MULTIPLE MATCHING AND RUNS BY THE SYMBOLIC METHOD

IRVING KAPLANSKY AND JOHN RIORDAN

New York City

1. Introduction. The two subjects in the title have generally been treated by distinct methods, an excellent summary of which is given by S. S. Wilks in Chapter X of [13]. For two-deck matching, an appreciable simplification over the classical work of MacMahon [7], which seems to underlie the generating function used by Wilks [12] and Battin [2], has been shown by one of us [5] to follow from symbolic methods. Here we give an elaboration of these methods to multiple matching and to runs.

The basis of the symbolic method in both problems has been given in [6], but for completeness a skeleton resume is given in Section 2 below. A new point is stressed: the relation of coefficients in polynomials of the symbolic method to factorial moments (cf. Fréchet [4]).

The emphasis for the most part is on showing the expedition of the symbolic method in reaching known results, but in several instances new results are obtained.

2. Symbolic expressions and moments. Let $A_1, \cdots, A_n$ be arbitrary events and let $p(A_{i_1}, \cdots, A_{i_k})$ denote the joint probability of $A_{i_1}, \cdots, A_{i_k}$; let P_r be the probability that exactly r of the events occur. Then

$$P_r = \sum_{k=0}^{n} (-1)^r {}_kC_r \Sigma(-1)^k p(A_{i_1}, \cdots, A_{i_k}) \tag{1}$$

and in particular

$$P_0 = \sum_{k=0}^{n} \Sigma(-1)^k p(A_{i_1}, \cdots, A_{i_k}),$$

or symbolically

$$P_0 = [1 - p(A_1)][1 - p(A_2)] \cdots [1 - p(A_n)]. \tag{2}$$

The cases to be studied will be exclusively ones where so-called *quasi-symmetry* holds, i.e., $p(A_{i_1}, \cdots, A_{i_k})$ is either 0 or a function ϕ_k of k alone. In that event (2) can be evaluated as follows: suppress all products that vanish, and form a polynomial $f(E)$ by replacing each surviving term $p(A_i)$ by E. Then $P_0 = f(E)\phi_0$ where E is a displacement operator: $E^k\phi_0 = \phi_k$.

The same polynomial $f(E)$ can also be used to obtain P_r and the moments of the distribution. From (1) we see that $P_r = f(E)\psi_0$, where $\psi_k = (-1)^r {}_kC_r\phi_k$. Again it is well known (Fréchet [4]) that the k-th factorial moment, defined by

$$M_{(k)} = \sum_{i=0}^{n} i(i-1) \cdots (i-k+1)P_i,$$

is also given by

$$M_{(k)} = k!\Sigma p(A_{i_1}, \cdots, A_{i_k}).$$

It follows that the terms of $f(E)\phi_0$ are essentially the factorial moments. More precisely, if

$$f(E) = \sum_{k=0}^{n} S_k(-E)^k,$$

then

$$M_{(k)} = k!S_k\phi_k. \quad (3)$$

3. Card matching. To avoid complications which add nothing to the fundamental idea, the case of three decks will be considered explicitly. As remarked by Battin [2], there is no loss of generality in supposing that the three decks have the same number of cards: let them be numbered from 1 to n. Let p_{ijk} denote the probability that the i-th, j-th, and k-th cards of the three decks are matched, that is, all occur in say the l-th place. The condition of quasi-symmetry is fulfilled, the (symbolic) product of k of the p's being either 0 or $\phi_k = [(n - k)!/n!]^2$.

The simplest problem is to find the probability that there be no triple matches of the form (i, i, i). Since no products of the expression

$$(1 - p_{111})(1 - p_{222}) \cdots (1 - p_{nnn})$$

vanish, the answer is $(1 - E)^n\phi_0$, in agreement with Anderson [1] (cf. also problem E 589 in the *American Mathematical Monthly*, p. 512, 1943; solution by John Riordan, p. 287, 1944).

