THE ASYMPTOTIC DISTRIBUTION OF RUNS OF CONSECUTIVE ELEMENTS

By Irving Kaplansky

New York City

In a permutation of $1, 2, \cdots, n$ let r denote the number of instances in which i is next to $i + 1$, i.e., in which either of the successions $(i, i + 1)$ or $(i + 1, i)$ occurs. Thus for the permutation 234651, $r = 3$. In [3] Wolfowitz[1] has proposed the use of r for significance tests in the non-parametric case, and in [4] he has shown that asymptotically r has the Poisson distribution with mean value 2. It is to be noted that $W(R)$, the number of runs as defined by Wolfowitz, is equal to $n - r$.

In this note we shall derive more explicit results concerning the asymptotic distribution of r. In a random permutation (all permutations being regarded as equally probable) let the probability of exactly r successions as above be $P(n, r)$, and let $M(n, k)$ denote the k-th factorial moment of the distribution, that is

$$M(n, k) = \Sigma_r r(r - 1) \cdots (r - k + 1)P(n, r).$$

We shall show that

$$(1) \qquad M(n, k) = 2^k\left[1 - \frac{k+1}{2k}\binom{k}{1}\frac{k}{n} + \frac{k+2}{2^2 k}\binom{k}{2}\frac{k(k-1)}{n(n-1)} - \cdots\right]$$

$$(2) \qquad P(n, r) = \frac{2^r e^{-2}}{r!}\left[1 - \frac{r^2 - 3r}{2n} + \frac{r^4 - 8r^3 + 9r^2 + 22r - 16}{8n(n-1)}\right] + 0(n^{-3}).$$

Since 2^k is the k-th factorial moment of the Poisson distribution with mean 2, either of these results serves to verify the asymptotic Poisson character of the distribution of r.

It would be possible to obtain some kind of explicit formula for the general term of (2), but there seems to be no reasonably simple form.

Proof of (1). Let A_i denote the event "$i + 1$ comes right after i" and B_i the event "i comes right after $i + 1$" $(i = 1, \cdots, n - 1)$. The joint probability of k of these $2n - 2$ events is either 0, if they are incompatible, or $(n - k)!/n!$ if they are compatible—for in the latter case we in effect assign positions for k of the elements and are then free to permute the $n - k$ others. Let $f(n, k)$ denote the number of ways of selecting k compatible events. Then it is known that ([1], eq. (40))

$$(3) \qquad M(n, k) = k!f(n, k)(n - k)!/n! = f(n, k)\Big/\binom{n}{k}.$$

[1] I am indebted to Dr. Wolfowitz for calling my attention to this problem, and to its identity with what I called the "n-kings problem" in [2].

The relations of incompatibility can be summarized by the statement that A_i is incompatible with B_j if $|i - j| \leqq 1$. In view of (3), our task thus reduces to the proof of the following combinatorial lemma.

LEMMA. *Suppose* $2n - 2$ *objects* $A_1, \cdots, A_{n-1}, B_1, \cdots, B_{n-1}$ *are given. Let* $f(n, k)$ *denote the number of ways of selecting* k *objects with the restriction that* A_i *and* B_j *must not both be chosen when* $|i - j| \leqq 1$. *Then*

$$\frac{f(n, k)}{2^k} = \sum_{i=0}^{k} (-1)^i \frac{k+1}{2^i k} \binom{k}{i} \binom{n-i}{k-i}. \tag{4}$$

PROOF. We split the acceptable selections into two subsets: those which include A_{n-1} and those which do not. Let the latter be $g(n, k)$ in number. Since the selections which include A_{n-1} must omit B_{n-1} and B_{n-2}, it is clear that they are $g(n - 1, k - 1)$ in number. Thus

$$f(n, k) = g(n, k) + g(n - 1, k - 1). \tag{5}$$

Similarly we split the selections which omit A_{n-1} according as they omit or include B_{n-1}; we obtain

$$g(n, k) = f(n - 1, k) + g(n - 1, k - 1). \tag{6}$$

Elimination of g from (5) and (6) yields[2]

$$f(n, k) = f(n - 1, k) + f(n - 1, k - 1) + f(n - 2, k - 1). \tag{7}$$

We can now make an inductive proof of (4). Assuming (4), we have

$$\frac{f(n, k) - f(n - 1, k)}{2^k} = \Sigma(-1)^i \frac{k+1}{2^i k} \binom{k}{i} \binom{n-i-1}{k-i-1}$$

