DISTRIBUTION OF SAMPLE ARRANGEMENTS FOR RUNS UP AND DOWN

By P. S. Olmstead

Bell Telephone Laboratories, Inc.

1. Summary. Using the notation of Levene and Wolfowitz [1], a new recursion formula is used to give the exact distribution of arrangements of n numbers, no two alike, with runs up or down of length p or more. These are tabled for n and p through $n = 14$. An exact solution is given for $p \geq n/2$. The average and variance determined by Levene and Wolfowitz are presented in a simplified form. The fraction of arrangements of n numbers with runs of length p or more are presented for the exact distributions, for the limiting Poisson Exponential, and for an extrapolation from the exact distributions. Agreement among the tables is discussed.

2. Introduction. Assume that

$$x_1, x_2, \cdots x_n$$

represent a series of repetitive measurements. In engineering work, experience has shown that, when the values of these measurements exhibit changes in level, trends, cycles, etc., it is usually indicative of the presence of findable causes. In general, the engineer becomes more confident that a findable cause exists for a change in level, a trend, or a cycle, when the change is large, the trend is long, or the cycle is regular.

On the basis of this experience, the engineer selects particular measures of change in level, length of trend, etc., to guide him in deciding when it is profitable to look for a cause. Having selected the measure, he is interested in knowing how often he may have to look for a cause that does not exist. One such measure is the length of the longest run up or down in a sample of n values. The chart in Figure 1, based on the analysis given here, applies when no two values are alike and indicates the fraction of all nonidentical arrangements that have runs up or down of length p or more.

Attention is directed to the distribution of sample arrangements that have at least one run up or down of length p or more. The distribution and the variances and covariances for lengths of runs up and down are given by Levene and Wolfowitz [1]. In addition, Wolfowitz [2] has shown that the limiting distribution for a particular length of run up or down is a Poisson Exponential.

The notation of Levene and Wolfowitz [1] will be used. Thus, let $a_1, a_2, \cdots, a_n$ be n numbers, no two alike, and let the sequence $S = (h_1, h_2, \cdots, h_n)$ be any permutation of $a_1, a_2, \cdots, a_n$, where S is to be considered a chance variable, and each of the $n!$ permutations of $a_1, a_2, \cdots, a_n$ is assigned the same probability. Consider the derived sequence R whose ith element is the sign ($+$ or $-$) of $h_{i+1} - h_i$, $(i = 1, 2, \cdots, n - 1)$. A sequence of p consecutive $+$ signs immediately preceded by a $-$ sign is called a run up of length p or more; a sequence of p consecutive $-$ signs immediately preceded by a $+$ sign is called a run down of length p or more. When such a run is both immediately preceded and immediately followed by an unlike sign, it is a run of length exactly p. The distribution of arrangements with at least one run up or down of length p or more is considered under five specific headings:

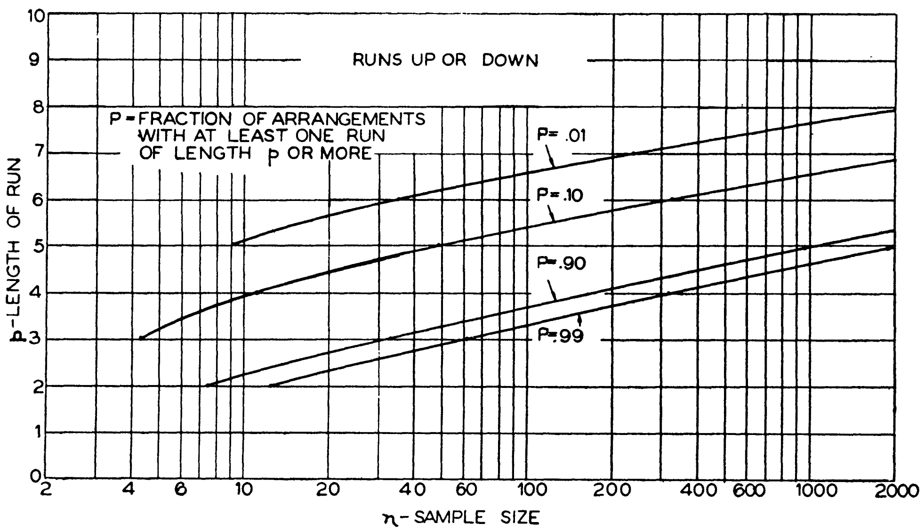

FIG. 1

1. An exact numerical solution for n small, i.e., computations have been completed up to and including $n = 14$.
2. An exact solution for $p \geq \frac{n}{2}$.
3. A limiting solution for $\frac{(p+1)!}{n}$ = constant.
4. An extrapolation from n small.
5. Constant probability relationships.

