ON CARD MATCHING

By T. W. Anderson
Princeton University

1. Introduction. Several authors have discussed the probability of obtaining a given number of matched pairs of cards under conditions of random pairing of two decks of arbitrary composition. The exact expression for this probability (equation (6)) is ordinarily too complicated for use in computing significance levels. This is especially true for certain practical applications. For example, in a square two-way contingency table in which the categories corresponding to rows are identical with those for columns, the sum of the entries in the diagonal cells has this distribution. Intuitively one would suspect that the distribution is asymptotically normal, as suggested by several authors. In the following section proof is given that the number of matched cards is asymptotically normally distributed when the total number of cards in each of the two decks approaches infinity with the proportion of cards in each suit of each deck remaining fixed. The form of the limiting distribution can then be used in computing approximate significance levels.

A problem of some interest to psychologists is that of determining whether an individual has matched two series of items better than could have been done "by chance"; for instance, whether a graphologist has matched personality descriptions with specimens of handwriting better than by chance. The problem can also be phrased in terms of card matching under random pairing of two identical decks each of a given number of different cards. This will be recognized as a classical problem of probability theory: Let tickets numbered from 1 to n be placed in a hat. If the tickets are drawn one by one from the hat, what is the probability that the number of the drawing will coincide with the number drawn a specified number of times? It is clear how the analagous problem of matching cards of three or more identical decks of a given number of different cards arises (e.g., matching appearance, personality, and handwriting). The latter part of the present paper is concerned with this problem. Battin [1] has displayed a generating function for the probability of obtaining a given number of matched cards between any number of decks of arbitrary composition. Battin's generating function is used to derive explicitly the probability of obtaining a specified number of matched cards and the moments of the distribution.

2. The Limiting Distribution of the Number of Matched Cards. In the ordinary card matching problem one is interested in the number of matchings when two decks, say D_1 and D_2, are paired randomly. Let D_1 consist of n_{11}, $n_{12}, \cdots, n_{1k}$ cards of suits $S_1, S_2, \cdots, S_k$, respectively, and let D_2 consist of $n_{21}, n_{22}, \cdots, n_{2k}$ cards of suits $S_1, S_2, \cdots, S_k$, respectively (any $n_{\alpha i}$ can be 0), where

$$\sum_{i=1}^{k} n_{1i} = \sum_{i=1}^{k} n_{2i} = n.$$

Let t_{ij} $(i, j = 1, 2, \cdots, k)$ be the number of pairings each involving a card from D_1 of suit S_i and a card from D_2 of suit S_j. It is easily seen that the probability of a set $\| t_{ij} \|$ under random pairing is the same as that associated with the entries $\| t_{ij} \|$ in a k by k contingency table [2] for which the row totals are fixed as $n_{11}, n_{12}, \cdots, n_{1k}$, and the column totals are fixed as $n_{21}, n_{22}, \cdots, n_{2k}$, i.e.

$$P(t_{ij}) = \frac{\prod_{i=1}^{k} n_{1i}! \prod_{j=1}^{k} n_{2j}!}{n! \prod_{i,j=1}^{k} t_{ij}!}. \tag{1}$$

The probability of obtaining h matchings is the same as that of the sum of diagonal terms in a square contingency table, i.e., $h = \sum_{i=1}^{k} t_{ii}$. In fact, in practical cases, the problem frequently arises in this manner: If two individuals each classify n objects into k categories, h is the number of objects on whose classification they agree.

The distribution (1) has essentially $(k - 1)^2$ variables since there are $2k - 1$ linear restrictions imposed on the t_{ij}. It is easy to verify that, for fixed $n_{1i}/n = m_{1i}$, say, and fixed $n_{2i}/n = m_{2i}$, say, the distribution (1) approaches the normal distribution in $(k - 1)^2$ linearly independent variables, as n approaches infinity. Let us substitute

$$x_{ij} = \frac{t_{ij} - nm_{1i}m_{2j}}{\sqrt{n}} \qquad (i, j = 1, 2, \cdots, k),$$

use Stirling's formula for each factorial in (1), and take the logarithm. The argument proceeds in a manner similar to the classical case of the limit of the binomial distribution.

