Non-Randomness in a Sequence of Two Alternatives: II. Runs Test
Author(s): D. E. Barton and F. N. David
Source: *Biometrika,* Vol. 45, No. 1/2 (Jun., 1958), pp. 253-256
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2333062
Accessed: 17-07-2019 09:29 UTC

Non-randomness in a sequence of two alternatives

II. Runs test

By D. E. BARTON and F. N. DAVID
University College London

It is assumed that there exists an infinite population composed of two characteristics in the proportions p and q. A sample of r is randomly drawn from this population, the elements being laid in a line according to the order of drawing. Under the null hypothesis the alternation of the two characteristics along the line will be random. Under the alternate hypothesis it will be supposed that the selector is influenced in the choice of the $(v+1)$st element by his knowledge of the vth element. Such a situation might arise if, confronted with a very large number of photographs of men and women and asked to pick out r in ascending age order, the selector tended to choose a photograph of a women if the one previously selected was also a woman. We have previously described this situation as one of persistence of type.

Let the suffix i denote the ith ranking position and denote the two characteristics by x and y. Then

$$P\{x_i\} = p = 1-q = 1-P\{y_i\} \quad (i = 1, 2, \ldots).$$

If in r drawings there are r_1 x's and r_2 y's the probability distribution of T, the number of runs of both alternatives, is, under the null hypothesis

$$P\{T=2t \mid r_1, r_2, H_0\} = 2\,\frac{{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}\,p^{r_1}q^{r_2}}{{}^{r}C_{r_1}\,p^{r_1}q^{r_2}} = \frac{2\,{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}}{{}^{r}C_{r_1}},$$

and

$$P\{T=2t+1 \mid r_1, r_2, H_0\} = \frac{({}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t}+{}^{r_1-1}C_{t}\,{}^{r_2-1}C_{t-1})\,p^{r_1}q^{r_2}}{{}^{r}C_{r_1}\,p^{r_1}q^{r_2}} = \frac{{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t}+{}^{r_1-1}C_{t}\,{}^{r_2-1}C_{t-1}}{{}^{r}C_{r_1}}.$$

This result is possibly due to Whitworth (*Choice and Chance*, Problems 193 and 194) but was probably known before him.

Under the alternate hypothesis we shall assume that persistence of type can be described by a simple Markoff chain. We suppose the probabilities for the single event are unaltered but that

$$P\{x_i \mid x_{i-1}\} = p+q\theta, \quad P\{y_i \mid x_{i-1}\} = q(1-\theta),$$
$$P\{x_i \mid y_{i-1}\} = p(1-\theta), \quad P\{y_i \mid y_{i-1}\} = q+p\theta \quad (i = 2, 3, \ldots).$$

It will be noticed that $\quad P\{x_i \mid x_{i-1}\} + P\{y_i \mid x_{i-1}\} = 1$

and $\quad P\{x_{i-1}\}\,P\{x_i \mid x_{i-1}\} + P\{y_{i-1}\}\,P\{x_i \mid y_{i-1}\} = p$

as expected. Under these assumptions the distribution of T in the non-null case, i.e. when $\theta \neq 0$, is the equilibrium position of that described by David (1947). We have, writing

$$S = \sum_t \left[\frac{pq(1-\theta)^2}{(p+q\theta)(q+p\theta)}\right]^t {}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}\left[\frac{\theta(1+\theta)\,t+(1-\theta)\,\{rpq+\theta(rp^2+r_2(q-p))\}}{t(1-\theta)\,(p+q\theta)\,(q+p\theta)}\right],$$

that

$$P\{T=2t \mid r_1, r_2, H_1\} = \frac{2}{S}\,{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}\,\frac{1}{1-\theta}\left[\frac{pq(1-\theta)^2}{(p+q\theta)(q+p\theta)}\right]^t,$$

and

$$P\{T=(2t+1) \mid r_1, r_2, H_1\} = \frac{1}{S}\,{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}\,\frac{1}{(p+q\theta)(q+p\theta)} \times\left[2pq(1+\theta)+\theta+\frac{rp(q+p\theta)+r_2(q-p)\,\theta}{t}\right]\left[\frac{pq(1-\theta)^2}{(p+q\theta)(q+p\theta)}\right]^t.$$

Power curves for varying p and θ have already been discussed by David (1947). When $\theta = 0$ the distribution reduces to that for H_0; the bounds for θ are ± 1. When θ is positive the critical region will be the lower tail of the distribution under H_0; when it is negative the upper tail will be appropriate.