3. Solution for n small. Starting with a single number, a_1, a second number, $a_2 > a_1$, may be placed before or after it to obtain the two independent arrangements of one run of length exactly 1. A third number, $a_3 > a_2 > a_1$, may be placed before, between, or after the preceding pair to obtain two independent arrangements of one run of length exactly 2 and four of two runs of length exactly 1. Continuing this process it is seen that, on the assumption that the distribution of independent arrangements for $(n - 1)$ numbers, $a_1 < a_2 < a_3 < \cdots < a_{n-1}$, is known, the distribution of independent arrangements for n numbers, $a_1 < a_2 < a_3 < \cdots < a_n$, can be found by using the following recursion formula:

$$
\begin{aligned}
&F_n[r_{n-1}, r_{n-2}, \cdots, r_h, \cdots, r_i, \cdots, r_j, \cdots, r_1] \\
&= \sum_{i=2}^{n-1} (r_{i-1} + 1)F_{n-1}[r_{n-2}, r_{n-3}, \cdots, (r_i - 1), (r_{i-1} + 1), \cdots r_1] \\
&+ 2F_{n-1}[r_{n-2}, r_{n-3}, \cdots, (r_1 - 1)] \\
(1)\quad &+ 2 \sum_{i=2}^{n-3} \sum_{j=1}^{i-1} (r_h + 1) \\
&\quad \cdot F_{n-1}[r_{n-3}, \cdots, (r_{h-i+j} + 1), \cdots, (r_i - 1), \cdots, (r_j - 1), \cdots, (r_1 - 1)] \\
&+ \sum_{i=1}^{n-3} (r_h + 1)F_{n-1}[r_{n-3}, \cdots, (r_{h-2i} + 1), \cdots, (r_i - 2), \cdots (r_1 - 1)]
\end{aligned}
$$

where r_i, etc., represents the number of runs either up or down of exactly length i in each arrangement of the n numbers designated F_n,

(2) $\Sigma_{i=1}^{n-1} r_i = r$, the total number of runs having lengths exactly i (from 1 to $n - 1$) for each arrangement included in F_n,

(3) $\Sigma_{i=1}^{n-1} i r_i = n - 1$, that is, the sum of the lengths of all such runs in any arrangement is one less than the total number of numbers,

$F_n[r_{n-1}, r_{n-2}, \cdots, r_h, \cdots, r_i, \cdots, r_j, \cdots, r_1]$, the total number of nonidentical sequences of the n numbers with exactly r_{n-1} runs of length exactly $(n - 1)$, $\cdots$ r_h runs of length exactly h, $\cdots$ r_i runs of length exactly i, $\cdots$ r_j runs of length exactly j, $\cdots$ r_1 runs of length exactly 1. Some of these r's are of course zero and their sum is that given in (2) above. Similar statements apply to the four F_{n-1}'s.

In the last two summations in (1), when $r_j = r_1$, $(r_j - 1)$ combines with $(r_1 - 1)$ to give $(r_1 - 2)$, and when $r_i = r_1$, $(r_i - 2)$ combines with $(r_1 - 1)$ to give $(r_1 - 3)$.

By using the above recursion formula, the exact number of arrangements with at least one run up or down of length p or more has been computed for $n = 2$ to $n = 14$, inclusive. This information is given in Table 1. In addition, it has been used to determine the probabilities of arrangements with runs up or down of length p or more as shown in Table 2. These tables provide a useful background for the limiting expressions considered in the next three sections.