Suppose now that the decks are given compositions in the usual fashion by having a_1, b_1, c_1 aces respectively, a_2, b_2, c_2 deuces, etc. We may number the cards so that $1, \cdots, a_1$ are aces, $a_1 + 1, \cdots, a_1 + a_2$ are deuces, and similarly in the other decks. The probability of precisely r matches among cards of the same denomination is then given by

$$F(a_1, b_1, c_1)F(a_2, b_2, c_2) \cdots \psi_0, \quad (4)$$

where

$$F(a, b, c) = \Pi(1 - p_{ijk})$$

the symbolic product being taken over ranges $i = 1, \cdots, a, j = 1, \cdots, b, k = 1, \cdots, c$.

A simple combinatorial argument reveals that

$$F(a, b, c) = \Sigma_t(a)_t(b)_t(c)_t(-E)^t/t! \quad (5)$$

where $(a)_t = a(a - 1) \cdots (a - t + 1)$ is the Jordan factorial notation. The problem of matching arbitrary decks is thus compactly solved by (4) and (5).

4. Examples. When decks of explicit structure are in question, the computation of probabilities and moments reduces to straightforward algebra, as is illustrated in the three following examples.

1. Suppose each of three decks has two suits of two cards each. Then, since

$$F(2, 2, 2)^2 = (1 - 8E + 4E^2)^2 = 1 - 16E + 72E^2 - 64E^3 + 16E^4,$$

it follows that

$$\begin{aligned}(4!)^2P_0 &= (4!)^2 - 16(3!)^2 + 72(2!)^2 - 64(1!)^2 + 16(0!)^2\\ &= 576 - 576 + 288 - 64 + 16 = 240,\end{aligned}$$

and the calculation of $(4!)^2P_r$ may be set forth as follows:

$$\begin{array}{lrrrrrrrrrrr} r & & & & & & & & & & & \\ 0 & 576 & - & 576 & + & 288 & - & 64 & + & 16 & = & 240 \\ 1 & & & 576 & - & 576 & + & 192 & - & 64 & = & 128 \\ 2 & & & & & 288 & - & 192 & + & 96 & = & 192 \\ 3 & & & & & & & 64 & - & 64 & = & 0 \\ 4 & & & & & & & & & 16 & = & 16 \end{array}$$

each column being obtained by multiplying its first row entry by a binomial coefficient. These results may be verified readily by direct enumeration.

2. In the case of three 5 by 5 decks, the polynomial is

$$\begin{aligned}F(5, 5, 5)^5 &= (1 - 125E + 4000E^2 - 36000E^3\\ &\qquad + 72000E^4 - 14400E^5)^5\\ &= 1 - 625E + 176{,}250E^2 - 29{,}711{,}250E^3\\ &\qquad + 3{,}346{,}063{,}125E^4 \cdots\end{aligned}$$

The factorial moments can be obtained using (3).

$$\begin{aligned}M_{(1)} &= 625/25^2 = 1,\\ M_{(2)} &= 2\cdot 176250/25^2\cdot 24^2 = 47/48,\\ M_{(3)} &= 7923/8464,\\ M_{(4)} &= 1784567/2048288,\end{aligned}$$

the first two in agreement with Battin [2].

3. The symbolic method can be applied to more intricate kinds of matching, as this final example shows. Suppose that the six matches represented by (123) and its permutations are forbidden, likewise the six matches represented by permutations of (456), and so on in groups of three. Then

$$\begin{aligned}(1 - p_{123})(1 - p_{132})(1 - p_{213})(1 - p_{231})(1 - p_{312})(1 - p_{321})&\\ = 1 - 6E + 6E^2 - 2E^3,&\end{aligned}$$

and so the answer is

$$(1 - 6E + 6E^2 - 2E^3)^{n/3}.$$

The analogous problem for 4 decks has the solution

$$(1 - 24E + 108E^2 - 96E^3 + 24E^4)^{n/4}.$$

The generalization to an arbitrary number of decks involves the enumeration of Latin rectangles, in itself a formidable problem.