Since there are imposed linear restrictions on the t_{ij}

$$\sum_{j=1}^{k} t_{ij} = n \cdot m_{1i} \qquad (i = 1, 2, \cdots, k),$$

$$\sum_{i=1}^{k} t_{ij} = n \cdot m_{2j} \qquad (j = 1, 2, \cdots, k),$$

there are also restrictions on the x_{ij}, namely,

$$\sum_{j=1}^{k} x_{ij} = \sum_{i=1}^{k} x_{ij} = 0.$$

Hence there are $(k - 1)^2$ linearly independent x_{ij}. If we choose x_{ij} $(i, j = 1, 2, \cdots, k - 1)$ as the linearly independent variables, the limiting probability element as n approaches infinity, is

$$\frac{1}{(2\pi)^{\frac{1}{2}(k-1)^2} \left(\prod_{i=1}^{k} m_{1i} \prod_{j=1}^{k} m_{2j} \right)^{\frac{1}{2}(k-1)}} e^{-\frac{1}{2}Q} \prod_{i,j=1}^{k-1} dx_{ij}, \tag{2}$$

where

$$Q \doteq \sum_{i,j=1}^{k} \frac{x_{ij}^2}{m_{1i} m_{2j}}$$

is written in terms of all the x_{ij} with the understanding that the linearly dependent variables are linear functions of the independent variables.

Now $h - E(h)$ is simply a linear combination of x_{ij}, namely,

$$h - E(h) = \sqrt{n} \sum_{i=1}^{k} x_{ii}.$$

Hence, it follows that

$$\frac{h - E(h)}{\sqrt{n}\,\sigma_h}$$

is asymptotically normally distributed with mean zero and variance unity. For large n, then, it is possible to use the normal distribution to approximate significance levels for h.

Of course, any other linear combination of the entries t_{ij} is asymptotically normally distributed. The quantity Q in (2) can be recognized as the Pearson χ^2 for contingency tables, and the above constitutes proof that it actually has the χ^2 distribution with $(k - 1)^2$ degrees of freedom.

3. Matchings between three or more decks. There are instances, such as the classification of n objects into k categories by 3 or more individuals, in which one is interested in the matchings of three decks or more. For any number of decks one can prove in a manner exactly analagous to §2 that the distribution of the number of matchings is asymptotically normal. Here the demonstration is indicated for three decks. Let us consider three decks D_α ($\alpha = 1, 2, 3$) with $n_{\alpha 1}, n_{\alpha 2}, \cdots, n_{\alpha k}$, cards of suits $S_1, S_2, \cdots, S_k$, respectively. Let t_{gij} be the number of triplets consisting of a card from S_g of D_1, a card from S_i of D_2, and a card from S_j of D_3 under random formation of triplets (i.e., laying down the three shuffled decks side by side).

The probability law of the set $\{t_{gij}\}$ can be derived by the consideration of the generating function,

$$(3) \quad \begin{aligned}(x_1 y_1 z_1 + x_1 y_1 z_2 + \cdots + x_1 y_2 z_1 + \cdots + x_2 y_1 z_1 + \cdots + x_k y_k z_k)^n \\ = \sum \frac{n!}{\prod_{g,i,j} t_{gij}!} \prod_{g,i,j} (x_g y_i z_j) t_{gij},\end{aligned}$$

where the summation extends over all the partitions $\{t_{gij}\}$ of n. The number of ways of deriving the set $\{t_{gij}\}$ is the coefficient of $\prod_{g,i,j} (x_g y_i z_j)^{t_{gij}}$, namely,

$$\frac{n!}{\prod_{g,i,j} t_{gij}!},$$

where $\sum_{i,j} t_{gij} = n_{1g}$, $\sum_{g,j} t_{gij} = n_{2i}$, and $\sum_{g,i} t_{gij} = n_{3j}$.