TABLE 1

Exact Numbers of Arrangements of n numbers with Runs of Length p or More

n \ p	1	2	3	4	5	6	7	8	9	10	11	12	13
2	2												
3	6	2											
4	24	14	2										
5	120	88	18	2									
6	720	598	156	22	2								
7	5,040	4,496	1,388	224	26	2							
8	40,320	37,550	13,334	2,352	304	30	2						
9	362,880	347,008	138,422	26,068	3,600	396	34	2					
10	3,628,800	3,527,758	1,554,854	309,178	44,640	5,220	500	38	2				
11	39,916,800	39,209,216	18,835,878	3,926,538	585,576	71,280	7,260	616	42	2			
12	479,001,600	473,596,070	245,249,548	53,333,016	8,159,498	1,021,680	108,240	9,768	744	46	2		
13	6,227,020,800	6,182,284,288	3,419,024,924	772,958,890	120,760,922	15,442,152	1,681,680	157,872	12,792	884	50	2	
14	87,178,291,200	86,779,569,238	50,852,433,294	11,920,405,298	1,895,856,108	246,427,634	27,387,360	2,642,640	222,768	16,380	1,036	54	2

TABLE 2

Exact Fraction of Arrangements of n numbers with Runs of Length p or More

n \ p	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	1.00000000	.00000000												
3	"	.33333333	.00000000											
4	"	.58333333	.08333333	.00000000										
5	"	.73333333	.15000000	.01666667	.00000000									
6	"	.83055556	.21666667	.03055556	.00277778	.00000000								
7	"	.89206349	.27539683	.04444444	.00515873	.00039683	.00000000							
8	"	.93129960	.33070437	.05833333	.00753968	.00074405	.00004960	.00000000						
9	"	.95626102	.38145392	.07183642	.00992063	.00109127	.00009369	.00000531	.00000000					
10	"	.97215553	.42847608	.08520117	.01230159	.00143849	.00013779	.00001047	.00000055	.0000000C				
11	"	.98227353	.47187846	.09836806	.01466991	.00178571	.00018188	.00001543	.00000105	.00000005	.00000000			
12	"	.98871501	.51200152	.11134204	.01703439	.00213294	.00022597	.00002039	.00000155	.00000010	.00000000	.00000000		
13	"	.99281574	.54906271	.12412981	.01939305	.00247986	.00027006	.00002535	.00000205	.00000014	.00000001	.00000000	.00000000	
14	"	.99542636	.58331533	.13673594	.02174688	.00282671	.00031415	.00003031	.00000256	.00000019	.00000001	.00000000	.00000000	.00000000

4. Solution for $p \geq \frac{n}{2}$. When $p \geq \frac{n}{2}$, it is clear that no sequence can contain more than one run of length p. Thus, the expected number of runs of length p or more in an arrangement is also the probability that an arrangement contains runs of length p or more. Writing Levene and Wolfowitz's [1] expression (4.2) in the simplified form previously published [3], we have

$$P(r_p') = E(r_p') = \frac{2[(n-p)(p+1)+1]}{(p+2)!} \quad \text{for} \quad \frac{n}{2} \leq p < n, \tag{4}$$

where r_p' represents the number of runs of length p or more. This expression checks exactly with Table 2 over the range to which it applies.

5. Solution for $\frac{(p+1)!}{n}$ = constant. As mentioned above, Wolfowitz [2] has shown that the limiting distribution for runs up and down is a Poisson Exponential. His proof applies specifically to the distribution of runs of length exactly p. However, the assumptions made in his derivation could have been applied to the distribution of runs of length p or more and would have led to identical conclusions for such runs. To see how closely this is approximated, it is possible to throw expression (4.17) for the variance of (r_p') derived by Levene and Wolfowitz [1] into the following simplified form:

$$\begin{aligned}\sigma^2(r_p') = \Bigg\{&\frac{2[(n-p)(p+1)+1]}{(p+2)!}\left[1 - \frac{2(p+1)^2[6p^2+7(p-1)]}{(p+2)!(2p+3)(2p+1)}\right. \\ &\left. - \frac{4(p+2)!}{(2p+3)!}\right] + \left[\frac{(p+1)[(2p+3)p(p-1)-6]}{p!(p+2)!(2p+3)(2p+1)}\right. \\ &\left. + \frac{2(p+1)^2+1}{(2p+3)!}\right]\Bigg\} \doteq [E(r_p')]\left[1 - \frac{3}{p!} - \frac{p!}{(2p)!}\right] + \frac{1}{2}\left[\frac{1}{(p!)^2} + \frac{1}{(2p)!}\right].\end{aligned} \tag{5}$$

