PSYCHOMETRIKA—VOL. 18, NO. 3
SEPTEMBER, 1953

NOTE ON MILLER'S "FINITE MARKOV PROCESSES IN PSYCHOLOGY"

RICHARD C. W. KAO
UNIVERSITY OF MICHIGAN

In his article "Finite Markov Processes in Psychology,"* G. A. Miller derived a least-squares "estimate" for a matrix of transitional probabilities. However, the mathematical proof is found to be invalid.

On page 158, Miller defined $\overline{N}$ by the equation

$$\overline{N} = N + C, \tag{19}$$

"where the elements of the matrix C are the corrections that must be added to the observed values in N to give the best estimate $\overline{N}$." He wished to determine $\overline{T}$, the "best" estimate of the transformation. From the definitions of $\overline{T}$, M, $\overline{N}$, N, and C, he argued that the following equation holds:

$$\overline{T}M = \overline{N} = N + C. \tag{20}$$

It is clear from this equation that $\overline{T}$ would be "best" in a trivial sense if C is assumed to be the zero matrix, i.e., $\overline{N} = N$. We shall show that Miller had in fact derived only this trivial estimate by means of his undefined mathematical technique†.

From equation (20) Miller obtained another expression for C:

$$C = -N + \overline{T}M. \tag{21}$$

For a least-squares solution, he argued that CC' must be a minimum. But this minimum cannot be obtained by simply "$\cdots$ putting the partial derivative with respect to $\overline{T}$ to zero:"

$$\frac{\partial}{\partial \overline{T}} CC' = MC' = 0, \tag{22}$$

for the operation of differentiating a function of a matrix with respect to the matrix has not been defined at all in this connection. It is obvious from equa-

Psychometrika, 1952, 17, 149-167.

†For a valid mathematical proof of a least-squares estimate in this connection, see Goodman's "A Further Note on 'Finite Markov Processes in Psychology.' " This issue, 245-248.

tion (21) that C' is as much a function of $\overline{T}$ as C. Hence, in differentiating the expression

$$CC' = (-N + \overline{T}M)C' = -NC' + \overline{T}MC',$$

one cannot assert that $\partial/\partial\overline{T}\ (-\ NC') = 0$. For then one would be asserting $\partial\overline{T}'/\partial\overline{T} = 0$, a result which is inconsistent with the undefined operation $\partial\overline{T}/\partial\overline{T} = 1$, in the case of symmetric matrices where one clearly has $\overline{T}' = \overline{T}$.* Hence, the first equality in equation (22) cannot be meaningful.†

The second equality in equation (22) in effect requires C to be a zero matrix. For in order that $MC' = 0$ for whatever M, C' or C must be zero. Granted that zero-divisors (i.e., $AB = 0$ for $A \neq 0$ and $B \neq 0$) are possible in dealing with matrices or rings in general, it is still true that the second equality in equation (22) holds only if C vanishes identically. This is so because no restriction has been placed upon M except that it be of order $a \times (n - 1)$, and consequently, one can choose a matrix M with all positive elements such that $MC' = 0$ only if $C' = C = 0$. In view of the fact that equation (22) is asserted to hold in general, we conclude that it does only if C is the zero matrix, from which the tautological nature of Miller's argument becomes clear.

*This argument is valid whether one chooses to use matrix or scalar notation for differentiation. Private communication with Professor Miller shows that he does the latter. In fact, he reasons that there are $(n - 1)$ equations in $(n - 1) + 2$ unknowns in C and $\overline{T}$, and the remaining two equations are obtained from setting the partial derivatives of $\Sigma_{i=1}^{n-1} c_i^2$ with respect to $\bar{t}_1$ and $\bar{t}_2$ to zero:

$$\frac{\partial}{\partial\bar{t}_j}\sum_{i=1}^{n-1} c_i^2 = 2\sum_{i=1}^{n-1} c_i \frac{\partial c_i}{\partial\bar{t}_j} = 0, \qquad (j = 1, 2);$$

whence $MC' = 0$ on substituting $\{\partial c_i/\partial\bar{t}_j\}$ by M. It is unclear why Professor Miller chooses one particular element in a matrix in his "matrix differentiation" and concludes that the whole matrix has thereby been minimized. In the case of two alternatives, various special assumptions lead to a matrix C of the form

$$\left\|\begin{matrix} \sum_{i=1}^{n-1} c_i^2 & -\sum_{i=1}^{n-1} c_i^2 \\ -\sum_{i=1}^{n-1} c_i^2 & \sum_{i=1}^{n-1} c_i^2 \end{matrix}\right\|.$$

If one minimizes, as Miller does, the upper left element with respect to $\bar{t}_j$, $j = 1, 2$, is one not simultaneously maximizing the upper right element with respect to the same thing? The fact remains that the elements of the matrix C are so functionally dependent on each other as not to permit the peculiar differentiation used by Miller. For a least-squares solution, it is sufficient to require only the elements on the principal diagonal of the matrix CC' to be a minimum. But this interpretation is a far cry from asserting that the whole matrix CC' is minimal. Indeed, the exact meaning of Miller's argument that CC' be a minimum is unclear.

†We note here the distinction between Miller's matrix differentiation and that of Dwyer and Macphail, Symbolic matrix derivatives. *Ann. math. Statist.*, 1948, 19, 517-534, esp. 523, 528-530.

Apart from all these considerations, Miller went on to substitute his equation (21) into equation (22) and obtained the expression

$$M(-N + \overline{T}M)' = -MN' + MM'\overline{T}' = 0.$$

By rearranging terms, he got what he called the "best" estimate of T:

$$\overline{T} = NM'(MM')^{-1}. \tag{23}$$

In case M, N are non-singular matrices—Miller's assumption that they be of order $a \times (n - 1)$ does not, of course, prevent a from being $(n - 1)$ —or M, N have inverses, we may show the tautological nature of Miller's argument by a different method.* We proceed to simplify equation (23) as follows:

$$\overline{T} = NM'(MM')^{-1} = NM'(M')^{-1}M^{-1} = NIM^{-1} = NM^{-1}$$

or

$$\overline{T}M = N,$$

which shows again that in equation (20) Miller had assumed $C = 0$ or $\overline{N} = N$. In case M, N are singular, this second proof would not apply; but the first still would. On the other hand, neither could Miller capitalize on the irrelevant fact that M, N are singular to prove the validity of his results. We have here something which is essentially a mathematical identity, the validity of which is independent of the choice of M and N. Hence, in order to show that equation (23) does not hold in general except in the trivial sense, it is sufficient to produce one counter-example where M, N are non-singular. For the logical denial of a proposition which reads, "for all x, $P(x)$ is true" is that "there exists one x such that $P(x)$ is false."

As a casual remark we note that setting the partial derivatives with respect to a variable to zero is only a necessary and not a sufficient condition for obtaining a minimum. For the latter, the second-order condition cannot be ignored. Granted that the experimental interpretation of CC' is such that a maximum is unlikely or even impossible, there is no assurance, on the other hand, that the stationary value obtained from using only the first-order condition is extremal at all. The mathematical problem of minimizing quadratic forms in general is not as simple as one may presume.

Finally, it is to be noted that the contention in this note refers only to Miller's "mathematical proof" of his best estimate for the matrix of transitional probabilities and not to his "psychological interpretation" of finite Markov processes.

Manuscript received 11/19/52

Revised manuscript received 4/4/53

*We note that this argument does not apply to Goodman's results, where M, N are column vectors.