The total number of ways of getting the marginal totals n_{1g}, n_{2i}, and n_{3j} is the coefficient of $\prod_{g,i,j} x_g^{n_{1g}} y_i^{n_{2i}} z_j^{n_{3j}}$ in (3); that is, in

$$\begin{aligned}(\sum_{g,i,j} x_g y_i z_j)^n &= (\sum_g x_g)^n (\sum_i y_i)^n (\sum_j z_j)^n \\ &= \sum_{g,i,j} \frac{n!}{\prod_g n_{1g}!} \cdot \frac{n!}{\prod_i n_{2i}!} \cdot \frac{n!}{\prod_j n_{3j}!} x_g^{n_{1g}} y_i^{n_{2i}} z_j^{n_{3j}}.\end{aligned}$$

The probability of getting the set $\{t_{gij}\}$ is the ratio of these expressions,

$$\begin{aligned}P(t_{gij}) &= \frac{n!}{\prod_{g,i,j} t_{gij}!} \Big/ \left[\frac{n!}{\prod_g n_{1g}!} \cdot \frac{n!}{\prod_i n_{2i}!} \cdot \frac{n!}{\prod_j n_{3j}!}\right] \\ &= \frac{\prod_g n_{1g}! \cdot \prod_i n_{2i}! \cdot \prod_j n_{3j}!}{(n!)^2 \prod_{g,i,j} t_{gij}!}\end{aligned} \tag{4}$$

This formula is analagous to (1) and, indeed, reduces to (1) for $n_{31} = n$, $n_{3j} = 0$ ($j = 2, 3, \cdots, k$). This is the probability associated with a three-way contingency table (k by k by k). For a contingency table, k by l by m, this probability would be (4) with the limits on g of 0 and k; on i, 0 and l; and on j, 0 and m.

For fixed values of the ratios $n_{\alpha i}/n = m_{\alpha i}$ ($\alpha = 1, 2, 3$; $i = 1, 2, \cdots, k$), say, the $k^3 - 3k + 2$ linearly independent variates in the set $\{t_{gij}\}$ are asymptotically normally distributed. To demonstrate this, substitute

$$x_{gij} = \frac{t_{gij} - nm_{1g} m_{2i} m_{3j}}{\sqrt{n}} \qquad (g, i, j = 1, 2, \cdots, k)$$

into (4) and use Stirling's approximation. There are $3k - 2$ independent linear restrictions on the x_{gij}, namely,

$$\sum_{i,j=1}^{k} x_{gij} = \sum_{g,j=1}^{k} x_{gij} = \sum_{g,i=1}^{k} x_{gij} = 0.$$

Therefore, there are $k^3 - 3k + 2$ x's which are unrestricted. Using these variables, we find that the limiting probability element of these x_{gij} is

$$\frac{1}{(2\pi)^{\frac{1}{2}(k^3-3k+2)} (\prod_g m_{1g} \prod_i m_{2i} \prod_j m_{3j})^{\frac{1}{2}(k^2-1)}} e^{-\frac{1}{2}Q} \prod dx_{gij}, \tag{5}$$

where

$$Q = \sum_{g,i,j=1}^{k} \frac{x_{gij}^2}{m_{1g} m_{2i} m_{3j}},$$

and the product of differentials is of $k^3 - 3k + 2$ variables. The number of matched triplets u, say, is the sum $\sum_{i=1}^{k} t_{iii}$, and we have

$$\frac{u - E(u)}{\sqrt{n}} = \sum_{i=1}^{k} x_{iii}.$$

From these facts it follows that $\frac{u - E(u)}{\sqrt{n}}$ is asymptotically normally distributed.

