# ON THE PROBLEM OF MULTIPLE MATCHING

BY I. L. BATTIN
*Drew University*

**1. Introduction.** The problem of determining the distribution of the number of "hits" or "matchings" under random matching of two decks of cards has received attention from a number of authors within the last few years. In 1934 Chapman [2] considered pairings between two series of $t$ elements each, and later [3] generalized the problem to series of $u$ and $t(\leq u)$ elements respectively. In the same paper he also considered the distribution of the mean number of correct matchings resulting from $n$ independent trials, and gave a method, and tables, for determining the significance of any obtained mean. In 1937 Bartlett [1] considered matchings of two decks of cards, using a number of interesting moment generating functions. In 1937 Huntington [12, 13] gave tables of probabilities for matchings between decks with the compositions $(5^5)$, $(4^4)$, and $(3^3)$, where $(s^t)$ denotes a deck consisting of $s$ of each of $t$ kinds of cards. More generally $(s_1 s_2 \cdots s_t)$ denotes $s_1$ cards of the first kind, $s_2$ of the second, etc. Sterne [16] has given the first four moments of the frequency distribution for the $(5^5)$ case and has fitted a Pearson Type I distribution function to the distribution. Sterne obtained his results by considering the probabilities in a $5 \times 5$ contingency table. He also considered the $4 \times 4$ and $3 \times 3$ cases. In 1938 Greville [7] gave a table of the exact probabilities for matchings between two decks of compositions $(5^5)$. Greenwood [4] obtained the variance of the distribution of hits for matchings between two decks having the respective compositions $(s^t)$ and $(s_1 s_2 \cdots s_t)$ with $s_1 + s_2 + \cdots + s_t = st = n$, and where it is not necessary that all the $s$'s should be different from zero. Earlier Wilks [19] had considered the same problem for $t = 5$ and $n = 25$.

In a very interesting paper Olds [15] in 1938 used permanents to express a moment generating function suitable for the problem in question. He obtained factorial moments and the first four ordinary moments about the mean, first for two decks with composition $(4^2)$, and then for two decks of composition $(s^t)$. In 1938 Stevens [17] considered a contingency table in connection with matchings between two sets of $n$ objects each, and gave the means, variances, and covariances of the single cell entries and various sub-totals of the cell entries. Stevens [18] also gave a treatment of the problem of matchings between two decks which was based on elementary considerations. In 1940 Greenwood [6] gave the first four moments of the distribution of hits between two decks of any composition whatever, generalizing the problem which had been treated earlier by Olds [15]. Finally in 1941, Greville [8] gave the exact distribution of hits for matchings between two decks of arbitrary composition. He also considered the problem from the standpoint of a contingency table, as had been done earlier by Stevens.

In 1939 Kullback [14] considered matchings between two sequences obtained by drawing at random a single element in turn from each of $n$ urns $U_i$ containing elements of $r$ types $E_j$ in the respective proportions $p_{ij}$. He showed that if the process of drawing were indefinitely repeated the distribution of hits would be that of a Poisson series.

The work which has been done thus far applies to the problem of matching two decks of cards. In the present paper a method is developed for obtaining the moments of the distribution of hits for matchings between three or more decks of cards of arbitrary composition.

**2. Matchings between two Decks of cards.** In the present paper it will be convenient to take as the point of departure the method used by Wilks [19] in his treatment of the problem of hits occurring under random matching of two decks of 25 cards each, namely a target deck with composition $(5^5)$ and a matching deck with composition $(s_i)$, $i = 1, 2, 3, \cdots, 5$, $\sum_i s_i = 25$. He showed that

$$(1) \quad \phi = \frac{1}{\begin{bmatrix} 25 \\ s_i \end{bmatrix}} (x_1 e^\theta + x_2 + \cdots + x_5)^5 (x_1 + x_2 e^\theta + x_3 + \cdots + x_5)^5 \cdots (x_1 + x_2 + \cdots + x_5 e^\theta)^5$$

where,

$$\begin{bmatrix} 25 \\ s_i \end{bmatrix} \equiv \frac{25!}{s_1! \, s_2! \cdots s_5!},$$

is a suitable generating function for obtaining the moments of the distribution. In fact, if we define an operator $K_{s_1 s_2 \cdots s_5}$ as

$$(2) \qquad K_{s_1 s_2 \cdots s_5} u \equiv \text{coefficient of } x_1^{s_1} x_2^{s_2} \cdots x_5^{s_5} \text{ in } u,$$

where $u = u(x_1, x_2, \cdots, x_5)$, and if $h$ denotes the number of hits, then for $r = 1, 2, \cdots, 5$,

$$(3) \qquad P(h = r) = \text{coefficient of } e^{r\theta} \text{ in } K_{s_1 s_2 \cdots s_5} \phi$$

And it is readily seen that

$$(4) \qquad E(h^p) = K_{s_1 s_2 \cdots s_5} \left. \frac{\partial^p \phi}{\partial \theta^p} \right|_{\theta = 0}.$$

Wilks' $\phi$ function involves a particular order for the target deck. If we are to generalize and obtain moments for matchings between more than two decks, it is obvious that we must devise a procedure which will, in the case of two decks, be perfectly symmetrical and not require that one deck be given a preferred status. In the case of two decks this is readily accomplished by the use of Kronecker deltas, and in the case of three or more decks by the use of obvious generalizations of these deltas.