Thus, $\sigma^2(r_p')$ is equal to $E(r_p')$ within one part in one thousand for $p \geq 7$ and it is apparent that the first two moments approximate those of a Poisson Exponential. Making use of this information, it is possible to prepare Table 3, which gives approximate values of the probabilities of arrangements with runs of length p or more based on

$$P(r_p') = 1 - e^{-E(r_p')} = 1 - e^{-(2[n-p)(p+1)+1])/(p+2)!}. \tag{6}$$

Comparison of Tables 2 and 3 shows agreement to closer than .0001 for $p \geq 6$, .001 for $p \geq 5$, .01 for $p \geq 4$, and .1 for $p \geq 3$ when $n \leq 14$. Similarly, the agreement for $p = 1$ is within .1 at $n \geq 4$, within .01 at $n \geq 8$, within .001 at $n \geq 11$ and .0001 at $n \geq 14$; the agreement for $p = 2$ is within .1 at $n \geq 10$. Possible agreement beyond $n = 14$ is of course subject to conjecture. However, it may be observed that the maximum difference for a given value of p was reduced from .2679 at $n = 2$, $p = 1$ to .1691 at $n = 6$, $p = 2$ indicating that closer agreement may be expected as p is increased.

6. Extrapolation from the exact solution for n small. Since the exponential in equation (6) may be written in the form:

$$e^{-(2[(n-p)(p+1)+1])/(p+2)!} = e^{(2[p(p+1)-1])/(p+2)!} \cdot e^{-(2n(p+1))/(p+2)!} \tag{7}$$

it follows that:

$$\frac{1 - P_{n+1}(r'_p)}{1 - P_n(r'_p)} = e^{-(2(p+1))/(p+2)!} \tag{8}$$

TABLE 3

Fraction of Arrangements of n numbers with Runs of Length p or More Based on Poisson Exponential

n \ p	1	2	3	4	5	6	7	8	9	10	>10
2	.7321	.0800									
3	.8111	.2835	.0165								
4	.9030	.4220	.0800	.0028							
5	.9502	.5654	.1393	.0165	.0004						
6	.9744	.6615	.1949	.0301	.0028	.0001					
7	.9869	.7364	.2467	.0435	.0052	.0004	.0000				
8	.9933	.7947	.2953	.0567	.0075	.0007	.0001	.0000			
9	.9965	.8401	.3408	.0697	.0099	.0011	.0001	.0000	.0000		
10	.9982	.8742	.3833	.0825	.0122	.0014	.0001	.0000	.0000	.0000	
11	.9991	.9030	.4230	.0952	.0146	.0018	.0002	.0000	.0000	.0000	.0000
12	.9995	.9244	.4603	.1076	.0169	.0021	.0002	.0000	.0000	.0000	.0000
13	.9997	.9412	.4951	.1200	.0193	.0025	.0003	.0000	.0000	.0000	.0000
14	.9999	.9542	.5276	.1321	.0216	.0028	.0003	.0000	.0000	.0000	.0000
15	.9999	.9643	.5581	.1441	.0239	.0032	.0004	.0000	.0000	.0000	.0000
20	1.0000	.9898	.6834	.2015	.0355	.0049	.0006	.0001	.0000	.0000	.0000
40	"	.9999	.9165	.3952	.0803	.0118	.0015	.0002	.0000	.0000	.0000
60	"	1.0000	.9780	.5419	.1231	.0186	.0023	.0003	.0000	.0000	.0000
80	"	"	.9942	.6530	.1639	.0254	.0032	.0004	.0000	.0000	.0000
100	"	"	.9985	.7371	.2030	.0322	.0041	.0005	.0000	.0000	.0000
200	"	"	1.0000	.9345	.3717	.0652	.0085	.0010	.0001	.0000	.0000
500	"	"	"	.9990	.6924	.1577	.0215	.0024	.0002	.0000	.0000
1000	"	"	"	1.0000	.9065	.2919	.0428	.0049	.0005	.0000	.0000
5000	"	"	"	"	1.0000	.8234	1976	.0245	.0025	.0002	.0000

showing that consecutive values of $1 - P(r'_p)$ are related by a constant of proportionality dependent only on p. Since this is true in the limit, Table 2 was examined to determine similar multipliers for extrapolation. The results of this examination are shown in Table 4 together with the values of (8). This table shows that the agreement between the value of $\frac{1 - P_{n+1}(r'_p)}{1 - P_n(r'_p)}$ for $n = 12$, e.g., and $e^{(-2(p+1))/(p+2)!}$ becomes closer the larger the value of p. The con-

