A BIVARIATE SIGN TEST

By J. L. Hodges, Jr.

University of California, Berkeley

1. Introduction. The sign test has proved to be a very useful means for judging the significance of treatments. Suppose that on each of n individuals (or pairs of individuals) measurements are made under two conditions, for example, before and after treatment (or on a treated and a control subject). Denote the two measurements for the ith individual (or pair of individuals) by x_i and x_i'. We formulate the null hypothesis that x_i and x_i' are identically and independently distributed, but wish to make no assumption concerning relations between the distributions of $x_1, x_2, \cdots, x_n$, nor concerning relations between those of $x_1', x_2', \cdots, x_n'$, save that each set is independent. The alternative to the null hypothesis is that the second measurements x_i' are generally shifted, with respect to the first measurements x_i, in the same direction for all (or most) of the individuals. The test is carried out by counting the number S of the differences $x_i' - x_i$ which have positive signs. Under the null hypothesis, S is binomially distributed with $p = \frac{1}{2}$, assuming there are no cases with $x_i' = x_i$, or that such cases of equality are broken randomly. Under the alternative, S would tend to have large values if the second measurements are generally increased relative to the first, small values if they are decreased. We may then reject for large S, small S, or either, according to the alternative against which we wish the test to have power. The great advantage of the test, aside from its simplicity, is the generality of the conditions under which it is valid.

The present paper proposes a bivariate analog of the two-sided sign test, which can be applied when two quantities are measured on each individual. We now have measurements x_i and y_i in a first circumstance, x_i' and y_i' in a second. Do the $4n$ measurements justify our concluding that the two circumstances differ? The null hypotheses is that the bivariate distribution for (x_i, y_i) is identical with that for (x_i', y_i'), and that these vectors are independent. The alternative of interest is that in the second circumstance the bivariate distribution has been shifted relative to the first, in generally the same direction for all individuals. The direction of this possible shift is, however, unknown.

To illustrate, suppose we measure blood pressure and blood sugar before and after treatment with a new drug on a number of individuals. We wish to know whether the drug influences these quantities, but have no preconceived notion concerning the direction or relative amount of the influence on either quantity, should it exist. The joint distribution of the quantities has an unknown form, and is presumably different in different individuals. The quantities are presumably dependent, but in an unknown way.

If we knew the direction of a possible shift, it would be easy to reduce our problem to the sign test. We could simply project the vectors of differences

Received October 14, 1954.

$(x_i' - x_i, \ y_i' - y_i)$ onto the given direction, and count the number S of projected vectors having the given sense. Our problem arises just because we do not have a given direction, but must derive one from the data.

The idea of the proposed test is to consider all possible directions, and calculate S for each. Let M be the maximum of the values thus calculated. We shall use M as our test statistic, rejecting the null hypothesis if M is too large. That is, we shall judge that a shift has occurred if there is *some* direction in which most of the measurement pairs have shifted; we shall judge that no shift has occurred if the shifts are in various directions with no heavy concentration.

The distribution theory for M under the null hypothesis is worked out in the following sections. Presumably it would be desirable to generalize the proposed test to more than two quantities. The multivariate analog of the statistic M is easily seen, though in more than three dimensions it would be difficult to compute M from the sample, and its null distribution might be troublesome.

2. Reduction to a combinatorial problem. We shall suppose that none of the n vectors $(x_i' - x_i, \ y_i' - y_i)$ lies on the same line, and take the n lines on which these vectors lie as given, with all probability calculations conditional on the given lines. Under the null hypothesis, the distribution of $(x_i' - x_i, \ y_i' - y_i)$ is the same as that of $(x_i - x_i', \ y_i - y_i')$, so that there is probability $\frac{1}{2}$ for the ith vector to be oriented in each of its two possible senses. As the n vectors are independent, we conclude that the 2^n possible orientations of the vectors are all equally likely.

It is easily seen that the value of M for a given set of orientations is independent of the angles between the lines and of the lengths of the vectors. Therefore, for simplicity we may suppose that the lines are equally spaced and the vectors all are of unit length. We imagine a circle on whose circumference $2n$ equally spaced loci are given. We are to distribute n plus signs and n minus signs among these loci, subject to the condition that diametrically opposed signs are opposite in sense. We shall call such an arrangement a *cycle*. We think of a cycle as being rotatable about its center into $2n$ positions, each being itself a cycle. For each position we count the number s of positive signs among the n uppermost signs; m is the maximum of the $2n$ values of s thus obtained. Our problem is to count the cycles having a given value of m.