For two decks of 25 cards each with compositions $(5^5)$ we need only let

$$\phi = (x_i\, y_j\, e^{\delta_{ij}\theta})^{25} \equiv \left( \sum_{i,j=1}^{5} x_i\, y_j\, e^{\delta_{ij}\theta} \right)^{25} \tag{5}$$

where $\delta_{ii} = 1$; $\delta_{ij} = 0$, $i \neq j$.
Then, if

(6) $K_{n_{11}n_{12}\cdots n_{15}\cdot n_{21}n_{22}\cdots n_{25}}\, u \equiv$ coefficient of $x_1^{n_{11}} x_2^{n_{12}} \cdots x_5^{n_{15}} y_1^{n_{21}} y_2^{n_{22}} \cdots y_5^{n_{25}}$ in $n$
where $u = u(x_1, x_2, \cdots, x_5, y_1, y_2, \cdots, y_5)$, it readily follows that

$$E(h^p) = \frac{K_{55555\cdot 55555} \left.\dfrac{\partial^p \phi}{\partial \theta^p}\right|_{\theta=0}}{K_{55555\cdot 55555}\, \phi \,\big|_{\theta=0}}. \tag{7}$$

More generally, for two decks of $n$ cards each, the cards being of $k$ types, and the decks having compositions $(n_{11}, n_{12}, \cdots, n_{1k})$, $(n_{21}, n_{22}, \cdots, n_{2k})$ respectively, we let

$$\phi = u^n \equiv (x_i\, y_j\, e^{\delta_{ij}\theta})^n \equiv \left( \sum_{i,j=1}^{k} x_i\, y_j\, e^{\delta_{ij}\theta} \right)^n. \tag{8}$$

The factors of $\phi$ are in one-to-one correspondence with the $n$ events of dealing a card from each of the two decks. The values which can be assumed by the subscripts $i$ and $j$ are in one-to-one correspondence with the $k$ types of cards. The symbol $x_i$ corresponds to the first deck, $y_j$ to the second, the subscripts $i$ and $j$ corresponding to the different types of cards in each deck. The expansion of $\phi$ consists of all products which can be formed by choosing one and only one pair $x_\alpha y_\beta$ from each factor of $\phi$ as a factor of the product. In forming any term of $\phi$, choosing $x_\alpha y_\alpha$ from any factor of $\phi$ corresponds to dealing a card of type $\alpha$ from both decks, and introduces $e^\theta$ into the coefficient of the term. Choosing $x_\alpha y_\beta$ from any factor corresponds to dealing a card of type $\alpha$ from the first deck, $\beta$ from the second. If $\alpha \neq \beta$, then, since $\delta_{ij} = 0$, $i \neq j$, $e^\theta$ is not introduced into the coefficient. Therefore in the coefficient of any term of $\phi$, $e^\theta$ will be raised to a power, say $s$, which is equal to the number of factors of $\phi$ from which pairs $x_\alpha y_\alpha$ have been chosen.

The total number of ways in which the term

$$x_1^{n_{11}} x_2^{n_{12}} \cdots x_k^{n_{1k}} y_1^{n_{21}} y_2^{n_{22}} \cdots y_k^{n_{2k}}$$

can arise is equal to the number of ways in which two decks of types $(n_{1i})$, $(n_{2j})$ respectively can be dealt, (where $(n_{1i}) \equiv (n_{11} n_{12} \cdots n_{1k})$ and similarly for $(n_{2j})$). But this is given by

$$\begin{aligned} K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\, \phi\, |_{\theta=0} &= K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}} \left( \sum_{i=1}^{k} x_i \right)^n \left( \sum_{i=1}^{k} y_i \right)^n \\ &= K_{n_{11}n_{12}\cdots n_{1k}} \left( \sum_{i=1}^{k} x_i \right)^n K_{n_{21}n_{22}\cdots n_{2k}} \left( \sum_{i=1}^{k} y_i \right)^n \\ &= \begin{bmatrix} n \\ n_{1i} \end{bmatrix} \begin{bmatrix} n \\ n_{2j} \end{bmatrix}. \end{aligned} \tag{9}$$

The coefficient of $e^{s\theta}$ in $K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi$ is the total number of ways in which the term $x_1^{n_{11}}x_2^{n_{12}}\cdots x_k^{n_{1k}}y_1^{n_{21}}y_2^{n_{22}}\cdots y_k^{n_{2k}}$ can be formed subject to the restriction that pairs $x_iy_j$ with $i = j$ are chosen from $s$ of the factors of $\phi$. But this is precisely the number of ways in which the two decks can be dealt so that there will be $s$ hits. Hence if, as above, $h$ is the number of hits, the probability that $h = s$, assuming all permutations in each deck to be equally likely, is given by

$$(10)\qquad P(h = s) = \frac{\text{coefficient of } e^{s\theta} \text{ in } K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi}{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi\,\big|_{\theta=0}}.$$