5. Moment formulas. It is possible to deduce from (4) and (5) fairly explicit formulas for the factorial moments. Let us define $u^{(t)} = (a)_t(b)_t(c)_t$. Then (5) may be written symbolically as

$$F(a, b, c) = \Sigma_t u^{(t)}(-E)^t/t! = \exp(-uE).$$

Writing $F(a_i, b_i, c_i) = \exp(-u_iE)$, we then have

$$P_0 = \exp[-(u_1 + u_2 + \cdots)E]\phi_0$$
$$= \Sigma_t(u_1 + u_2 + \cdots)^t \frac{(-E)^t}{t!} \phi_0,$$

or finally, if $m + 1$ decks are being matched,

$$P_0 = \Sigma_t(-)^t(u_1 + u_2 + \cdots)^t/t!(n)_t^m. \tag{6}$$

It is to be borne in mind that after expansion of $(u_1 + u_2 + \cdots)^t$ by the multinomial theorem, the term $u_1^x u_2^y u_3^z \cdots$ is replaced by $u_1^{(x)}u_2^{(y)}u_3^{(z)} \cdots$ with the u's defined as above.

By (3), factorial moments corresponding to (6) are given by

$$M_{(t)} = (u_1 + u_2 + \cdots)^t/(n)_t^m. \tag{7}$$

Thus in particular

$$n^m M_{(1)} = u_1 + u_2 + \cdots = \Sigma_i a_i b_i \cdots$$
$$n^m(n - 1)^m M_{(2)} = (u_1 + u_2 + \cdots)^2$$
$$= \Sigma_i a_i(a_i - 1)b_i(b_i - 1) \cdots + 2\Sigma_{i \neq j} a_i a_j b_i b_j \cdots$$

the cases $m = 1, 2$ in agreement with Battin [2].

In the simple case where $m = 1$ (two decks), $a_i = b_i = a$ and $n = sa$, we have $u^{(t)} = (a)_t^2$ and

$$(n)_t M_{(t)} = (u + u + \cdots u)^t \tag{8}$$

with su's in the parenthesis. The right of (8) is the multi-variable polynomial of E. T. Bell [3], $Y_t(y_1, y_2, \cdots, y_t)$ with $y_k = (s)u^{(k)}$ and (s) a symbolic factorial such that $y_iy_j = (s)_2u^{(i)}u^{(j)}$, etc. Instances of (8) may be compared with Olds [9].

Expanding (8) we obtain

$$(n)_t M_{(t)} = (s)_t [u^{(1)}]^t + {}_tC_2(s)_{t-1}u^{(2)}[u^{(1)}]^{t-2} + \cdots$$
$$= (s)_t a^{2t} + {}_tC_2(s)_{t-1}a^{2t-2}(a-1)^2 + \cdots$$

and, since $(s)_t/(n)_t \to a^{-t}$ as $n \to \infty$, it follows that $M_{(t)} \to a^t$, i.e., the limiting distribution is Poisson with mean a. As indicated in [6] one may proceed to obtain successive terms of an asymptotic series for the distribution. These results generalize to the case where $M_{(1)} = \Sigma a_i b_i/n$ approaches a finite limit as $n \to \infty$. In certain instances where $M_{(1)} \to \infty$, asymptotic normality can be proved (cf. [1] and [8]).

6. Successions and runs. As shown in [6], enumeration of permutations with a specified number of 2-successions like 12, 42, $\cdots$ may be accomplished by introduction of symbols like q_{12}, q_{42}, denoting probabilities that 1 immediately precede 2, 4 precede 2, resp. For permutations of objects a_1 of which are of one kind, a_2 of a second, $\cdots$ with $a_1 + a_2 + \cdots a_s = n$, the probability of exactly r 2-successions is ([6] p. 914)

$$P_r = G(a_1)G(a_2) \cdots G(a_s)\psi_0 \tag{9}$$

with $\psi_k = (-1)^r_k C_r(n-k)!/n!$ and

$$G(a) = \sum_{t=0}^{a-1} (a)_t(a-1)_t(-E)^t/t!.$$

It is to be noted that in deriving (9), elements of the first kind are numbered 1 to a_1, of the second $a_1 + 1$ to $a_1 + a_2$, $\cdots$ and a succession occurs if either i precedes j or j precedes i with i and j in the same set.

