tends to work well in practice.

From this teaching experiment, one may guess that runs are of interest for statistical inference. Run-based tests of hypotheses—or simply *runs tests*—are based on the size of the longest run or, more often, on the number of runs in a sequence. Runs tests figure among the oldest nonparametric procedures, as evidenced by, e.g., [2]. They exhibit many advantages : they remain valid (in the sense that they meet the level constraint) under very mild assumptions, they are often distribution-free, they are very robust against possible outliers, they do not require any moment conditions, and they are easy to calculate and implement. Their main drawback, at least in the standard univariate case, is that they are poorly efficient compared to parametric Gaussian procedures, unless observations originate from a distribution with very heavy tails.

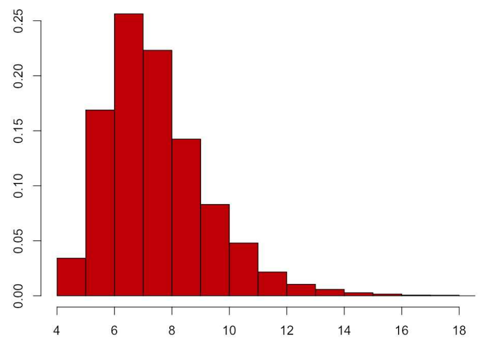

Figure 1 Histogram of the distribution of the longest run in a collection of $10,000$ head and tail sequences of length 200.

History of runs in a snapshot

A considerable amount of research on the theory of runs and runs tests was achieved around 1940, by [3–8] or [9]; see also the extensive bibliography in [10]. The popularity of runs tests has continued to increase until the late 1950's, where it seems to reach a peak: [10] contains 24 references for the period between 1955 and 1959. Notable contributions in the less active times are [11, 12], which contain several distributional results on the total number of runs and on the length of the longest run. This area of research has become active again during the 1980's and 1990's (see, e.g., [13]). The traditional combinatorial approach was then partially replaced by a Markov-chain approach, as described in [14] or [15]. A precise account on the development of the distribution theory of runs can be found in [16]. Modern research on runs tests mainly focuses on the definition of multivariate runs tests. Besides the slightly older proposal by [17], recent works in this direction are [18–20] and [21].

Domains of application

The high popularity of runs tests is explained by their many nice properties (see the bottom of Page 1) and the fact that they can deal with both qualitative and quantitative data. Even more important, arguably, is the fact they happen to be useful in many different areas and setups. [22, 23] and [24] describe economic and econometric applications of runs tests. [9, 25, 26], and [27] apply the concept of runs to quality control charts, [28] to contingency tables, [29] to regression problems, and [30] to multi-sample problems.

The famous paper Wald and Wolfowitz (1940), that deals with multi-sample problems, proposes a runs test to detect whether two samples are drawn from the same population. This work deserves some special attention, as it is actually the basis of most of the subsequent developments. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two independent samples of i.i.d. univariate observations. In order to test whether the common distribution of the X_i's coincides with the common distribution of the Y_j's, Wald and Wolfowitz' statistic is the number of runs in the sequence obtained by replacing X's with ones and Y's with zeros in the order statistic of the pooled sample. Clearly, the null hypothesis that both parent distributions do coincide is to be rejected for small values of this runs statistic.

Runs tests of the Wald and Wolfowitz type have also been used to address the problem of testing randomness of a series of observations, which is one of the most important problems in time series analysis. This problem is arguably the most natural field of applications of runs tests, be it in testing against trend ([8],

[24] or [29]), first-order serial dependence ([31–33] or [34]), higher-order Markov dependence ([34]) or in the study of comovements between two series ([35, 36]). Gibbons [11, 12] provides some insight into the uses of runs in this context. To be more specific, classical runs tests typically reject the null hypothesis of randomness, in a sample of serial observations X_t, $t = 1, \ldots, n$, when the number of runs in the sequence $S_t(\theta_n)$, $t = 1, \ldots, n$ is either too small or too big; here, $S_t(a) := \mathrm{Sign}(X_t - a)$ stands for the sign of X_t with respect to a (that is, the sign of $X_t - a$), and θ_n denotes the sample median of the n observations. Clearly, small (resp., large) numbers of runs indicate positive (resp., negative) serial dependence, hence are incompatible with the null of randomness. Under the assumption that the observations share the same population median and are almost surely different from this common median value, the runs statistic rewrites

$$R^{(n)}_{\mathrm{rand}} = 1 + \sum_{t=2}^{n} \frac{1 - S_t(\theta_n) S_{t-1}(\theta_n)}{2}, \tag{1}$$

which shows that the classical runs test is based on a first-order (sign) autocorrelation quantity, hence might fail to detect higher-order dependence. In order to improve on this, Dufour, Hallin and Mizera [37] propose generalized runs tests for randomness, that are based on higher-order (sign) autocorrelations. Their contribution can be considered as an important step towards a general methodology for the analysis of time series with nonhomogeneous innovations (the runs test for randomness does not assume that observations are identically distributed, but only that they share the same population median).