The complete description of any sequence is given by the number of elements of either kind and the two compositions of r_1 x's and r_2 y's, together with the information as to the nature of the first observation. When r_1 and r_2 are decomposed into the same number of components (say t) then $T = 2t$ is the number of runs. When the number of components differs by 1 the total number of runs is odd. Let us specify the two compositions by $(a_1, \ldots, a_k)$, $(b_1, \ldots, b_l)$, where

$$r_1 = \Sigma a_i, \quad r_2 = \Sigma b_i \quad \text{and} \quad |\,l-k\,| \leqslant 1.$$

Let α be a characteristic random variable taking the value 1 if the sequence starts with an x and zero otherwise. Various functions of the compositions have been used as test functions for randomness in the sequence, generally conditional on r_1 and r_2.

Now the probability of any sequence of events (each of which must be either x or y) of specification $[r_1, r_2, \{a_i\}, \{b_i\}, \alpha]$, under the null hypothesis of a sequence of r independent trials is $p^{r_1}q^{r_2}$ and is independent of the other variables in the specification. This probability is $(\frac{1}{2})^r$ if $p = q$. Under the hypothesis H_1 for persistence of type, but with the additional specification that $p = \frac{1}{2}$, the probability of any given sequence is

$$(\tfrac{1}{2})^r (1-\theta)^{k+l-1} (1+\theta)^{r-k-l}, \tag{1}$$

that is to say it depends just on θ, r, and $k+1$ (the total number of components of r_1 and r_2). We have called $T = k+1$ in the preceding paragraph. It follows that under these circumstances, for a fixed sample size r, we have that T is sufficient for θ.

We note that, summing the probabilities of each sequence over all sequences of the same value of T, and writing $\pi_0(z)$ to denote the probability generating function of T under the null hypothesis

$$P(T; H_1) = \left(\frac{1-\theta}{1+\theta}\right)^T P(T; H_0) \Big/ \pi_0\left(\frac{1-\theta}{1+\theta}\right).$$

It will also be noted that this result is true if we consider the distribution conditional on r_1 and r_2 since these do not enter into (1), and we may therefore consider a subset of the sample space to obtain

$$P(T \mid r_1, r_2; H_1) = \left(\frac{1-\theta}{1+\theta}\right)^T P(T \mid r_1, r_2; H_0) \Big/ \pi_0\left(\frac{1-\theta}{1+\theta} \middle| r_1, r_2\right),$$

where $\pi_0(z \mid r_1, r_2)$ is the null hypothesis probability generating function of T conditional on r_1 and r_2. Assuming θ positive, values of the power function for two different sequences and three different values of θ are given in Table 1. The critical region was made exactly 0·05 in each case, for purposes of comparison, by taking a proportion of a frequency block.

Table 1. *Powers of T, S and b under the alternate hypothesis of dependence*

r	r_1	r_2	θ	1/4	3/5	7/9	0
10	5	5	T S b	0·178 0·126 0·055	0·587 0·363 0·175	0·842 0·593 0·288	0·05 — —
10	6	4	T S	0·178 0·124	0·587 0·362	0·842 0·593	0·05 —

For sequences of reasonable length and r_1 and r_2 not very different it is possible to assume normality for T under H_0. The distribution of under H_1 may, for values of θ not very different from 0, also be assumed normal, but the mean and variance will be different. The moments of T under the alternate hypothesis, when $p = \frac{1}{2}$, can be found by using the same device as in our earlier paper (Barton, David & Mallows (1958)). Let $\kappa_v(\theta)$ and κ_v denote the vth cumulants under the alternate and the null hypothesis respectively. If

$$\delta = \log[(1-\theta)/(1+\theta)]$$

then

$$\kappa_v(\theta) = \kappa_v + \delta . \kappa_{v+1} + \frac{\delta^2}{2}\kappa_{v+2} + \frac{\delta^3}{3!}\kappa_{v+3} + \dots$$

and in particular

$$\kappa_1(\theta) \doteqdot \kappa_1 + \delta . \kappa_2 + \tfrac{1}{2}\delta^2 . \kappa_3 + \tfrac{1}{6}\delta^3 . \kappa_4,$$
$$\kappa_2(\theta) \doteqdot \kappa_2 + \delta . \kappa_3 + \tfrac{1}{2}\delta^2 . \kappa_4,$$
$$\kappa_3(\theta) \doteqdot \kappa_3 + \delta . \kappa_4,$$
$$\kappa_4(\theta) \doteqdot \kappa_4.$$