For $s = 2$, i.e., two kinds of elements, there is a simpler formula due to Stevens [10], but for the general case (9) seems to be the only reasonably explicit solution known. In particular, for the function $F(a_1, \cdots, a_s)$ of Mood [8] which enumerates the number of permutations with no 2-successions, we have

$$F(a_1, \cdots, a_s) = n!G(a_1) \cdots G(a_s)\phi_0.$$

Factorial moments for 2-successions are given at once by (7):

$$M_{(t)} = (u_1 + u_2 + \cdots + u_s)^t/(n)_t \tag{10}$$

with $u_i^{(j)} = (a_i)_j(a_i - 1)_j$.

It is more usual to classify permutations according to the number of runs, say r', a run consisting of a succession of i like elements ($i = 1, 2, \cdots$). Since every 2-succession causes the loss of a potential run, we have $r' = n - r$, i.e. the number of runs is n diminished by the number of 2-successions. Factorial moments $\bar{M}_{(t)}$ for runs are then given by the usual formula for change of origin:

$$\bar{M}_{(t)} = \sum_{i=0}^{t} (-1^i {}_tC_i(n-i)_{t-i}M_{(i)}. \tag{11}$$

Examples. 1. Introducing α_i for the i-th elementary symmetric function of the a's,

$$\alpha_1 = a_1 + a_2 + \cdots a_s = n,$$
$$\alpha_2 = a_1a_2 + a_1a_3 + \cdots + a_{s-1}a_s,$$
$$\alpha_3 = a_1a_2a_3 + \cdots,$$

we may derive from (10) and (11) the formula

$$\bar{M}_{(1)} = 1 + 2\alpha_2/n \tag{12}$$

for the mean number of runs. The variance σ^2, the same for runs and 2-successions, is given by

$$\sigma^2 = M_{(2)} + M_{(1)} - M_{(1)}^2 = \frac{2\alpha_2(2\alpha_2 - n) - 6n\alpha_3}{n_2(n-1)}. \tag{13}$$

For runs of two kinds of elements, formulas (12) and (13) specialize to those given by Wald and Wolfowitz [11].

2. For runs of elements of a single kind, factors in (9) pertaining to other elements are suppressed. Thus if a is written for a_1, and terms in $a_2, \cdots, a_s$ are suppressed, (9) and (10) become

$$P_r = G(a)\psi_0,$$
$$M_{(t)} = (a)_t(a-1)_t/(n)_t.$$

Moments for runs are given by

$$\bar{M}_{(t)} = \sum_{i=0}^{t} (-1)^i {}_tC_i(n-i)_{t-i}M_{(i)} = (a)_t(n-a+1)_t/(n)_t$$

in agreement with Mood [8].

REFERENCES

[1] T. W. ANDERSON, "On card matching," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 426–435.
[2] I. L. BATTIN, "On the problem of multiple matching," *Annals of Math Stat.*, Vol. 13 (1942), pp. 294–305.
[3] E. T. BELL, 'Exponential polynomials," *Annals of Math.*, Vol. 35 (1934), pp. 258–277.
[4] M. FRÉCHET, "Les probabilités associées à un système d'événements compatibles et dépendants," *Actualités Scientifiques et Industrielles*, no. 859, Paris, 1940.
[5] I. KAPLANSKY, "On a generalization of the probleme des rencontres," *Amer. Math. Monthly*, Vol. 46 (1939), pp. 159–161.
[6] I. KAPLANSKY, "Symbolic solution of certain problems in permutations," *Bull. Amer. Math. Soc.*, Vol. 50 (1944), pp. 906–914.
[7] P. A. MACMAHON, *Combinatory Analysis*, Cambridge 1915, especially Vol. I, pp. 99–114.
[8] A. M. MOOD, "The distribution theory of runs," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 367–392.
[9] E. G. OLDS, "A moment-generating function which is useful in solving certain matching problems," *Bull. Amer. Math. Soc.*, Vol. 44 (1938), pp. 407–413.
[10] W. L. STEVENS, "Distribution of groups in a sequence of alternatives," *Annals of Eugenics*, Vol. 9 (1939).
[11] A. WALD AND J. WOLFOWITZ, "On a test whether two samples are from the same population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 147–162.
[12] S. S. WILKS, *Statistical Aspects of Experiments in Telepathy*, a lecture delivered to the Galois Institute of Mathematics, Long Island University, Dec. 4, 1937.
[13] S. S. WILKS, *Mathematical Statistics*, Princeton, 1943.