Since this is true for all values of $s$ it follows that

$$(11)\qquad E(h^p) = \frac{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\dfrac{\partial^p\phi}{\partial\theta^p}\Big|_{\theta=0}}{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi\,\big|_{\theta=0}}.$$

Since

$$\frac{\partial\phi}{\partial\theta}\Big|_{\theta=0} = nu^{n-1}\left[\sum_{i,j=1}^{k}\delta_{ij}x_iy_je^{\delta_{ij}\theta}\right]\Big|_{\theta=0} = n\left[\sum_{i=1}^{k}x_iy_ie^{\theta}\right]u^{n-1}\Big|_{\theta=0}$$
$$= n\left[\sum_{i=1}^{k}x_iy_i\right]\left(\sum_{i=1}^{k}x_i\right)^{n-1}\left(\sum_{j=1}^{k}y_j\right)^{n-1}$$

we have at once

$$E(h) = \frac{n}{\begin{bmatrix} n \\ n_{1i}\end{bmatrix}\begin{bmatrix} n \\ n_{2j}\end{bmatrix}}\sum_{i=1}^{k}K_{n_{11}n_{12}\cdots n_{1i-1}(n_{1i}-1)n_{1i+1}\cdots n_{1k}}\left(\sum_{i=1}^{k}x_i\right)^{n-1}$$
$$\cdot K_{n_{21}n_{22}\cdots n_{2i-1}(n_{2i}-1)n_{2i+1}\cdots n_{2k}}\left(\sum_{j=1}^{k}y_j\right)^{n-1}$$
$$(12)\qquad = \frac{n}{\begin{bmatrix} n \\ n_{1i}\end{bmatrix}\begin{bmatrix} n \\ n_{2j}\end{bmatrix}}\sum_{i=1}^{k}\left[\frac{(n-1)!}{n_{11}!\cdots n_{1i-1}!(n_{1i}-1)!n_{1i+1}!\cdots n_{1k}!}\right]$$
$$\cdot\left[\frac{(n-1)!}{n_{2i}!\cdots n_{2i-1}!(n_{2i}-1)!n_{2i+1}!\cdots n_{2k}!}\right]$$
$$= \sum_{i=1}^{k}\frac{n_{1i}n_{2i}}{n}.$$

It is an equally straightforward matter to show that

$$(13)\qquad E(h^2) = \sum_i\left[\frac{n_{1i}n_{2i}}{n} + \frac{n_{1i}(n_{1i}-1)n_{2i}(n_{2i}-1)}{n(n-1)}\right] + \sum_{i\neq j}\frac{n_{1i}n_{1j}n_{2i}n_{2j}}{n(n-1)}$$

and that

$$(14)\qquad \sigma_h^2 = \sum_i\left[\frac{n_{1i}n_{2i}}{n} - \frac{n_{1i}^2n_{2i}^2}{n^2} + \frac{n_{1i}^{(2)}n_{2i}^{(2)}}{n^{(2)}}\right] + \sum_{i\neq j}\frac{n_{1i}n_{1j}n_{2i}n_{2j}}{n^2(n-1)}.$$

Evidently any of the $n_{1i}$ and $n_{2j}$ may be zero, provided only that $\sum_{i=1}^{k} n_{1i} = \sum_{j=1}^{k} n_{2j} = n$. The case of two decks with unequal numbers of cards $m$ and $n$, $(m < n)$, is readily handled by substituting for the smaller deck one obtained by adding $n$-$m$ "blank" cards—that is, cards of any type not already appearing in either deck, as indicated by Greville [8], who however obtained his results by considering a preferred order for one of the decks.

EXAMPLE 1. In the case of the decks treated by Wilks [19], $n = 25$, $k = 5$, $n_{1i} = n_{2j} = 5$. Hence from (12)

$$E(h) = \sum_{i=1}^{5} \left\{\frac{5\cdot 5}{25}\right\} = 5,$$

and from (14)

$$\sigma_h^2 = \sum_{i=1}^{5} \left\{\frac{5\cdot 5}{25} - \frac{25\cdot 25}{(25)^2} + \frac{5\cdot 4\cdot 5\cdot 4}{25\cdot 24}\right\} + \sum_{\substack{i,j=1\\ i\neq j}}^{5} \frac{5\cdot 5\cdot 5\cdot 5}{(25)^2 24}$$

$$= \sum_{i=1}^{5} \frac{16}{24} + \sum_{\substack{i,j=1\\ i\neq j}}^{5} \frac{1}{24} = 4\frac{1}{6}.$$

EXAMPLE 2. Suppose we have two decks as shown by the scheme

| | Type of card 1 | 2 | 3 | 4 | 5 | Total of all types |
|---|---|---|---|---|---|---|
| No. in deck A | 5 | 7 | 8 | 0 | 0 | 20 |
| No. in deck B | 0 | 3 | 4 | 6 | 2 | 15 |