The above results may be easily generalized. In a q-way contingency table with fixed marginal totals $n \cdot m_{\alpha i}$ ($\alpha = 1, 2, \cdots, q$; $i = 1, 2, \cdots, k$), the probability of a set $\{t_{gi \ldots j}\}$ is

$$\frac{\prod_{\alpha=1}^{q} \prod_{i=1}^{k} (n \cdot m_{\alpha i})!}{(n!)^{q-1} \prod_{g, i, \cdots, j=1}^{k} t_{gi \cdots j}!}.$$

The entries minus their respective means and divided by $\sqrt{n}$, namely,

$$x_{gi \cdots j} = \frac{t_{gi \cdots j} - n m_{1g} m_{2i} \cdots m_{qj}}{\sqrt{n}}$$

are asymptotically normally distributed according to

$$(2\pi)^{-\frac{1}{2}(k^q - qk + q - 1)} \left(\prod_{\alpha=1}^{q} \prod_{i=1}^{k} m_{\alpha i} \right)^{-\frac{1}{2}(k^{q-1} - 1)} e^{-\frac{1}{2}Q},$$

where

$$Q = \sum_{g, i, \cdots, j=1}^{k} \frac{x_{gi \cdots j}^2}{m_{1g} m_{2i} \cdots m_{qj}}.$$

The generalization of Pearson's χ^2, namely Q, has the χ^2-distribution with $k^q - qk + q - 1$ degrees of freedom. Finally,

$$s = \sum_{i=1}^{k} t_{ii \cdots i},$$

the number of matched q-tuplets, under random formation of q-tuplets is asymptotically normally distributed.

4. Matching cards of identical decks, each of n different cards. The probability of obtaining a given number of pairs of matched cards under random pairing of two identical decks each of n different cards has been derived by Chapman [3] by a straightforward method and, of course, the solution of the classical problem mentioned in the introduction is this probability. Another technique involving the use of the general expression for the number of matchings of two decks of arbitrary composition can be easily generalized to three or more decks.

Before discussing this method, let us derive this general expression first by the use of the generating function discussed by Battin. Consider the multinomial

$$(x_1y_1e^{\theta} + x_1y_2 + \cdots + x_2y_1 + x_2y_2e^{\theta} + \cdots + x_ky_ke^{\theta})^n.$$

The coefficient of $e^{h\theta}x_1^{n_{11}} \cdots x_k^{n_{1k}}y_1^{n_{21}} \cdots y_k^{n_{2k}}$ (where k is the number of suits; n_{1i} the number of cards of suit S_i in the first deck; n_{2i} the number of cards of suit S_i in the second deck; and $n = \Sigma n_{1i} = \Sigma n_{2i}$) is the number of ways the cards may be arranged so that there are h matchings. After expanding the multinomial

$$[\sum_i x_i y_i e^{\theta} + (\sum_i x_i)(\sum_j y_j) - \sum_i x_i y_i]^n$$

in powers of x_i and y_i, taking the proper coefficient, and dividing by the total number of ways the cards can be arranged, one arrives at the probability law of h [4],

$$P(h) = \frac{\prod_i n_{1i}! \prod_j n_{2j}!}{(n!)^2} \sum_{g=0}^{n-h} (-1)^{n-h-g} \binom{n}{h}\binom{n-h}{g} T_g, \tag{6}$$

where

$$T_g = \sum \frac{(g!)^2(n-g)!}{\prod_{i=1}^{k} [(n_{1i} - s_i)!\,(n_{2i} - s_i)!\,s_i!]}, \tag{7}$$

where the summation is extended over all s_i, satisfying the following conditions:

$$\Sigma s_i = n - g, \qquad n_{1i} - s_i \geq 0, \qquad n_{2i} - s_i \geq 0, \qquad s_i \geq 0$$
$$(i = 1, 2, \cdots, k).$$

From (6) one can easily derive the distribution of the number of matchings when two identical decks of n different cards are randomly paired. Let $n_{1i} = 1$, $n_{2i} = 1$, and $n = k$. Then T_g as defined in (7) is

$$T_g = \sum \frac{(g!)^2(n-g)!}{(0!0!1!)^{n-g}(1!1!0!)^g} = \frac{n!}{g!(n-g)!}(g!)^2(n-g)!$$

for s_i can equal 0 or 1 and there are ${}_nC_g$ choices of the 0's. Hence, we find the probability of the number of matchings v to be

$$P(v) = \frac{1}{v!}\sum_{j=0}^{n-v} \frac{(-1)^j}{j!}. \tag{8}$$

This result has been given by Chapman [3]. It is, in fact, a classical probability law.