It is clear that $\frac{1}{2}n \leqq m \leqq n$. We shall denote $n - m$ by k; thus k is the smallest number of minus signs which can be uppermost. The operation of rotation carries one cycle into another, generating equivalence classes of cycles. The largest possible class has $2n$ members. Smaller classes are possible, since there may exist cycles which are carried into themselves by a rotation through r positions, $0 < r < 2n$. However, the smallest such r must be of the form $rc = 2n$ where $3 \leqq c$ is odd (since opposite signs are of opposite sense); thus cycles in an equivalence class smaller than $2n$ will have $k \geqq n/3$. As our interest is primarily in the tail of the distribution (k small), we shall simplify by restricting $k < n/3$, whence we can assume every class to have $2n$ members.

To count the classes, we shall select from each class a representative member, called the *pattern* for the class. This member is the unique one which satisfies two conditions, which can be expressed in terms of the n uppermost signs. These signs are arranged in a semi-circle, and we are particularly interested in the signs forming a consecutive set of fewer than n signs at either extreme of the semi-circle; we call such a set a (right or left) *tail*. The two conditions are:

(a) There is no right tail in which there is a majority of minus signs.

(b) There is no left tail in which the plus signs are not in the majority.

The conditions serve to insure that the pattern has the maximum number m of positive signs uppermost; if it were possible to rotate it into a position with more positive signs uppermost, there would have to be a tail with a majority of minus signs. The conditions also insure that only one pattern is selected from each class; if there were two representatives of the class, (i.e., a cycle appearing in two positions) one of these would contradict condition (a). In general it is not true that every class has a member satisfying these conditions (consider a cycle with alternating signs), but it is true under the restriction $k < n/3$.

3. Counting the patterns. We may obtain a formula for the number $P(n, k)$ of patterns most easily by identifying our problem with the classical problem of gambler's ruin. A pattern, read from right to left, may be interpreted as the record of a penny tossing game in which a gambler with initial capital $h = n - 2k$, playing against an adversary with unit initial capital, is ruined at the nth toss. The probability of such ruin is on the one hand $P(n, k)/2^n$; but on the other hand formulae for it are well known (see, for example, [1], p. 304, problem 6). In fact,

$$P(n, k) = (w_h + w_{3h+2} + w_{5h+4} + \cdots) - (w_{h+2} + w_{3h+4} + w_{5h+6} + \cdots), \tag{1}$$

where $h = n - 2k$, and

$$w_z = \frac{z}{n}\binom{n}{\frac{1}{2}(n - z)}$$

is the number of ways in which a gambler with initial capital z can be ruined at the nth toss when playing against an infinitely rich adversary.

If we take advantage once more of the restriction $k < n/3$, only two terms of (1) differ from zero, so that

$$P(n, k) = \frac{n - 2k}{n}\binom{n}{k} - \frac{n - 2k + 2}{n}\binom{n}{k - 1}. \tag{2}$$

Let $Q(n, k)$ denote the number of patterns with at most k minus signs uppermost. Summing (2) we obtain

$$Q(n, k) = \frac{n - 2k}{n}\binom{n}{k} = \binom{n - 1}{k} - \binom{n - 1}{k - 1}.$$

Recalling that there are $2n$ cycles for each pattern and 2^n cycles in all, while under the null hypothesis these 2^n cycles are equally likely, we find

$$\Pr\{K \leqq k\} = (n - 2k)\binom{n}{k} \Big/ 2^{n-1}$$

The table gives values of $\Pr\{K \leqq k\}$ to $5D$ for $n = 1(1)30$, and $k < n/3$.