Here deck B has five fewer cards than deck A. Hence we must presume that there are six types of cards in all, and that the decks have the respective distributions (578000) and (034625). We then have at once

$$E(h) = \sum_{i=1}^{6} \frac{n_{1i} n_{2i}}{n} = \frac{1}{20}[0 + 3\cdot 7 + 4\cdot 8 + 0 + 0 + 0]$$

$$= 2.65$$

$$\sigma_h^2 = \sum_{i=1}^{6} \left\{\frac{n_{1i} n_{2i}}{n} - \frac{n_{1i}^2 n_{2i}^2}{n^2} + \frac{n_{1i}^{(2)} n_{2i}^{(2)}}{n^{(2)}}\right\} + \sum_{\substack{i,j=1\\ i\neq j}}^{6} \frac{n_{1i} n_{2i} n_{1j} n_{2j}}{n^2(n-1)}$$

$$= 2.65 - \frac{1}{400}\{3^2\cdot 7^2 + 4^2\cdot 8^2\} + \frac{1}{20\cdot 19}\{3\cdot 2\cdot 7\cdot 6 + 4\cdot 3\cdot 8\cdot 7\}$$

$$+ \frac{1}{400\cdot 19}\{3\cdot 7\cdot 4\cdot 8 + 4\cdot 8\cdot 3\cdot 7\}$$

**3. Matchings between three decks.** Let the three decks be of types $(n_{11}n_{12} \cdots n_{1q})$, $(n_{21}n_{22} \cdots n_{2q})$, $(n_{31}n_{32} \cdots n_{3q})$ respectively, with $\sum_{i=1}^{q} n_{1i} = \sum_{j=1}^{q} n_{2j} = \sum_{k=1}^{q} n_{3k} = n$, and consider the function

$$(15) \qquad \phi = \left[ \sum_{i,j,k=1}^{q} x_i y_j z_k e^{\delta_{ijk}\theta_{123}+\delta_{ij}\theta_{12}+\delta_{ik}\delta_{13}+\delta_{jk}\theta_{23}} \right]^n \equiv u^n,$$

where

$$(16) \qquad \delta_{iii} = 1, \qquad \delta_{ijk} = 0 \qquad i,\ j,\ k \text{ not all equal,}$$

and the other deltas are the usual Kronecker symbols.

Each factor of $\phi$ corresponds to one deal from each of the three decks. The symbols $x$, $y$, and $z$ correspond respectively to cards in the first, second, and third decks. The subscripts $i, j, k, = 1, 2, \cdots, q$ correspond to the types of cards—there being $q$ distinct types.

Choosing $x_\alpha y_\alpha z_\alpha$ from a factor of $\phi$ corresponds to a deal in which a card of type $\alpha$ is dealt from all three decks, and introduces $e^{\theta_{123}+\theta_{12}+\theta_{13}+\theta_{23}}$ into the coefficient of the corresponding term in the expansion of $\phi$. Similarly, choosing $x_\alpha y_\alpha z_\beta$, $\beta \neq \alpha$, corresponds to a hit between the first and second decks, and introduces $e^{\theta_{12}}$ into the coefficient. Similarly choosing $x_\alpha y_\beta z_\alpha$ introduces $e^{\theta_{13}}$; $x_\beta y_\alpha z_\alpha$ introduces $e^{\theta_{23}}$. Choosing $x_\alpha y_\beta z_\gamma$, $\alpha \neq \beta \neq \gamma \neq \alpha$ corresponds to a deal with no hits, and introduces no powers of $e$ into the coefficient, since all the $\delta$'s are zero.

Let $K_{n_{1i}\cdot n_{2j}\cdot n_{3k}}$ be defined by

$$(17) \quad K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} u \equiv \text{coefficient of } x_1^{n_{11}} \cdots x_q^{n_{1q}} y_1^{n_{21}} \cdots y_q^{n_{2q}} z_1^{n_{31}} \cdots z_q^{n_{3q}} \text{ in } u.$$

Then the coefficient of $e^{h_{123}\theta_{123}}$ in $K_{n_{1i}\cdot n_{2j}\cdot n_{3k}}\phi \mid_{\theta_{12}=\theta_{13}=\theta_{23}=\theta}$ is the number of ways in which the cards can be dealt so as to yield precisely $h_{123}$ triples, or hits between all three decks. Similarly the coefficient of $e^{h_{12}\theta_{12}}$ in $K_{n_{1i}\cdot n_{2j}\cdot n_{3k}}\phi \mid_{\theta_{12}=\theta_{13}=\theta_{23}=0}$ is the number of ways in which the cards can be dealt so as to yield precisely $h_{12}$ hits between the first and second decks, with corresponding results for the first and third $(h_{13})$ and second and third $(h_{23})$ decks.