The moment generating function is

$$\varphi(\theta) = \sum_{v=0}^{n}\sum_{j=0}^{n-v} \frac{(-1)^j e^{v\theta}}{v!\,j!} = \sum_{g=0}^{n} \frac{(e^{\theta} - 1)^g}{g!}.$$

From this expression it is easy to verify that

$$E(v) = 1, \qquad \sigma_v^2 = 1, \qquad E(v^{(r)}) = 1 \ (r \leq n).$$

It is interesting to observe that as n approaches infinity, the moment generating function approaches

$$\sum_{g=0}^{\infty} \frac{(e^\theta - 1)^g}{g!} = e^{(e^\theta - 1)}. \tag{9}$$

It therefore follows that the limiting form of the distribution is the Poisson distribution with parameter unity, namely,

$$\frac{e^{-1}1^x}{x!} = \frac{1}{e}\frac{1}{x!}. \tag{10}$$

If one writes the moment generating function as

$$\varphi(\theta) = \sum_{g=0}^{n} \frac{\left(\frac{\theta}{1!} + \frac{\theta}{2!} + \cdots\right)^g}{g!} \tag{11}$$

one can see that the first n powers of θ in (9) are the same as in (11). Hence, the first n moments of the distribution (8) are the same as those of the Poisson distribution (10). In particular it is interesting to observe that in the random pairing of any two series, such as the serial numbers and order numbers in the Selective Service drawing, the expected number of matchings is exactly 1.

In applications of this method of matching (e.g., matching individuals and handwriting), the experiment may be repeated several times. It would be desirable, therefore, to have the probability law of the mean of a sample. The exact distribution, however, is too complicated to use. It follows from the central limit theorem that the mean of a sample of N observations from this distribution is asymptotically normally distributed as $N \to \infty$. It can also be shown by using the moment generating function that if the observations are from distributions with different n (i.e., the i-th observation from a pair of decks of n_i cards, $n_i \geq 2$), the distribution of the mean of the sample is asymptotically normal.

Now let us consider the analogue for three decks of cards. The generating function [1] for the number of matchings of three cards, one from each of three decks of arbitrary composition as defined in §3 is

$$(x_1y_1z_1e^\theta + x_1y_1z_2 + \cdots + x_1y_2z_1 + \cdots + x_2y_1z_1 + \cdots + x_2y_2z_2e^\theta + \cdots + x_ky_kz_ke^\theta)^n.$$

The probability of obtaining t matched triplets found after expanding this expression is

$$P(t) = \frac{\prod_{\alpha=1}^{3}\prod_{i=1}^{k} n_{\alpha i}!}{(n!)^3} \sum_{g=0}^{n-t} \binom{n}{t}\binom{n-t}{g}(-1)^{n-t-g}T_g, \tag{12}$$

where

$$T_g = \sum \frac{(g!)^3(n-g)!}{\prod_{i=1}^{k}\left[\prod_{\alpha=1}^{3}(n_{\alpha i}-s_i)!\cdot s_i!\right]},$$

where

$$\sum_{i=1}^{k} s_i = n - g, \qquad s_i \geq 0,$$

$$n_{\alpha i} - s_i \geq 0 \quad (\alpha = 1, 2, 3; i = 1, 2, \cdots, k).$$

To specialize (12) for the case to be considered here, namely, three identical decks of n different cards each, we let

$$n_{\alpha i} = 1 \quad (\alpha = 1, 2, 3; i = 1, 2, \cdots, k),$$

$$n = k.$$

Then, observing that

$$T_g = (g!)^2 n!,$$

one finds that the probability of t matchings is

$$(13) \qquad P(t) = \frac{1}{n!t!}\sum_{g=0}^{n-t}\frac{(-1)^{n-t-g}g!}{(n-t-g)!} = \frac{1}{n!t!}\sum_{j=0}^{n-t}\frac{(-1)^j(n-t-j)!}{j!}.$$