Table of $\Pr\{K \leqq k\}$, for $k < n/3$.

n	k = 0	1	2	3	4	5	6	7	8	9
1	1.00000									
2	1.00000									
3	.75000									
4	.50000	1.00000								
5	.31250	.93750								
6	.18750	.75000								
7	.10938	.54688	.98438							
8	.06250	.37500	.87500							
9	.03516	.24609	.70312							
10	.01953	.15625	.52734	.93750						
11	.01074	.09668	.37598	.80566						
12	.00586	.05859	.25781	.64453						
13	.00317	.03491	.17139	.48877	.87280					
14	.00171	.02051	.11108	.35547	.73315					
15	.00092	.01190	.07050	.24994	.58319					
16	.00049	.00684	.04395	.17090	.44434	.79980				
17	.00026	.00389	.02698	.11414	.32684	.66095				
18	.00014	.00220	.01634	.07471	.23346	.52295				
19	.00007	.00123	.00978	.04805	.16264	.39922	.72450			
20	.00004	.00069	.00580	.03044	.11089	.29572	.59143			
21	.00002	.00038	.00340	.01903	.07420	.21347	.46575			
22	.00001	.00021	.00198	.01175	.04883	.15068	.35578	.65057		
23	.00001	.00012	.00115	.00718	.03167	.10429	.26474	.52605		
24	0	.00006	.00066	.00434	.02027	.07094	.19254	.41259		
25		.00003	.00038	.00260	.01282	.04750	.13723	.31517	.58020	
26		.00002	.00021	.00155	.00802	.03137	.09606	.23525	.46559	
27		.00001	.00012	.00092	.00497	.02045	.06616	.17202	.36390	
28		.00001	.00007	.00054	.00305	.01318	.04491	.12350	.27789	.51490
29		0	.00004	.00031	.00186	.00841	.03008	.08722	.20786	.41055
30			.00002	.00018	.00112	.00531	.01991	.06067	.15263	.31987

Although $\Pr\{K < n/3\}$ tends to 0 as $n \to \infty$, it does not fall below 5 percent until $n = 72$, or below 1 percent until $n = 102$. If the test proves useful, it may be desirable to consider the distribution of K for $k \geqq n/3$, where the results are likely to be less simple and neat.

REFERENCE

[1] WILLIAM FELLER, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York, 1950.

ON THE CONVERGENCE OF EMPIRIC DISTRIBUTION FUNCTIONS[1]

BY J. R. BLUM

Indiana University

1. Summary. Let μ be a probability measure on the Borel sets of k-dimensional Euclidean space E_k. Let $\{X_n\}$, $n = 1, 2, \cdots$, be a sequence of k-dimensional independent random vectors, distributed according to μ. For each $n = 1, 2, \cdots$ let μ_n be the empiric distribution function corresponding to $X_1, \cdots, X_n$, i.e., for every Borel set $A \ \varepsilon \ E_k$, we define $\mu_n(A)$ to be the proportion of observations among $X_1, \cdots, X_n$ which fall in A.

Let $\mathfrak{A}$ be the class of Borel sets in E_k defined below. The object of this paper is to prove that $P\{\lim_{n\to\infty} \sup_{A \varepsilon \mathfrak{A}} |\mu_n(A) - \mu(A)| = 0\} = 1$.

2. Introduction. Let $F(x)$ be a distribution function on the real line and let $\{X_n\}$, $n = 1, 2, \cdots$, be a sequence of independent random variables distributed according to F. For each $n = 1, 2, \cdots$ let $F_n(x)$ be the empiric distribution function corresponding to $X_1, \cdots, X_n$. The well-known theorem of Glivenko-Cantelli (see, e.g., Fréchet [1]) states that

$$P\{\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0\} = 1.$$

Fortet and Mourier [2] have proved several theorems on the convergence of empiric distribution functions in a separable metric space E. In particular, they show that if E is a Euclidean space and μ is a probability measure on E which is absolutely continuous with respect to Lebesgue measure, then

$$P\{\lim_{n\to\infty} \sup_{A \varepsilon \mathfrak{A}} |\mu_n(A) - \mu(A)| = 0\} = 1, \tag{2.1}$$

where $\mathfrak{A}$ is the collection of open half-spaces in E. Wolfowitz [3] proved that (2.1) holds without any assumptions on μ. In this note we prove that if μ is absolutely continuous with respect to Lebesgue measure, then (2.1) holds for a considerably more general class of sets.

To avoid repetition we shall assume from now on that every set considered is

Received October 29, 1954.

[1] Work done with the support of the Office of Ordnance Research, U.S. Army.