By the same reasoning as before then, we have

$$(18) \qquad E(h_{123}^r) = \frac{K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} \left.\dfrac{\partial^r \phi}{\partial \theta_{123}^r}\right|_{\theta's=0}}{K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} \phi \mid_{\theta's=0}},$$

$$(19) \qquad E(h_{12}^r) = \frac{K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} \left.\dfrac{\partial^r \phi}{\partial \theta_{12}^r}\right|_{\theta's=0}}{K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} \phi \mid_{\theta's=0}},$$

with similar results for $h_{13}$ and $h_{23}$. And it is a straightforward matter to show that

$$(20)\quad E(h_{123}) = n \sum_{i=1}^{q} \left( \prod_{\alpha=1}^{3} \frac{n_{\alpha i}}{n} \right)$$

$$(21)\quad \begin{aligned} E(h_{123}^2) = n \sum_{i=1}^{q} \left( \prod_{\alpha=1}^{3} \frac{n_{\alpha i}}{n} \right) + n(n-1) \sum_{i=1}^{q} \left( \prod_{\alpha=1}^{3} \frac{n_{\alpha i}^{(2)}}{n^{(2)}} \right) \\ + n(n-1) \sum_{i,j=1,(i\neq j)}^{q} \left( \prod_{\alpha=1}^{3} \frac{n_{\alpha i} n_{\alpha j}}{n^{(2)}} \right). \end{aligned}$$

$$(22)\quad E(h_{12}) = \frac{1}{n^2} \sum_{i,k=1}^{q} n_{1i} n_{2j} n_{3k}$$

$$(23)\quad E(h_{13}) = \frac{1}{n^2} \sum_{j,k=1}^{q} n_{1k} n_{2j} n_{3k}$$

$$(24)\quad E(h_{23}) = \frac{1}{n^2} \sum_{i,j=1}^{q} n_{1i} n_{2j} n_{3j}$$

$$(25)\quad \begin{aligned} E(h_{12}^2) = \frac{1}{n^2} \sum_{i,k} n_{1i} n_{2i} n_{3k} + \frac{1}{n^2(n-1)^2} \Big[ \sum_{i,k} n_{1i}^{(2)} n_{2i}^{(2)} n_{3k}^{(2)} \\ + \sum_{i,\, k\neq r} n_{1i}^{(2)} n_{2i}^{(2)} n_{3k} n_{3r} + \sum_{k,\, i\neq l} n_{1i} n_{1l} n_{2i} n_{2l} n_{3k}^{(2)} \\ + \sum_{i\neq l,\, k\neq r} n_{1i} n_{1l} n_{2i} n_{2l} n_{3k} n_{3r} \Big] \end{aligned}$$

with corresponding results for other moments. It is understood each summation index takes values from 1 to $q$.

As before, if the decks do not all have the same total number of cards it is merely necessary to introduce one or more sets of "blank" cards. Thus we would replace decks with the compositions (57800), (03462), (00335) by hypothetical decks (5780000), (0346250), (0033509) and proceed as before.

EXAMPLE 3. For three decks of 25 cards, consisting of five of each of five kinds we have $n = 25$, $n_{\alpha i} = 5$, $\alpha = 1, 2, 3$, $i = 1, 2, \cdots, 5$. Hence

$$E(h_{123}) = 25 \sum_{i=1}^{5} \prod_{\alpha=1}^{3} \frac{5}{25} = 1$$

$$\begin{aligned} E(h_{123}^2) &= 25 \sum_{i=1}^{5} \left(\frac{5}{25}\right)^3 + 25\cdot 24 \sum_{i=1}^{5} \left(\frac{5\cdot 4}{25\cdot 24}\right)^3 + 25\cdot 24 \sum_{\substack{i,j=1\\ i\neq j}}^{5} \left(\frac{5^2}{25\cdot 24}\right)^3 \\ &= 1\frac{47}{48} \end{aligned}$$

$$\sigma_{h_{123}}^2 = \frac{47}{48}.$$

$$\begin{aligned} E(h_{12}) &= \frac{1}{(25)^2} \sum_{i,k=1}^{5} 5^3 \\ &= 5 \end{aligned}$$

$$E(h_{12}^2) = \frac{1}{(25)^2} \sum_{i,k=1}^{5} 5^3 + \frac{1}{(25)^2(24)^2}\left[\sum_{i,k=1}^{5} 5^3 4^3 + \sum_{\substack{i,k,r=1\\ k\neq r}}^{5} 5^4 4^2 + \sum_{\substack{i,l,k=1\\ i\neq l}}^{5} 5^5 4 + \sum_{\substack{i,l,k,r=1\\ i\neq l\\ k\neq r}}^{5} 5^6\right]$$

$$= 29\tfrac{1}{6},$$

$$\sigma_{h_{12}}^2 = 4\tfrac{1}{6}.$$

with similar results for $E(h_{13})$, $E(h_{23})$, $\sigma_{h_{13}}^2$, and $\sigma_{h_{23}}^2$.

**4. Generalization to any number of decks.** If the moments of the distribution of hits—doubles, triples, quadruples, ...—in matching any number of decks is desired, these can be obtained by using an obvious generalization of (15). Thus for four decks we would define $\delta_{iiii} = 1$, $\delta_{ijkl} = 0$, $i, j, k, l$ not all equal, and use

$$(26) \qquad \phi = \left[\sum_{i,j,k,l=1}^{q} x_i y_j z_k w_l e^{\delta_{ijkl}\theta_{1234}+\delta_{ijk}\theta_{123}+\delta_{ijl}\theta_{124}+\cdots+\delta_{ij}\theta_{12}+\cdots+\delta_{kl}\theta_{34}}\right]^n$$

However, it is evident that as the number of decks is increased the summations involved and the manipulation of the (generalized) $K$ operators rapidly become complicated.

