STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

S 317 (VP 22)

MUTUAL CHOICES

by

F. Göbel

October 1963

Mutual Choices

1. Introduction

Let a group of n persons be given. Each person chooses k other persons ($1 \leqslant k \leqslant n-2$) at random, i.e. such that the $\binom{n-1}{k}$ ways in which a person can choose, have the common probability $\binom{n-1}{k}^{-1}$; the persons choose independently of each other. We will consider the problem of determining the distribution of the number of mutual choices.

It was a remark in Hofstätter's booklet [1] that led to the formulation of this problem.[1] Hofstätter refers to his book [2] for the probabilistic background, but the treatment given there is limited to the derivation of the (binomial) distribution of the number of times that a given person is chosen.

It is immediately clear [2] that the expected value $\mathcal{E}\,\underline{w}$ of the number w of mutual choices is given by

$$\mathcal{E}\,\underline{w} = \binom{n}{2}\left(\frac{k}{n-1}\right)^2 = \frac{n\,k^2}{2(n-1)} \qquad (1)$$

In the greater part of this preliminary report, our considerations will be limited to the case k = 1.

2. The distribution of $\underline{w}$ for k = 1

On determining the distribution of $\underline{w}$ for k=1, we will make use of some well-known identities, which we will

---

1) p.112
2) Stochastic variables will be denoted by underlined letters.

recapitulate here. The notation is virtually the same as in Feller's book (Chapter IV in the second edition).

Let $A_1,\dots,A_i$ be an arbitrary collection of events, and let $\{i_1,\dots,i_r\}$ be any subset of $\{1,\dots,N\}$. We then define

$$p_{i_1,\dots,i_r} = P[A_{i_1},\dots,A_{i_r}] \tag{2}$$

$$S_r = \sum p_{i_1,\dots,i_r} \tag{3}$$

where the summation is over all r - tuples from $\{1,\dots,N\}$. Then each of the quantities S_r and the probability Q_r that exactly r of the events $A_1,\dots,A_n$ are realized, can be expressed in the other, as follows:

$$Q_r = \sum_{j\geqslant r} (-1)^{j+r} \binom{j}{r} S_j \tag{4}$$

$$S_r = \sum_{i\geqslant r} \binom{i}{r} Q_i \tag{5}$$

In order to apply these identities, we define $A_{(i,j)}$ as the event "persons i and j choose each other". (Hence a "p with r indices" will become a "p with r pairs of indices"). The probability that two given persons choose each other is $(n-1)^{-2w}$, and therefore, because of the independency, the p's with w pairs of indices are equal to $(n-1)^{-2w}$ provided all 2w indices are different. Obviously, if not all indices are different, p will be zero. S_w will therefore be equal to $(n-1)^{-2}$ times the number of ways in which w disjoint pairs may be chosen from the set of all persons:

$$S_w = \frac{n!}{2^w \; w! \; (n-2w)! \; (n-1)^{2w}} \tag{6}$$

Formula (4) can now be applied to yield the following expression for the distribution of $\underline{w}$:

$$\boxed{P[\underline{w} = w] = \frac{n!}{w!} \sum_{j \geq w} \frac{(-1)^{j+w}}{(j-w)! \; (n-2j)! \; 2^j (n-1)^{2j}}} \tag{7}$$

3. The limiting distribution of w for arbitrary k.

For k=1, the limiting distribution of $\underline{w}$ when n tends to infinity can be found from (5) and (6) as follows. According to (5), the quantities $r! \; S_r$ are just the factorial moments of $\underline{w}$, whereas from (6) we have

$$\lim_{n \to \infty} r! \; S_r = \frac{1}{2^r}$$

Hence, the factorial moments of the limiting distribution are $\frac{1}{2^r}$, i.e. the limiting distribution is Poisson with mean $\frac{1}{2}$.