The moment generating function is

$$(14) \qquad \varphi(\theta) = \frac{1}{n!}\sum_{t=0}^{n}\sum_{j=0}^{n-t}\frac{e^{t\theta}(-1)^j(n-t-j)!}{t!j!}$$

$$= \frac{1}{n!}\sum_{g=0}^{n}\frac{(n-g)!}{g!}(e^{\theta}-1)^g.$$

One can readily verify that

$$(15) \qquad E(t) = \frac{1}{n},$$

$$\sigma_t^2 = \frac{n^2 - n + 1}{n^2(n-1)}.$$

Since both $E(t)$ and σ_t^2 approach 0, as n approaches infinity, by Tchebycheff's inequality we can see that the probability approaches 1 that there will be no matched triplet as n increases without bound. As in the case of two decks, the result that the mean of a sample from this population is asymptotically normally distributed follows from the central limit theorem.

For the general case of q identical decks each of n different cards we can generalize (13), (14), and (15) immediately. First, let us note that the probability of s matched cards for q decks of arbitrary composition is

$$P(s) = \frac{\prod_{\alpha=1}^{q} \prod_{i=1}^{k} n_{\alpha i}!}{(n!)^q} \sum_{g=0}^{n-s} \binom{n}{s} \binom{n-s}{g} (-1)^{n-s-g} T_g,$$

where

$$T_g = \sum \frac{(g!)^q (n-g)!}{\prod_{i=1}^{k} \left[\prod_{\alpha=1}^{q} (n_{\alpha i} - s_i)! \cdot s_i! \right]},$$

where

$$\sum_{i=1}^{k} s_i = n - g, \qquad s_i \geq 0,$$

$$(n_{\alpha i} - s_i) \geq 0$$

$$(\alpha = 1, 2, \cdots, q; i = 1, 2, \cdots, k).$$

The probability of w, the number of matchings when each of the q decks consists of n different cards, is

$$P(w) = \frac{1}{(n!)^{q-2}} \sum_{j=0}^{n-w} \frac{(-1)^j [(n-w-j)!]^{q-2}}{j!}.$$

The moment generating function is

$$\frac{1}{(n!)^{q-2}} \sum_{g=0}^{n} \frac{[(n-g)!]^{q-2}}{g!} (e^\theta - 1)^g.$$

Finally, the mean and variance are

$$E(w) = \frac{1}{n^{q-2}},$$

$$\sigma_w^2 = \frac{n^{q-2}(n-1)^{q-2} + n^{q-2} - (n-1)^{q-2}}{n^{2(q-2)}(n-1)^{q-2}}.$$

5. Summary. Two distinct problems associated with card matching have been considered in this paper. In the first place it has been shown that the distribution of the number of matched pairs obtained under conditions of random pairing of two decks of arbitrary composition is asymptotically normal when the number of cards in each deck approaches infinity and the proportion of cards in each suit remains fixed. This demonstration was extended to the cases of matchings between three or more decks. The second problem treated in the present paper is concerned with the matchings between identical decks, each of n different cards. The probability law for the case of two decks was derived by

the use of a generating function. When n approaches infinity the limiting distribution was shown to be Poisson. The case of three or more decks was treated in similar manner, with the probability law and the moments given.

REFERENCES

[1] I. L. Battin, "On the problem of multiple matching", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 294–305.

[2] F. Yates, "Contingency tables involving small numbers and the χ^2-test", *Jour. Roy. Stat. Soc.* (*Supple.*), Vol. 1 (1934), pp. 217–235.

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

[4] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, (1943) pp. 208–213.