In the late 1980's, runs have also proved useful for the problem of testing symmetry of a univariate (absolutely continuous) distribution, as is shown by Cohen and Menjoge [38] and McWilliams [39]; both works independently propose a common runs test that is usually referred to as McWilliams test in the literature. Assuming that a sample of i.i.d. observations $X_1, \ldots, X_n$ is available, McWilliams' test rejects the null that the parent distribution is symmetric about a fixed centre (θ, say) for small values of

$$R^{(n)}_{\mathrm{symm}} = 1 + \sum_{i=2}^{n} \frac{1 - S_{A_i}(\theta) S_{A_{i-1}}(\theta)}{2}, \tag{2}$$

where $A_1, \ldots, A_n$ are defined through $|X_{A_1} - \theta| < |X_{A_2} - \theta| < \ldots < |X_{A_n} - \theta|$ (absolute continuity almost surely provides this strict ordering). In other words, the runs test statistic is computed first by permuting the observations according to the distances from the null symmetry centre θ, then by recording the number of runs in the resulting sequence of signs. Clearly, a small value of $R^{(n)}_{\mathrm{symm}}$ indicates different tail weights on both sides of θ, hence is incompatible with the null of symmetry about θ. Cohen and Menjoge [38] derive the asymptotic null distribution of $R^{(n)}_{\mathrm{symm}}$, while McWilliams [39] makes use of results from [3] to prove that $R^{(n)}_{\mathrm{symm}} - 1$ is binomial with parameters $n - 1$ and $1/2$. This runs test for symmetry is shown to be universally consistent in [40]. Finally, weighted and conditional versions of this test are proposed in [41] and in [42], respectively.

Extension to the multivariate setup

To describe the recent extensions of the concept of runs to the multivariate setup, we focus on the multivariate versions of two of the three testing problems considered above, namely (i) the problem of testing for randomness in a multivariate series and (ii) the problem of testing symmetry of a multivariate distribution; we refer to [17] for the construction of multivariate runs tests in the third problem, namely the two-sample location problem.

(i) Consider the problem of testing the null hypothesis of randomness in the k-dimensional series of observations $\mathbf{X}_t$, $t = 1, \ldots, n$. Assuming that the observations share a common distribution that is spherically symmetric about $\boldsymbol{\theta}$, Marden [18] proposes—by analogy with (1)—to reject the null of randomness for either small or large values of

$$R^{(n)}_{\rm rand} = 1 + \sum_{t=2}^{n} \frac{1 - \mathbf{S}'_t(\boldsymbol{\theta})\mathbf{S}_{t-1}(\boldsymbol{\theta})}{2}, \tag{3}$$

where the so-called *spatial sign* $\mathbf{S}_t(\mathbf{a})$ of $\mathbf{X}_t$ with respect to $\mathbf{a}$ is defined as $(\mathbf{X}_t - \mathbf{a})/\|\mathbf{X}_t - \mathbf{a}\|$ if $\mathbf{X}_t \neq \mathbf{a}$ and as $\mathbf{0}$ otherwise. Paindaveine [19] further considers "full-rank" runs tests obtained by replacing inner products with outer products in a normalized version of (3); an appropriate matrix norm is then needed to decide how "large" the resulting runs statistic is. In order to make the test valid under a weaker symmetry assumption, standardized multivariate runs are defined there, either by using *standardized spatial signs* (see, e.g., [43]) or Randles' *interdirections* (see [44]). Quite surprisingly, relative efficiencies with respect to Gaussian parametric competitors improve as the dimension k increases.