For $k > 1$ it is no longer true that all non-zero p's with w indices are equal to each other. As a consequence, it is no longer feasible to give manageable formulae for S_w. (Even a seemingly simple case like n=7, k=3 leads to a total of about 150 different values of p's). However, it is still possible to find the limiting distribution of $\underline{w}$, because for every value of $\underline{w}$, there is only one value of p that makes a significant contribution in the limiting case, as will be shown now.

Suppose a_j given persons are involved in exactly j mutual choices ($j=0,\ldots, k$; $\sum j a_j = 2w$). In such a situation we have

$$p_{i_1 \dots i_w} = \prod_{j=0}^{k} \left\{ \frac{\binom{n-1-j}{k-j}}{\binom{n-1}{k}} \right\}^{a_j} = \mathcal{O}(n^{-2w})$$

Consider the persons that are involved in at least one mutual choice. This group consists of $a_1+\dots+a_k = n-a_o$ persons, who may be chosen in $\binom{n}{n-a_o}$ ways from the whole group. As $n-a_o = a_1+\dots+a_k \leqslant 2w$ is bounded, we have

$$\binom{n}{n-a_o} = \mathcal{O}(n^{n-a_o})$$

Finally, the number of ways in which the above-mentioned $a_1 + \dots + a_k$ persons can allocate their mutual choices is independent of n, and hence S_w is the sum of a number of terms that are $\mathcal{O}(n^{-2w+n-a_o})$, and <u>not</u>, as is easily verified $o(n^{-2w+n-a_o})$. The exponent is maximal when a_o is minimal, i.e. when $a_o=n-2w$, $a_1=2w$, $a_2=\dots= a_k = 0$, as is easily seen. Now the leading term in S_w can be determined in exactly the same way as was done for k=1, and we find

$$\lim_{n\to\infty} r!\, S_r = \left(\frac{k^2}{2}\right)^r \qquad (8)$$

Therefore, also for $k>1$ the limiting distribution of <u>w</u> is Poisson (with mean $\frac{k^2}{2}$)

4. The variance of w

For very low values of w it is still possible to determine S_w, for arbitrary values of k and n. For example, if one wants to calculate S_2, only two cases have to be considered:

1. A and B choose each other, and C and D choose each other.
2. A and B choose each other, and A and C choose each other.

When S_2 is known, σ^2 can be found from the formula

$$\sigma^2 = 2S_2 + S_1 - S_1^2,$$

and after some calculations one finds

$$\boxed{\sigma^2 = \frac{n\,k^2(n-k-1)^2}{2(n-1)^3}} \qquad (9)$$

This expression remains unchanged when k is replaced by n-k-1, as is proper.

5. Extension to cycles of length greater than 2. (k=1)

Beside mutual choices, which can be interpreted as cycles of length 2, one might also consider cycles of length 3 (A chooses B, B chooses C, C chooses A) or greater. If $\underline{m}_i$ denotes the number of cycles of length i[1], the following relations can be proved.

$$\mathcal{E}\underline{m}_i = \frac{n!}{(n-i)\; i\; (n-1)^i} \qquad (10)$$

$$P\left[\underline{m}_i = m\right] = \frac{n!}{m!} \sum_{j \geqslant m} \frac{(-1)^{j+m}}{(j-m)!\;(n-ij)!\;\left\{i(n-1)^i\right\}^j} \qquad (11)$$

$$\lim_{n\to\infty} \sum_{j \geqslant m} \frac{j!}{(j-m)!}\; P\left[\underline{m}_i = j\right] = \frac{1}{i^m} \qquad (12)$$

From (12) it follows that the limiting distribution of the number of cycles of lenght i is Poisson with mean $\frac{1}{i}$ $(i \geqslant 2)$ Hence, the number of persons involved in a cycle of length $i \geqslant 2$ is 1 on the everage in the limiting case.

---

1) It will be convenient to have both $\underline{m}_2$ and $\underline{w}$ as symbols for the number of mutual choices.