**5. Application of our moment-generating technique to two-way contingency tables.** The moment-generating technique which we have discussed has wider applications than merely to matching problems. As an example of considerable interest we shall consider the contingency problem. Consider the array

$$(27) \qquad \begin{array}{c|c} n_{\alpha\beta\cdot} & n_{\alpha\cdot} \\ \hline n_{\cdot\beta} & n \end{array} \qquad \begin{array}{c} \alpha = 1, 2, \cdots, r \\ \beta = 1, 2, \cdots, s \\ \sum_{\alpha,\beta} n_{\alpha\beta} = \sum_{\alpha} n_{\alpha\cdot} = \sum_{\beta} n_{\beta\cdot} = n \end{array}$$

and also the function

$$(28) \qquad \phi = \prod_{\alpha=1}^{r} (x_\beta e^{\theta_{\alpha\beta}})^{n_{\alpha\cdot}} \equiv \prod_{\alpha=1}^{r} \left(\sum_{\beta=1}^{s} x_\beta e^{\theta_{\alpha\beta}}\right)^{n_{\alpha\cdot}}.$$

If $i$ and $j$ are particular values of $\alpha$ and $\beta$ respectively, then to the $i$-th row of the array corresponds the product $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$, consisting of $n_{i\cdot}$ identical factors $x_\beta e^{\theta_{i\beta}}$, one such factor corresponding to each of the $n_{i\cdot}$ elements in the row. To the $j$-th column of the array corresponds the $x_j$ which appears in each of the factors of $\phi$. To the $ij$-th cell of the array corresponds $e^{\theta_{ij}}$ which appears only in the factors $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$, and in each of these only as the coefficient of $x_j$.

The expansion of $\phi$ consists of all products which can be formed by taking as factors one and only one element $x_\beta e^{\theta_{\alpha\beta}}$ (not summed) from each factor of $\phi$. But taking $x_j e^{\theta_{ij}}$ from one of the factors $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$ of $\phi$ corresponds exactly to putting an element in the $ij$-th cell of a lattice such as (27). Hence every term in the expansion of $\phi$ corresponds to a particular distribution in such a lattice. Moreover, all terms of $\phi$ correspond to distributions in which the row totals are $n_{\alpha\cdot}$, for we must take $n_{i\cdot}$ elements from the product $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$. Further, those terms in which the $x_\beta$ appear in the particular product $x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}}$ correspond to distributions in which the column totals are $n_{\cdot 1}, n_{\cdot 2}, \cdots, n_{\cdot s}$, since choosing $n_{\cdot j}$ elements $x_j e^{\theta_{\alpha j}}$ corresponds to putting $n_{\cdot j}$ elements in the $j$-th column and some row of the lattice.

Expanding $\phi$ we obtain

$$\phi = \cdots + \left[ \sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}} \right] x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}} + \cdots \tag{29}$$

where the summation is over all partitions $(n_{\alpha 1} n_{\alpha 2} \cdots n_{\alpha s})$ of the $n_{\alpha\cdot}$ such that $(n_{1\beta} n_{2\beta} \cdots n_{r\beta})$ is also a partition of $n_{\cdot\beta}$. It is clear that since every set of values of the $n_{\alpha\beta}$ subject to the partition restrictions $\sum_\beta n_{\alpha\beta} = n_{\alpha\cdot}$, $\sum_\alpha n_{\alpha\beta} = n_{\cdot\beta}$ corresponds to a particular distribution of $n$ elements in the lattice (27), every particular product $\prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}$ corresponds to such a distribution, and represents the number of ways in which it can arise. Further, the total coefficient displayed (29), namely $\sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}$, represents the total number of ways in which distributions with row totals $n_{\alpha\cdot}$ and column totals $n_{\cdot\beta}$ can arise. Setting all the $\theta_{\alpha\beta} = 0$ we readily find

$$\begin{aligned} \sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} &= K_{n_{\cdot 1} n_{\cdot 2} \cdots n_{\cdot s}} \phi \,|_{\theta_{\alpha\beta}=0} = K_{n_{\cdot 1} n_{\cdot 2} \cdots n_{\cdot s}} (x_1 + x_2 + \cdots + x_s)^n \\ &= \begin{bmatrix} n \\ n_{\cdot\beta} \end{bmatrix}. \end{aligned} \tag{30}$$

Hence the probability of any particular distribution $\| n_{\alpha\beta} \|$ with fixed row totals $n_{\alpha\cdot}$ and fixed column totals $n_{\cdot\beta}$ is

$$P(\| n_{\alpha\beta} \| \,|\, n_{\alpha\cdot}, n_{\cdot\beta}) = \frac{\prod_\alpha \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}}{\begin{bmatrix} n \\ n_{\cdot\beta} \end{bmatrix}}. \tag{31}$$