(ii) Parallel to the above, Marden [18] also proposes to test the null that the common distribution of $\mathbf{X}_i$, $i = 1, \ldots, n$ is spherically symmetric (about $\boldsymbol{\theta}$) by means of a multivariate runs statistic of the form

$$R^{(n)}_{\rm symm} = 1 + \sum_{i=2}^{n} \frac{1 - \mathbf{S}'_{A_i}(\boldsymbol{\theta})\mathbf{S}_{A_{i-1}}(\boldsymbol{\theta})}{2}, \tag{4}$$

where $A_1, \ldots, A_n$ are defined through $\|\mathbf{X}_{A_1} - \boldsymbol{\theta}\| < \|\mathbf{X}_{A_2} - \boldsymbol{\theta}\| < \ldots < \|\mathbf{X}_{A_n} - \boldsymbol{\theta}\|$; compare with (2). More recently, Ley [20] defines runs tests for the broader null hypothesis of central symmetry (about $\boldsymbol{\theta}$) by introducing multivariate runs that are more nonparametric in nature. Denoting by $\text{Simplex}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ the convex hull of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ in $\mathbb{R}^k$ and by $D^{(n)}_{\boldsymbol{\theta}}(\mathbf{x})$ the depth of $\mathbf{x}$ with respect to the empirical distribution of the sample $\mathbf{X}_1, \ldots, \mathbf{X}_n, 2\boldsymbol{\theta} - \mathbf{X}_1, \ldots, 2\boldsymbol{\theta} - \mathbf{X}_n$, these runs are defined as

$$R^{(n)}_{\rm symm} := 1 + \sum_{i=k+1}^{n} \mathbb{I}_{[\boldsymbol{\theta} \in \text{Simplex}(\mathbf{X}_{A_i}, \mathbf{X}_{A_{i-1}}, \ldots, \mathbf{X}_{A_{i-k}})]},$$

where $A_1, \ldots, A_n$ are now defined through $D^{(n)}_{\boldsymbol{\theta}}(\mathbf{X}_{A_1}) > \ldots > D^{(n)}_{\boldsymbol{\theta}}(\mathbf{X}_{A_n})$; for all classical statistical depth functions (halfspace depth, simplicial depth, simplicial volume depth, etc.; see [45]), this runs statistic reduces to the McWilliams runs statistic in (2) in the univariate case. An extensive study of the resulting tests in the bivariate case is provided in [21].

References

[1] Révész P. & Csörgö M. (1981). *Strong Approximations in Probability and Statistics*, Academic Press.

[2] Fisher R.A. (1926). On random sequence, *Quaterly Journal of the Royal Meteorological Society* 52, 250.

[3] Wishart J. & Hirschfeld H.O. (1936) A theorem concerning the distribution of joins between line segments. *The Journal of the London Mathematics Society* 11, 227–235.

[4] Kermack W.O. & McKendrick A.G. (1937). Tests for randomness in a series of numerical observations, *Proceedings of the Royal Society of Edinburgh* 57, 228–240.

[5] Mood A.M. (1940). The distribution theory of runs, *The Annals of Mathematical Statistics* 11, 367–392.

[6] Olmstead P.S. (1940) Note on theoretical and observed distributions of repetitive occurrences, *The Annals of Mathematical Statistics* 11, 363–366.

[7] Wald A. & Wolfowitz J. (1940). On a test whether two samples are from the same population, *The Annals of Mathematical Statistics* 11, 147–162.

[8] Wallis W.A. & Moore G.H. (1941). A significance test for time series analysis, *Journal of the American Statistical Association* 36, 401–409.

[9] Wolfowitz J. (1943). On the theory of runs with some applications to quality control, *The Annals of Mathematical Statistics* 14, 280–288.

[10] Dufour J.-M., Lepage Y. & Zeidan H. (1982). Nonparametric testing for time series: a bibliography, *Canadian Journal of Statistics* 10, 1–38.

[11] Gibbons J.D. (1971). *Nonparametric statistical inference*, McGraw-Hill, New York.

[12] Gibbons J.D. (1976). *Nonparametric methods for quantitative analysis*, Holt, Rinehart & Winston, New York.

[13] Koutras M.V. & Papastavridis S.G. (1993). On the number of runs and related statistics, *Statistica Sinica* 3, 277–294.

[14] Schwager S.J. (1983). Run probabilities in sequences of Markov-dependent trials, *Journal of the American Statistical Association* 78, 168–175.

[15] Fu J.C. & Koutras M.V. (1994). Distribution theory of runs: a Markov chain approach, *Journal of the American Statistical Association* 89, 1050–1058.

[16] Fu J.C. (1996). Distribution theory of runs and patterns associated with a sequence of multi-state trials, *Statistica Sinica* 6, 957–974.

[17] Friedman J.H. & Rafsky L.C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests, *The Annals of Statistics* 7, 697–717.