6. Persons that do not belong to a cycle (k=1)

The expected value, $\mathcal{E}\underline{m}_1$, of the number of persons that do not belong to a cycle, can immediately be found from (10):

$$\mathcal{E}\,\underline{m}_1 = n-\sum_2^n i\ \mathcal{E}\,\underline{m}_i = n-\sum_2^n \frac{n!}{(n-i)!(n-1)^i}$$

This expression can also be written as follows

$$\mathcal{E}\underline{m}_1 = \frac{n^2}{n-1} - \frac{n!}{(n-1)^n}\ \psi(n-1) \qquad (13)$$

where $\psi(n) = \sum_0^n \frac{n^j}{j!}$

It can be verified by partial integration that

$$\sum_0^n \frac{e^{-n}n^j}{j!} = 1-\frac{1}{\Gamma(n+1)}\int_0^n t^n e^{-t}dt$$

The right hand is asymptotically $\frac{1}{2} + \frac{1}{3\sqrt{2\pi n}} + \mathcal{O}(\frac{1}{n})$,

whence

$$\psi(n)\approx(\frac{1}{2} + \frac{1}{3\sqrt{2\pi n}})\ e^n.$$

From (13) we then have

$$\boxed{\mathcal{E}\ \underline{m}_1 = n-\sqrt{\frac{\pi n}{2}} + \mathcal{O}(1)} \qquad (14)$$

To determine the distribution of $\underline{m}_1$, we need the following lemmata.

Lemma 1. If m is the number of persons that do not belong to a cycle, then these persons can choose in $\gamma(m,n)$ ways, where γ is given by

$$\gamma(m,n) = n^m-m.n^{m-1}\ . \qquad (15)$$

<u>Proof:</u> Let A be the set of all persons that do not belong to a cycle, C is the set of the persons that do belong to a cycle, and B the subset of A of persons that choose somebody in C. Suppose A is not empty. As no cycles occur in A, neither B nor C can be empty. Let b be the number of elements in B. Then there are $\binom{m}{b}$ possibilities for the set B, and the persons belonging to it can make their choices in $(n-m)^b$ ways. In order to determine the number of ways in which the m-b persons belonging to A-B can choose, we notice that this question is equivalent to the original problem, A-B playing the role of A, and B that of C. For, the persons belonging to A-B choose nobody from C (or else b would be larger), and they choose without cycles. Hence we may apply induction [1]:

$$\gamma(m,n) = \sum_{b=1}^{m} \binom{m}{b} (n-m)^b \gamma(m-b,m) \qquad (16)$$

Now, by substituting (15) into the right hand side of (16), the desired expression for $\gamma(m,n)$ reappears. We still have to show the correctness of (15) for m=0 and all n. This, however, is trivial.

<u>Corrolary.</u> Cayley's formula for the number of rooted trees that can be formed with m points, can be obtained from (15) as follows. The number is the same as the number of ways in which m persons can choose, without cycles, if there is exactly one "outward" choice, i.e. $m.\gamma(m-1,m) = m^{m-1}$.

---

1) For this to be possible, it is also necessary that $m-b<m$ and $m<n$, which follows immediately from the non-emptiness of B and C.

<u>Lemma 2</u>. n persons who choose by cycles can do this in D_n ways, where D_n is n subfactorial [1)]

<u>Proof</u> Number the n persons 1,...,n. It is easily seen that there is a one-to-one correspondence between patterns consisting of cycles only and permutations without invariant elements, and the lemma follows.

Using the lemmata, and the fact that the group of n can be divided in two groups of m and n-m persons in $\binom{n}{m}$ ways, we find

$$P[\underline{m}_1=m] = (n^m - m.n^{m-1})\binom{n}{m} D_{n-m}(n-1)^{-n}$$

or, equivalently

$$\boxed{P[\underline{m}_1=m] = \binom{n-1}{m} n^m D_{n-m}(n-1)^{-n}} \tag{17}$$

As $\sum_{m=0}^{n-2} P[\underline{m}_1=m] = 1$, we have the corrolary

$$\sum_{m=0}^{n-2} \binom{n-1}{m} n^m D_{n-m} = (n-1)^n \tag{18}$$

The asymptotic behaviour of $\underline{m}_1$ may be determined as follows

$$P[\underline{m}_1 \leqslant m] = \sum_{0}^{m} \binom{n-1}{j} n^j D_{n-j}(n-1)^{-n}$$