TABLE 4

Determination of Extrapolation Constant

p \ n	2		3		4		5		6	
	$1 - P_n(r_2')$	$\frac{1 - P_{n+1}(r_2')}{1 - P_n(r_2')}$	$1 - P_n(r_3')$	$\frac{1 - P_{n+1}(r_3')}{1 - P_n(r_3')}$	$1 - P_n(r_4')$	$\frac{1 - P_{n+1}(r_4')}{1 - P_n(r_4')}$	$1 - P_n(r_5')$	$\frac{1 - P_{n+1}(r_5')}{1 - P_n(r_5')}$	$1 - P_n(r_6')$	$\frac{1 - P_{n+1}(r_6')}{1 - P_n(r_6')}$
2	1.00000000	.66666667								
3	.66666667	.62500000	1.00000000	.91666667						
4	.41666667	.64000000	.91666667	.92727273	1.00000000	.98333333				
5	.26666667	.63541667	.85000000	.92156863	.98333333	.98587571	1.00000000	.99722222		
6	.16944444	.63700234	.78333333	.92502533	.96944444	.98567336	.99722222	.99761242	1.00000000	.99960317
7	.10793651	.63648898	.72460317	.92367196	.95555556	.98546511	.99484127	.99760670	.99960317	.99965264
8	.06870040	.63666266	.66929563	.92417469	.94166667	.98566044	.99246032	.99760096	.99925595	.99965252
9	.04373898	.63660534	.61854608	.92397954	.92816358	.98560087	.99007937	.99759518	.99890873	.99965240
10	.02784447	.63662455	.57152392	.92405851	.91479883	.98560679	.98769841	.99760218	.99856151	.99965228
11	.01772647	.63661817	.52812154	.92402684	.90163194	.98561056	.98533009	.99760032	.99821429	.99965215
12	.01128499	.63662030	.48799848	.92405470	.88865796	.98561002	.98296561	.99760046	.99786706	.99965234
13	.00718426	.63661959	.45093729	.92404128	.87587019	.98560731	.98060695	.99759961	.99752014	.99965228
14	.00457364		.41668467		.86326406		.97825312		.99717329	
Chosen value of $\frac{1 - P_{n+1}(r_p')}{1 - P_n(r_p')}$		.63662 $\doteq \frac{2}{\pi}$		.92404		.98561		.99760		.999652
Value of $e^{-(2(p+1))/(p+2)!}$		.77880		.93550		.98629		.99762		.9996528

stancy of the ratio for a given value of p is such as to permit calculation of probabilities for any value of n to a minimum of three or possibly four decimal places. Such calculations have been made and recorded in Table 5. The following formulae[1] were used for these calculations:

$$
\begin{aligned}
P_n(r_1') &= 1 \\
P_n(r_2') &= 1 - (.00437364)\left(\frac{2}{\pi}\right)^{n-14} \\
P_n(r_3') &= 1 - (.45093729)(.92404)^{n-13} \\
P_n(r_4') &= 1 - (.87587019)(.98561)^{n-13} \\
P_n(r_5') &= 1 - (.98060695)(.99760)^{n-13} \\
P_n(r_6') &= 1 - (.99752014)(.999652)^{n-13}
\end{aligned} \tag{9}
$$

or in general

$$P_n(r_p') = 1 - [1 - P_{n_0}(r_p')][\text{Constant}_p]^{n-n_0}. \tag{10}$$

Comparison of Table 3 with Tables 2 and 5 shows that the difference for given p and n has a maximum for each value of p and that this maximum decreases with increase in p. The maximum values of the difference shown in the tables are: $p = 1, n = 2$, .2679; $p = 2, n = 6$, .1691; $p = 3, n = 20$, .0572; $p = 4, n = 80$, .0154; $p = 5, n = 500$, .0033; and $p = 6, n = 5000$, .0007. Thus, it is apparent that the agreement beyond $p = 6$ should be within .0001 and the method of Section 5 used for Table 3 is satisfactory for these probabilities.