*Moments of the* $n_{ij}$. Consider now the result of differentiating $\phi$ with respect to a particular $\theta_{\alpha\beta}$, say $\theta_{ij}$. We obtain

$$\frac{\partial\phi}{\partial\theta_{ij}} = \cdots + \sum_a n_{ij} \prod_\alpha \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}} x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}} + \cdots \tag{32}$$

where $\sum_a$ denotes summation over indices such that $\sum_{\alpha} n_{\alpha\beta} = n_{.\beta}$, $\sum_{\alpha \neq i} n_{\alpha j} + n_{ij} = n_{.j}$ $(\beta \neq j)$. Now $n_{ij} \leq \min (n_{i.}, n_{.j})$, but also $n_{ij}$ can never be less than $n_{.j} - (n - n_{i.})$. For $n_{.j} = n_{ij} + \sum_{\alpha \neq i} n_{\alpha j}$. Since the maximum value of $n_{\alpha j} \leq n_{\alpha .}$, the maximum value of $\sum_{\alpha \neq i} n_{\alpha j} \leq \sum_{\alpha \neq i} n_{\alpha .}$. Hence

$$n_{ij} = n_{.j} - \sum_{\alpha \neq i} n_{\alpha j} \geq n_{.j} - \sum_{\alpha \neq i} n_{\alpha .} = n_{.j} - (n - n_{i.}).$$

Therefore

$$\max (0, n_{.j} - n + n_{i.}) \leq n_{ij} \leq \min (n_{i.}, n_{.j}).$$

Accordingly, combining all the terms of (32) in which $n_{ij}$ has a particular value, $\gamma$, we have

$$\frac{\partial \phi}{\partial \theta_{ij}} = \cdots + \sum_{\gamma = \max(0, n_{.j} - n + n_{i.})}^{\min(n_{i.}, n_{.j})} \gamma \sum^*_{\sum_{\alpha} n_{\alpha\beta} = n_{.\beta}} \prod_{\alpha}^* \begin{bmatrix} n_{\alpha .} \\ n_{\alpha\beta} \end{bmatrix} \cdot e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}} x_1^{n_{.1}} x_2^{n_{.2}} \cdots x_s^{n_{.s}} + \cdots. \tag{33}$$

where $\Sigma^*$ denotes summation and $\Pi^*$ multiplication with $n_{ij} = \gamma$.

Since $\sum^*_{\sum_{\alpha} n_{\alpha\beta} = n_{.\beta}} \prod_{\alpha}^* \begin{bmatrix} n_{\alpha .} \\ n_{\alpha\beta} \end{bmatrix}$ is precisely the number of distributions $\| n_{\alpha\beta} \|$ for which $n_{ij} = \gamma$, it follows that

$$E(n_{ij} \mid n_{\alpha .}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial \phi}{\partial \theta_{ij}} \right|_{\theta_{\alpha\beta} = 0}. \tag{34}$$

Similarly it follows that

$$E(n_{ij}^p \mid n_{\alpha .}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial^p \phi}{\partial \theta_{ij}^p} \right|_{\theta_{\alpha\beta} = 0} \tag{35}$$

$$E(n_{ij}^p n_{kl}^q \mid n_{\alpha .}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial^{p+q} \phi}{\partial \theta_{ij}^p \, \partial \theta_{kl}^q} \right|_{\theta_{\alpha\beta} = 0} \tag{36}$$

where we may have $i = k$ or $i \neq k$, and $j = l$ or $j \neq l$.

By straightforward differentiation and reduction we find that for the array (27) with given marginal totals $n_{\alpha .}, n_{.\beta}$

$$E(n_{ij}) = \frac{n_{i.} n_{.j}}{n} \tag{37}$$

$$E(n_{ij}^2) = \frac{n_{i.}^{(2)} n_{.j}^{(2)}}{n^{(2)}} + \frac{n_{i.} n_{.j}}{n} \tag{38}$$

$$\sigma^2_{n_{ij}} = \frac{[n^2 - n(n_{i.} + n_{.j}) + n_{i.}n_{.j}]n_{i.}n_{.j}}{n^2(n-1)} \tag{39}$$

$$E(n_{ij}^3) = \frac{n_{i.}^{(3)}n_{.j}^{(3)}}{n^{(3)}} + 3\frac{n_{i.}^{(2)}n_{.j}^{(2)}}{n^{(2)}} + \frac{n_{i.}n_{.j}}{n} \tag{40}$$

$$E(n_{ij}^4) = \frac{n_{i.}^{(4)}n_{j.}^{(4)}}{n^{(4)}} + 6\frac{n_{i.}^{(3)}n_{.j}^{(3)}}{n^{(3)}} + 7\frac{n_{i.}^{(2)}n_{.j}^{(2)}}{n^{(2)}} + \frac{n_{i.}n_{.j}}{n}, \tag{41}$$

and if $i$ and $k$, $j$ and $l$ are distinct

$$E(n_{ij}^2n_{kj}^2) = \frac{n_{i.}^{(2)}n_{k.}^{(2)}n_{.j}^{(4)}}{n^{(4)}} + (n_{i.}^{(2)}n_{k.} + n_{i.}n_{k.}^{(2)})\frac{n_{.j}^{(3)}}{n^{(3)}} + \frac{n_{i.}n_{k.}n_{.j}^{(2)}}{n^{(2)}} \tag{42}$$