If n-j is large for all j between 0 and m, i.e. if n-m is large, D_{n-j} can be approximated by $(n-j)!\, e^{-1}$. Substituting this, we find after some calculations

$$P[\underline{m}_1 \leqslant m] \approx \frac{(n-1)!}{e(n-1)^n} \frac{n^{m+1}}{m!}$$

---

1) A for our purpose, convenient way to define D_n is: the number of permutations of n elements 1,...,n such that element i is not in the i-th place, i=1,...,n.

For n and m large, Stirling's formula can be used:

$$P[\underline{m}_1 \leq m] \approx \left(\frac{n}{m}\right)^{m+\frac{1}{2}} e^{-(n-m)} .$$

By making the change of variable $\underline{m} = n-\underline{y}\sqrt{n}$, we find

$$P[\underline{y} \geq y] \approx \left(\frac{n}{n-y\sqrt{n}}\right)^{n-y\sqrt{n}+\frac{1}{2}} e^{-y\sqrt{n}} \approx e^{-y^2} ,$$

provided $y/\sqrt{n}$ tends to zero when n tends to infinity. To ensure that all approximations made are valid, it is sufficient to restrict the range of y by

$$n^{-\frac{1}{4}} \leq y \leq n^{-\frac{1}{4}} .$$

Indeed: $n-m=y\sqrt{n} \geq n^{\frac{1}{4}}$; $m=n-y\sqrt{n} \geq n-n^{\frac{3}{4}}$; $y/\sqrt{n} \leq n^{\frac{1}{4}}$.

Finally, by making the change of variable $\underline{y}^2=z$, we conclude that the asymptotic distribution of $\frac{(n-\underline{m}_1)^2}{n}$ is exponential (with mean 1).

7. Persons that are not chosen

Katz [3] gives the following formula for the distribution of the number $\underline{r}_0$ of persons that are not chosen (arbitrary k)

$$P[\underline{r}_0=r]= \sum_{j=r}^{n-1-k} (-1)^{r+j}\binom{j}{r}\binom{n}{j}\binom{n-j}{k}^{j}\binom{n-j-1}{k}^{n-j}\binom{n-1}{k}^{-n} .$$

For k=1, r=0 this expression reduces to

$$P[\underline{r}_0=0]= \sum_{j=0}^{n-2} (-1)^{j}\binom{n}{j}(n-j)^{j}(n-j-1)^{n-j}(n-1)^{-n} .$$

A second formula for this probability can be found using (17). For, though $\underline{r}_0$ is generally less then $\underline{m}_1$, we still have

$$\underline{r}_0 = 0 \Leftrightarrow \underline{m}_1 = 0,$$

as is easily seen. Therefore,

$$P[\underline{r}_0=0] = P[\underline{m}_1=0] = D_n(n-1)^{-n},$$

and we have the identity

$$\sum_{j=0}^{n-2} (-1)^j \binom{n}{j}(n-j)^j(n-j-1)^{n-j} = D_n,$$

which can also be proved in more direct way by using $D_n = nD_{n-1}+(-1)^n$ and Abel's generalized binomial formula.

Katz & Powell [4] have also considered a more general case (where the number of choices made by the i-th person is a given number depending on i), and give some very complicated formulae for this case.

8. Some numerical results

Table 1 shows the distribution of the number of mutual choices for k=1 and some values of n. The distribution for n=3,4,5,6,7 was calculated with formula 20, section 9. For n=20, formula (7) was used. Of course, $n=\infty$ corresponds to the Poisson distribution of the limiting case.

n ↓ \ m_2 →	0	1	2	3	4
3	.250	.750	-	-	-
4	.370	.593	.037	-	-
5	.434	.508	.059	-	-
6	.471	.459	.069	.001	-
7	.495	.428	.074	.002	-
20	.574	.337	.079	.009	.001
∞	.607	.303	.076	.013	.002

Table 1. $P[\underline{m}_2=m_2]$, k=1.