Since the distribution of T under H_0 is quickly normal with increasing r so that κ_v $(v > 2)$ tends reasonably quickly to zero these expansions should be adequate as regards order for θ small. The factorial moments of T under the null hypothesis are of reasonably succinct algebraic form (Barton & David (1957)), but the central moments and cumulants do not appear to reduce easily. The first four cumulants under H_0 are

$$\kappa_1 = \frac{2r_1r_2}{r}+1, \quad \kappa_2 = \frac{2r_1r_2}{r^{(2)}}\left(\frac{2r_1r_2}{r}-1\right),$$

$$\kappa_3 = \frac{2r_1r_2}{r^{(3)}}\left(\frac{16r_1^2r_2^2}{r^2}-\frac{4r_1r_2(r+3)}{r}+3r\right),$$

$$\kappa_4 = \frac{2r_1r_2}{r^{(4)}}\left(\frac{48(5r-6)\,r_1^3r_2^3}{r^{(2)}.r^2}-\frac{48(2r^2+3r-6)\,r_1^2r_2^2}{r^{(2)}.r} + \frac{2(4r^3+45r^2-37r-18)\,r_1r_2}{r^{(2)}}-(7r^2+13r-6)\right).$$

When $r_1 = r_2$ the above reduce to

$$\kappa_1 = \frac{r+2}{2}, \quad \kappa_2 = \frac{r(r-2)}{4(r-1)}, \quad \kappa_3 = 0, \quad \kappa_4 = \frac{-r(r-2)(r^2-4r+6)}{8(r-1)^2(r-3)},$$

$$\gamma_2 = -2\left(\frac{1}{r}+\frac{1}{(r-2)(r-3)}\right).$$

Table 2. *Bivariate distribution of b and T for a sequence* (7, 3)

b \ T	2	3	4	5	6	7	Total
0	1	2	4	2	1	—	10
1	—	2	8	16	14	10	50
2	—	2	8	16	14	10	50
3	1	2	4	2	1	—	10
Total	2	8	24	36	30	20	120

For $\theta < 0{\cdot}6$ a satisfactory agreement was found between the true mean and variance of T (as found from the calculated probability distribution function) and from the series expansion or the cumulants. This is not, however, true for the $\kappa_3(\theta)$ and the $\kappa_4(\theta)$ of T, and it is clear that for the series expansion to be useful in these two cases higher moments of T in the null case must be calculated. To cut the series short, as has been done in the previous section, will be adequate for $\kappa_3(\theta)$ and $\kappa_4(\theta)$ only when θ is, very approximately, $< 0{\cdot}2$. In this latter case the distribution of T_θ will be approximated to by the normal distribution. For large values of θ the distribution of T is J-shaped and in order to approximate to it by (say) a Pearson curve, the series expansions for $\kappa_3(\theta)$ and $\kappa_4(\theta)$ will need to be extended.

In our previous paper the powers of S, the sum of the ranks of one characteristic, of b, the number of one characteristic below the median of the sequence, and of T, were compared under the same alternate hypothesis. Using the same arguments as we set forward there, we may compare the powers of S and T, and of b and T, under the dependence alternative, using the bivariate distribution. The arrays of the distribution of S for T fixed are weighted by $[(1-\theta)/(1+\theta)]^T$ and then added for T keeping S fixed. The critical region for S under the null hypothesis is the sum of the two tail areas. The power of S to detect $\theta \neq 0$ is given in Table 1.