$$E(n_{ij}^2n_{il}^2) = \frac{n_{i.}^{(4)}n_{.j}^{(2)}n_{.l}^{(2)}}{n^{(4)}} + (n_{.j}^{(2)}n_{.l} + n_{.j}n_{.l}^{(2)})\frac{n_{i.}^{(3)}}{n^{(3)}} + \frac{n_{i.}^{(2)}n_{.j}n_{.l}}{n^{(2)}} \tag{43}$$

$$\begin{aligned} E(n_{ij}^2n_{kl}^2) = {} & \frac{n_{i.}^{(2)}n_{k.}^{(2)}n_{.j}^{(2)}n_{.l}^{(2)}}{n^{(4)}} + \frac{n_{i.}^{(2)}n_{k.}n_{.j}^{(2)}n_{.l}}{n^{(3)}} \\ & + \frac{n_{i.}n_{k.}^{(2)}n_{.j}n_{.l}^{(2)}}{n^{(3)}} + \frac{n_{i.}n_{k.}n_{.j}n_{.l}}{n^{(2)}} \end{aligned} \tag{44}$$

*Moments of the distribution of Chi Square.* For the array (27)

$$\begin{aligned} \chi^2 &= \sum_{\alpha,\beta} \frac{\left(n_{\alpha\beta} - \dfrac{n_{\alpha.}n_{.\beta}}{n}\right)^2}{\dfrac{n_{\alpha.}n_{.\beta}}{n}} \\ &= \sum_{\alpha,\beta}\left[\frac{n}{n_{\alpha.}n_{.\beta}}n_{\alpha\beta}^2 - 2n_{\alpha\beta} + \frac{n_{\alpha.}n_{.\beta}}{n}\right]. \end{aligned} \tag{45}$$

Hence, using the above results we can, theoretically, find all the moments of the exact distribution of $\chi^2$. It is not difficult to show that

$$E(\chi^2) = \frac{n}{n-1}(r-1)(s-1). \tag{46}$$

The value of $E[(\chi^2)^2]$ and the variance of $\chi^2$ were found by straightforward application of our methods and the results agreed with those given by Haldane [10].

The writer is indebted to Professor Wilks for helpful criticisms and suggestions.

## REFERENCES

[1] M. S. Bartlett, "Properties of sufficiency and statistical tests," *Proc. Roy. Soc.*, A, Vol. 160 (1937), pp. 268–282.

[2] Dwight Chapman, "The statistics of the method of correct matchings," *Amer. Jour. Psych.*, Vol. 46 (1934), pp. 287–298.

[3] Dwight Chapman, "The generalized problem of correct matchings," *Annals of Math. Stat.*, Vol. 6 (1935), pp. 85–95.

[4] J. A. GREENWOOD, "Variance of a general matching problem," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 56–59.

[5] J. A. GREENWOOD, "Variance of the ESP call series," *Jour. of Parapsychology*, Vol. 2 (1938), pp. 60–64.

[6] J. A. GREENWOOD, "The first four moments of a general matching problem," *Annals of Eug.*, Vol. 10 (1940), pp. 290–292.

[7] T. N. E. GREVILLE, "Exact probabilities for the matching hypothesis," *Jour. of Parapsychology*, Vol. 2 (1938), pp. 55–59.

[8] T. N. E. GREVILLE, "The frequency distribution of a general matching problem," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 350–354.

[9] J. B. S. HALDANE, "The mean and variance of Chi square, when used as a test of homogeneity, when expectations are small," *Biometrika*, Vol. 31 (1940).

[10] J. B. S. HALDANE, "The first six moments of Chi-square for an $n$-fold table with $n$ degrees of freedom when some expectations are small," *Biometrika*, Vol. 29 (1939), p. 389.

[11] J. B. S. HALDANE, "The exact value of the moments of the distribution of chi-square, used as a test of goodness of fit, when the expectations are small," *Biometrika*, Vol. 29 (1939), p. 133.

[12] E. V. HUNTINGTON, "A rating table for card matching experiments," *Jour. of Parapsychology*, Vol. 4 (1937), pp. 292–294.

[13] E. V. HUNTINGTON, "Exact probabilities in certain card-matching problems," *Science*, Vol. 86 (1937), pp. 499–500.

[14] SOLOMON KULLBACK, "Note on a matching problem," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 77–80.

[15] 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.

[16] T. E. STERNE, "The solution of a problem in probability," *Science*, Vol. 86, (1937) pp. 500–501.

[17] W. L. STEVENS, "Distribution of entries in a contingency table," *Annals of Eugenics*, Vol. 8 (1938), pp. 238–244.

[18] W. L. STEVENS, "Tests of significance for extra sensory perception data," Psychological Review, Vol. 46 (1938), pp. 142–150.

[19] S. S. WILKS, "Statistical aspects of experiments in telepathy," a lecture delivered to the Galois Institute of Mathematics, Long Island University, December 4, 1937.