Table 2 gives the distribution of the number of mutual choices for $n=5,6,7,\infty$, and $k=2$. The computations were done using formulae (2),(3) and (4). As we have said before, these computations are complicated by the fact that not all non-zero p's are equal for $k>1$.
For low values of n, the distribution has a very small variance.

$n \downarrow$ \ $m_2 \rightarrow$	0	1	2	3	4	5	6	7
5	.003	.089	.401	.421	.085	.002	-	-
6	.014	.144	.386	.347	.101	.008	.000	-
7	.027	.176	.368	.308	.106	.014	.001	.000
∞	.135	.271	.271	.180	.090	.036	.012	.003

Table 2. $P\left[\underline{m}_2=m_2\right]$, $k=2$.

Table 3 shows the distribution of the number of persons who do not belong to a cycle, for $n=3,4,5,6,7$ and $k=1$. In spite of the fact that all rows in table 3 are increasing, the distribution has, for large values of n, its mode at approximately $n-\sqrt{n}$.

$n \downarrow$ \ $m_1 \rightarrow$	0	1	2	3	4	5
3	.250	.750	-	-	-	-
4	.111	.296	.593	-	-	-
5	.043	.176	.293	.488	-	-
6	.017	.084	.207	.276	.415	-
7	.007	.040	.116	.221	.257	.360

Table 3. $P\left[\underline{m}_1=m_1\right]$, $k=1$.

9. A different formula for the distribution of w (k=1)

Suppose n persons choose in such a way that m_i cycles of length i result ($i \geq 2$), and that m_1 persons do not belong to any cycle. Using lemma 1 (section 6) it is easily seen that the number of ways in which this can be done, is given by

$$f(\varphi)=\frac{n!\,(n^{m_1}-m_1 \cdot n^{m_1-1})}{(1^{m_1}2^{m_2}\ldots)(m_1!m_2!\ldots)} .$$

Hence

$$P[\underline{w}=w]=\sum_{\substack{\varphi(n)\\ m_2=w}} f(\varphi).\ (n-1)^{-n} \qquad (19)$$

By collecting terms that correspond to the same value of m_1, (19) can be reduced to

$$\boxed{P[\underline{w}=w]=\frac{n!}{(n-1)^n 2^w w!}\sum_{v}\frac{(n-v)n^{v-1}}{v!}\,\frac{T(n-v-2w)}{(n-v-2w)!}} \qquad (20)$$

where $$T(j) = j!\sum{}^{*}\frac{1}{(3^{m_3}4^{m_4}\ldots)(m_3!\,m_4!\ldots)} \qquad (21)$$

The summation $\sum^{*}$ in (21) is over all partitions of j for which $m_1=m_2=0$.

According to (21), T(j) can be interpreted as the number of permutations of j in which no cycles of lengths 1 or 2 occur. Applying the principle of inclusion and exclusion to the permutations of n without cycles of length 1, it can be shown that

$$T(n)= n!\sum_{i}\left(-\frac{1}{2}\right)^{i}\frac{D_{n-2i}}{i!(n-2i)!}$$

The first few non-trivial [1] values T(n) are:
T(6) = 160, T(7) = 1140, T(8) = 8988.

This approach, however, seems to be even less promising than the one chosen in section 2, when results for $k \geq 1$ are desired.

Acknowledgement. I thank W. van Zwet, J. Kriens and F.W. Steutel for many helpful discussions and constructive remarks.

---

1) If $n \leq 5$, $T(n)=D_n$, because the exclusion of the partitions with 1-cycles, or 2-cycles, or both, excludes all but the permutation consisting of one cycle of length n.

References

1. P.R. Hofstätter, Gruppendynamik, R.D.E. 38.

2. P.R. Hofstätter, Sozialpsychologie, 1956.

3. L. Katz, The distribution of the number of isolates in a social group, A.M.S. 23 (1952), p.271-276.

4. L. Katz & J.H. Powell, A.M.S. 28 (1957), p.442-448.