[18] Marden J. (1999). Multivariate rank tests, *In Multivariate Analysis, Design of Experiments and Survey Sampling, ed. S. Ghosh, New York: Marcel Dekker*, 401–432.

[19] Paindaveine D. (2009). On multivariate runs tests for randomness, *Journal of the American Statistical Association* 104, 1525–1538.

[20] Ley C. (2010). Univariate and Multivariate Symmetry: Statistical Inference and Distributional Aspects, *Unpublished thesis, Univ. Libre de Bruxelles, Brussels*.

[21] Ley C. & Paindaveine D. (2011). Depth-based runs tests for bivariate central symmetry, ECARES working paper 2011-06.

[22] Karsten K.G. (1926). The Harvard business indexes: a new interpretation, *Journal of the American Statistical Association* 21, 399–418.

[23] Cowles A. & Jones H.E. (1937). Some a posteriori probabilities in stock market action, *Econometrica* 5, 280–294.

[24] Mann H.B. (1945) Nonparametric tests against trend, *Econometrica* 13, 245–259.

[25] Mosteller F. (1941). Note on an application of runs to quality control charts, *The Annals of Mathematical Statistics* 12, 228–232.

[26] Weiler H. (1953). The use of runs to control the mean in quality control, *Journal of the American Statistical Association* 48, 816–825.

[27] Weiler H. (1954). A new type of control chart limits for means, ranges, and sequential runs, *Journal of the American Statistical Association* 49, 298–314.

[28] Moore P.G. (1949). A test for randomness in a sequence of two alternatives involving a 2×2 table, *Biometrika* 36, 305–316.

[29] Stuart A. (1956). The efficiencies of tests of randomness against normal regression, *Journal of the American Statistical Association* 51, 285–287.

[30] Barton D.E. & David F.N. (1957). Multiple runs, *Biometrika* 44, 168–178.

[31] David F.N. (1947). A power function for tests of randomness in a sequence of alternatives, *Biometrika* 34, 335–339.

[32] Goodman L.A. (1958). Simplified runs tests and likelihood ratio tests for Markov chains, *Biometrika* 45, 181–197.

[33] Granger C.W.J. (1963). A quick test for serial correlation suitable for use with non-stationary time series, *Journal of the American Statistical Association* 58, 728–736.

[34] Denny J.L. & Yakowitz S.J. (1978). Admissible run-contingency type tests for independence and Markov dependence, *Journal of the American Statistical Association* 73, 177–181.

[35] Goodman L.A. & Grunfeld Y. (1961). Some nonparametric tests for comovements between time series, *Journal of the American Statistical Association* 56, 11–26.

[36] Yang M.C.K. & Schreckengost J.F. (1981). Difference sign test for comovements between two time series, *Communications in Statistics, Theory and Methods* A10, 355–369.
[37] Dufour J.-M., Hallin M. & Mizera I. (1998). Generalized runs tests for heteroskedastic time series, *Journal of Nonparametric Statistics* 9, 39–86.
[38] Cohen J.P. & Menjoge S.S. (1988). One-sample run tests of symmetry, *Journal of Statistical Planning and Inference* 18, 93–100.
[39] McWilliams T.P. (1990). A distribution-free test for symmetry based on a runs statistic, *Journal of the American Statistical Association* 85, 1130–1133.
[40] Henze N. (1993). On the consistency of a test for symmetry based on a runs statistic, *Journal of Nonparametric Statistics* 3, 195–199.
[41] Modarres R. & Gastwirth J.L. (1996). A modified runs test for symmetry, *Statistics and Probability Letters* 31, 107–112.
[42] Baklizi A. (2003). A conditional distribution-free runs test for symmetry, *Journal of Nonparametric Statistics* 15, 713–718.
[43] Taskinen S., Oja H. & Randles R.H. (2005), Multivariate nonparametric tests of independence, *Journal of the American Statistical Association* 100, 916–925.
[44] Randles R.H. (1989). A distribution-free multivariate sign test based on interdirections, *Journal of the American Statistical Association* 84, 1045–1050.
[45] Zuo Y. & Serfling R.J. (2000). General notions of statistical depth function, *The Annals of Statistics* 28, 461–482.

Christophe Ley
Université Libre de Bruxelles, E.C.A.R.E.S. and Département de Mathématique, Boulevard du Triomphe, Campus Plaine, CP210, B-1050 Brussels, Belgium

Davy Paindaveine
Université Libre de Bruxelles, E.C.A.R.E.S. and Département de Mathématique, Av. F.D. Roosevelt, 50, CP114/04, B-1050 Brussels, Belgium