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

A FURTHER NOTE ON "FINITE MARKOV PROCESSES IN PSYCHOLOGY"*

LEO A. GOODMAN
THE UNIVERSITY OF CHICAGO

In his interesting article "Finite Markov Processes in Psychology," G. A. Miller derived a "least-squares estimate" for a matrix of transitional probabilities (1). However, the mathematical proof seems to be unclear. Since this proof is considered invalid (2), we shall present a somewhat clearer version of the proof of this result. We shall also examine the general problem in some detail.

In the proof we shall assume that the reader is familiar with matrix notation, which enables a considerably shorter presentation. We shall follow the matrix conventions and the terminology adopted by Miller (1).

Let m_{ik} $(i = 1, 2, \cdots, a)$ represent the observed distribution on the kth trial $(k = 1, 2, \cdots, n)$. That is, m_{ik} is the proportion observed in the ith alternative quantity on the kth trial. There are a alternative quantities and n such trials. Let

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1,n-1} \\ m_{21} & m_{22} & \cdots & m_{2,n-1} \\ m_{31} & m_{32} & \cdots & m_{3,n-1} \\ \vdots & \vdots & & \vdots \\ m_{a1} & m_{a2} & & m_{a,n-1} \end{bmatrix}$$

be the $a \times (n - 1)$ matrix formed by placing in successive columns the distributions observed on successive trials, from trial 1 through trial $n - 1$. Following the notation adopted by Miller, we let t_{ij} be the transitional probability that an observation which is in the jth alternative quantity $(j = 1, 2, \cdots, a)$ at a given trial, will be in the ith alternative quantity

*This report was prepared in connection with research supported by the Office of Naval Research.

$(i = 1, 2, \cdots, a)$ on the following trial. We define the row vectors $T_i = [t_{i1}, t_{i2}, \cdots, t_{ia}]$ and the $a \times a$ transformation matrix

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1a} \\ t_{21} & t_{22} & \cdots & t_{2a} \\ t_{31} & t_{32} & \cdots & t_{3a} \\ \vdots & \vdots & & \vdots \\ t_{a1} & t_{a2} & & t_{aa} \end{bmatrix}.$$

We also define the row vectors $N_i = [m_{i2}, m_{i3}, m_{i4} \cdots, m_{in}]$ and $C_i = T_iM - N_i$. The problem is then to determine a matrix

$$\overline{T} = \begin{bmatrix} \bar{t}_{11} & \bar{t}_{12} & \cdots & \bar{t}_{1a} \\ \bar{t}_{21} & \bar{t}_{22} & \cdots & \bar{t}_{2a} \\ \bar{t}_{31} & \bar{t}_{32} & \cdots & \bar{t}_{3a} \\ \vdots & \vdots & & \vdots \\ \bar{t}_{a1} & \bar{t}_{a2} & & \bar{t}_{aa} \end{bmatrix}.$$

such that C_iC_i' is minimized for all values of i $(i = 1, 2, \cdots, a)$ when T is taken equal to $\overline{T}$. By the usual proof in the theory of least squares (cf. [3], 55), we see that C_iC_i' is minimized when T_i is taken equal to

$$\overline{T}_i = N_iM'(MM')^{-1} \qquad \text{(if } MM' \text{ is nonsingular)}.$$

Hence, the "least-squares estimate" is

$$\overline{T} = \begin{bmatrix} \overline{T}_1 \\ \overline{T}_2 \\ \overline{T}_3 \\ \vdots \\ \overline{T}_a \end{bmatrix}$$

which is, in fact,

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

where N is defined as

$$N = \begin{bmatrix} N_1 \\ N_2 \\ N_3 \\ \vdots \\ N_a \end{bmatrix}. \qquad Q.E.D.$$

We might wish our estimates $\bar{T}$ of the parameter T to have some of the same properties as the parameter T. For example, it may sometimes be desirable to require $[1]\bar{T} = [1]$, since we have that $[1]T = [1]$, where $[1]$ is the a-dimensional row vector $[1, 1, \cdots, 1]$. We shall now prove that the "least-squares estimate" $\bar{T}$ has this desirable property.

THEOREM: *We have that* $[1]\bar{T} = [1]$, *where* $\bar{T} = NM'(MM')^{-1}$.

PROOF: From the definition of N, we see that

$$[1]N = [1].$$

Hence, it is sufficient to prove that

$$[1]M'(MM')^{-1} = [1],$$

or

$$[1]M' = [1]MM'.$$

From the definition of M, we have that $[1]M = [1]$. Hence we have that

$$[1]MM' = [1]M'. \qquad Q.E.D.$$

We might also have obtained these results using the general regression methods presented by S. S. Wilks (4). The problem is that of estimating $a \times a$ parameters which are subject to a linear restraints. We shall be interested in minimizing $\sum_{i=1}^{a} C_i C_i'$ in order to obtain the "least-squares estimate." In other words, we wish to estimate a parameter which is a point in an $a(a - 1)$ dimensional space. Since $t_{ij} \geq 0$, the parameter will lie in a subset of this space. If we also wish our estimate $\bar{T}$ to lie in this same subset, the method of estimation is still quite straightforward but sometimes tedious. We first obtain the "least-squares estimate" $\bar{T}$. If this estimate lies in the subset (i.e., $\bar{t}_{ij} \geq 0$), then $\bar{T}$ is used to estimate T. If $\bar{T}$ is not included in the subset, then the appropriate estimate will lie on the boundary of the subset. We then use that estimate on the boundary of the subset which minimizes $\sum_{i=1}^{a} C_i C_i'$.

The numerical examples in (1) illustrate how this result is used in a learning experiment in a T-maze. As the author himself has pointed out, the least-squares fit described in (1) is not most efficient for Markov processes.

If the observed transitional proportions are available, they would clearly be more appropriate in the estimation of transitional probabilities.

REFERENCES

1. Miller, G. A. Finite Markov processes in psychology. *Psychometrika*, 1952, **17**, 149–167.
2. Kao, Richard. Note on Miller's "Finite Markov processes in Psychology." *Psychometrika*, 1953, **18**, 241–243.
3. Kempthorne, Oscar. The design and analysis of experiments. New York, John Wiley and Sons, Inc., 1952.
4. Wilks, S. S. Mathematical statistics. Princeton: Princeton Univ. Press, 1944.

Manuscript received 2/15/53

Revised manuscript received 3/14/53