The joint distribution of T and b may be written down in explicit algebraic form and the bivariate table constructed. For the sake of illustration we given the bivariate distribution for $r_1 = 7$ and $r_2 = 3$. When the total number in the sequence is even it is possible to make a dichotomy at the median and the table is symmetrical about $\mathscr{E}(b)$. It follows that $\mathscr{E}(b \mid T)$ is constant. When r, the number in the sequence, is odd, we make a dichotomy between the Rth and the $(R+1)$st observations ($R = \frac{1}{2}(r-1)$), and the symmetry of the table disappears. The regression of T on b is quadratic under the null hypothesis and may be found either from the joint probability distribution function or from the following considerations. Let T_1 be the number of runs of both characteristics below the point of dichotomy and T_2 the number above. Then

$$T = T_1 + T_2 - \alpha,$$

where $\alpha = 1$ if the last element below the point of dichotomy and the first above are of like characteristics and $\alpha = 0$ otherwise. The conditional expectations we write as

$$\mathscr{E}(T \mid b) = \mathscr{E}(T_1 \mid b) + \mathscr{E}(T_2 \mid b) - \mathscr{E}(\alpha \mid b).$$

It is immediate that

$$\mathscr{E}(T_1 \mid b) = 1 + \frac{2b(R-b)}{R}, \quad \mathscr{E}(T_2 \mid b) = 1 + \frac{2(r_2-b)\,(r-R-(r_2-b))}{r-R},$$

$$\mathscr{E}(\alpha \mid b) = \frac{b(r_2-b)+(R-b)\,(r-R-r_2+b)}{R(r-R)},$$

whence, on substitution,

$$\mathscr{E}(T \mid b) = 1 + 2r_2 - \frac{1}{R(r-R)}\{2b^2(r-1) - b(r-2R+2r_2(2R-1)) + Rr_2(2r_2-1)\}.$$

The maximum value of $\mathscr{E}(T \mid b)$ will be when

$$b = \frac{1}{4(r-1)}(r-2R+2r_2(2R-1)),$$

which for a median dichotomy of an even number of observations reduces to $b = \frac{1}{2}r_2$. The regression of T on b under the alternate hypothesis is approximately, following the argument already set out,

$$\mathscr{E}(T \mid b, H_1) \doteqdot \mathscr{E}(T \mid b, H_0) + \delta \,.\, \sigma^2_{T\mid b}(H_0),$$

and may therefore be calculated once $\sigma^2_{T\mid b}(H_0)$ is found. This second moment will be of the fourth power in b, but since we do not need the regression under H_1 we have not calculated it.

The power of b under the alternate hypothesis can be found in precisely the same way as we have put forward for calculating the power of S. Because b can take few values in a short sequence the power has been found for a (5^2) sequence only. The critical region, the sum of the two tail areas of the b-distribution, has been forced to be 0·05. It will be noticed that b appears to be of little value to detect dependence in a sequence. For the moments of b under the alternate hypothesis we have not been able to find simple expressions. Since the denominator of the general expressions is the probability density function of the number of events in a Markoff chain of two alternatives, and this has not yet been found expressible in terms of elementary functions, it seems unlikely that the moments of b will be tractable.

In a previous paper we put forward an alternate ranking hypothesis for which S was a sufficient test statistic. In this paper we have put forward an alternate hypothesis against which T is a sufficient statistic. The powers of S, T and b have been compared for each model.

REFERENCES

Barton, D. E. & David, F. N. (1957). *Biometrika*, **44**, 168.

Barton, D. E., David, F. N. & Mallows, C. L. (1958). *Biometrika*, **45**, 166.

David, F. N. (1947). *Biometrika*, **34**, 335.

Note on multiple comparisons for adjusted means in the analysis of covariance

By MAX HALPERIN* and S. W. GREENHOUSE†

National Institutes of Health

1. Introduction

The analysis of covariance, in the simple application to a one-way classification, deals with the problem of comparing k-class means of a variable y in the presence of a covariate x. If we observe

$$(y_{i1}, x_{i1}), \quad (y_{i2}, x_{i2}), \quad \ldots, \quad (y_{in_i}, x_{in_i})$$

in the ith class, the usual assumption (using $\mathscr{E}$ to denote expected value of) is that $\mathscr{E}y_{ij} = a_i + bx_{ij}$ as opposed to the customary situation in the analysis of variance where $\mathscr{E}y_{ij} = a_i$. Since the mean of the

* Division of Biologics Standards.

† National Institute of Mental Health.