7. Constant probability relationships. From Tables 2, 3 and 5, it is possible to make interpolations for the values of n required to have a probability of at least $P(r_p')$ that an arrangement will have a run of length p or more. When the conditions of Section 5 apply, the value of n is, of course:

$$n = p - \frac{1}{p+1} - \frac{p+2}{2}\, p! \log_e [1 - P(r_p')]. \tag{11}$$

---

[1] It will be noted that the constant for $p = 2$ has been taken to be $\frac{2}{\pi}$, whereas the last value shown in Table 4 is .63661959. However, alternate values in this series are converging. Comparing these subseries shows that by $n = 16$, the values would agree with $\frac{2}{\pi}$ to eight decimal places. An analytic proof that $\frac{2}{\pi}$ is the limiting value of the constant has recently been found by J. W. Tukey.

While reading the manuscript J. Riordan observed that the number of arrangements with longest length 1, say $f(n, 1)$ has the generating function

$$\Sigma f(n, 1) \frac{t^n}{n!} = 2(\sec t + \tan t)$$

hence is twice the Euler number for n even and twice the tangent number for n odd, a result given essentially by Netto [4]. These observations lead directly to the limiting value, $\frac{2}{\pi}$ noted above.

TABLE 5

Fraction of Arrangements of n Numbers with Runs of Length p or More Based on Extrapolation with Extrapolation Constant

n \ p	1	2	3	4	5	6
14	1.0000	.9954	.5833	.1367	.0217	.0028
15	"	.9971	.6150	.1492	.0241	.0032
20	"	.9997	.7406	.2086	.0358	.0049
40	"	1.0000	.9466	.4078	.0810	.0118
60	"	"	.9890	.5568	.1241	.0187
80	"	"	.9977	.6684	.1652	.0255
100	"	"	.9995	.7518	.2044	.0322
200	"	"	1.0000	.9418	.3743	.0653
500	"	"	"	.9992	.6957	.1580
1000	"	"	"	1.0000	.9085	.2925
5000	"	"	"	"	1.0000	.8241

TABLE 6

Sample Size for Constant Probability Based on Poisson Exponential

P \ p	1	2	3	4	5	6	7	8
$\leq .99$	7	20	71	335	1939	13268		
$\leq .95$	5	13	47	219	1263	8633		
$\leq .90$	3	10	37	169	971	6637		
$\leq .10$	0	2	4	11	49	309	2296	
$\leq .05$	0	1	3	7	26	153	1170	10350
$\leq .01$	0	1	2	4	9	34	235	2036

TABLE 7

Sample Size for Constant Probability Based on Extrapolation

P \ p	1	2	3	4	5	6
$\leq .99$	—	12	61	321	1923	13239
$\leq .95$	—	8	40	210	1253	8614
$\leq .90$	—	7	32	162	964	6622
$\leq .10$	—	(2)	4	11	48	308
$\leq .05$	—	(2)	(3)	7	26	153
$\leq .01$	—	(2)	(3)	(4)	9	34

Similarly, it may be obtained from the extrapolation formulae of Section 6 in the form:

$$n = n_0 + \frac{\log [1 - P_n(r'_p)] - \log [1 - P_{n_0}(r'_p)]}{\log [\text{Constant}_p]}. \tag{12}$$

Results of computations based on (11) and (12), are given in Tables 6 and 7, respectively for particular values of $P(r'_p)$. It will be noted that Table 7 is in exact agreement with Table 2 and that it differs but little in a practical sense from Table 6.

References

[1] H. LEVENE AND J. WOLFOWITZ, "The covariance matrix of runs up and down," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 58–69.

[2] J. WOLFOWITZ, "Asymptotic distribution of runs up and down," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 163–172.

[3] P. S. OLMSTEAD, Review of "A Significance Test for Time Series" by W. A. Wallis and G. H. Moore, *Am. Stat. Assn. Jour.*, Vol. 37 (1942), p. 152.

[4] E. NETTO, *Lehrbuch der Combinatorik*, Leipzig, 1901, pp. 110–112.