$$\frac{f(n-2, k-1)}{2^{k-1}} = \Sigma(-1)^i \frac{k+i-1}{2^i(k-1)} \binom{k-1}{i} \cdot \left[\binom{n-i-1}{k-i-1} - \binom{n-i-2}{k-i-2}\right]$$

$$= \Sigma(-1)^i \binom{n-i-1}{k-i-1} \cdot \left[\frac{k+i-1}{2^i(k-1)} \binom{k-1}{i} + \frac{k+i-2}{2^{i-1}(k-1)} \binom{k-1}{i-1}\right].$$

In view of the identity

$$\frac{k+i}{k} \binom{k}{i} = \frac{k+i-1}{k-1} \binom{k-1}{i} + \frac{k+i-2}{k-1} \binom{k-1}{i-1}$$

we now readily verify that the right hand side of (4) satisfies (7). To complete the induction we must check the appropriate boundary conditions. According to (4) we have

$$\frac{f(k, k)}{2^k} = \sum_{i=0}^{k} (-1)^i \frac{k+i}{2^i k} \binom{k}{i} = 0,$$

$f(n, 1) = 2n - 2$, both as they should be.

[2] This recursion formula is essentially the same as equation (20) in [2].

Note. There are various other formulas for $f(n, k)$; we have selected (4) as it exhibits the asymptotic behaviour best. In an unpublished investigation John Riordan obtained a neat representation as a hypergeometric function:

$$f(n, k) = 2(n - k)F(1 - k, 1 + k - n; 2; 2)$$

and derived corresponding recursion formulas. Essentially the same result was given by Wolfowitz [3]. Still another formula given by Riordan is

$$f(n, k) = 2\sum_{i=0}^{k-1} \binom{k-1}{i}\binom{n-1-i}{k}.$$

A symbolic version is given in §5 of [2].

Proof of (2). From the formula of Poincaré ([1], eq. (29))

$$r!P(n, r) = \sum_{k=r}^{n} (-1)^{k+r} M(n, k)/(k - r)!$$

or, in a cabalistic symbolic form, $P(n, r) = M^r e^{-M}/r!$. We substitute the successive terms of (1) and we may let the sum run to infinity at a cost of $0(n^{-m})$ for any positive m. The first term contributes[3]

$$\sum_{k=r}^{\infty} (-1)^{k+r} 2^k/(k - r)! = 2^r \sum_{i=0}^{\infty} (-2)^i/i! = 2^r e^{-2}.$$

Again since

$$k^2 + k = (k - r)(k - r - 1) + (2r + 2)(k - r) + r^2 + r,$$

the next term yields

$$\sum_{k=r}^{\infty} (-1)^{k+r}(k^2 + k)2^{k-1}/(k - r)! = 2^r e^{-2}\left(2 - 2r - 2 + \frac{r^2 + r}{2}\right),$$

and so on in obvious fashion.

Some indication of the asymptotic behavior of $P(n, r)$ is afforded by the following table for $n = 10$. It is to be noted that, because of the form of (2), the approach to Poisson is much more rapid for $r = 0$ and 3 than for other r.

r	$P(10, r)$	Poisson	First two terms of (2)
0	.132	.135	.135
1	.300	.271	.298
2	.305	.271	.298
3	.179	.180	.180
4	.065	.090	.072
5	.015	.036	.018
6	.002	.012	.001
7	.000	.003	.001

[3] My thanks are due to Mr. Riordan for correcting an error in this section, and for many helpful suggestions concerning the entire paper.

REFERENCES

[1] M. FRÉCHET, "Les probabilités associées a un système d'événements compatibles et dépendants," *Actualités Scientifiques et Industrielles*, no. 859, Paris, 1940.
[2] I. KAPLANSKY, "Symbolic solution of certain problems in permutations," *Bull. Amer. Math. Soc.*, Vol. 50, (1944) pp. 906–914.
[3] J. WOLFOWITZ, "Additive partition functions and a class of statistical hypotheses," *Annals of Math. Stat.*, Vol. 13, (1942) pp. 247–279.
[4] J. WOLFOWITZ, "Note on runs of consecutive elements," *Annals of Math. Stat.*, Vol. 15, (1944) pp. 97–98.