A new computational method for the maximum likelihood solution in factor analysis is presented. This method takes into account the fact that the likelihood function may not have a maximum in a point of the parameter space where all unique variances are positive. Instead, the maximum may be attained on the boundary of the parameter space where one or more of the unique variances are zero. It is demonstrated that such improper (Heywood) solutions occur more often than is usually expected. A general procedure to deal with such improper solutions is proposed. The proposed methods are illustrated using two small sets of empirical data, and results obtained from the analyses of many other sets of data are reported. These analyses verify that the new computational method converges rapidly and that the maximum likelihood solution can be determined very accurately. A by-product obtained by the method is a large sample estimate of the variance-covariance matrix of the estimated unique variances. This can be used to set up approximate confidence intervals for communalities and unique variances.

1. Introduction

We shall present a new computational method for the maximum likelihood solution in factor analysis. The maximum likelihood solution was first given by Lawley [1940] and was further developed in two papers by him [1942, 1943]. A more condensed derivation of the method appears in a book by Lawley and Maxwell [1963]. Rao [1955] related the maximum likelihood method to canonical correlation analysis, and Bargmann [1957] related it to the problem of testing partial independence in multivariate statistical analysis. The estimates obtained by Rao and Bargmann, though derived from principles other than the maximum likelihood principle, satisfy Lawley's likelihood equations, thus constituting another set of maximum likelihood estimates. If identification conditions are imposed, these estimates all become identical.

All the above-mentioned derivations show that the maximum likelihood estimates are determined as the solution of two matrix equations. These

---

*The first part of this work was done at the University of Uppsala, Sweden and supported by the Swedish Council for Social Science Research. The second part was done at Educational Testing Service and supported by a grant (NSF-GB-1985) from National Science Foundation to Educational Testing Service.

The author is deeply indebted to Dr. D. N. Lawley whose contributions amounted nearly to coauthorship. The author also wishes to thank Mr. G. Gruvaeus for much valuable assistance in constructing the computer program.
equations cannot be solved algebraically; instead some iterative procedure has to be used, such as the procedures proposed by Lawley [1942], Rao [1955], Howe [1955], and Bargmann [1957]. In carrying out these iterative procedures, certain difficulties have been involved. The situation might best be characterized by some quotations.

About the procedure suggested by Lawley [1942], used by Emmett [1949], and investigated by Howe [1955], Lawley and Maxwell [1963] say "It has not been found possible to establish exact conditions under which the above procedure converges, but in practice this is usually the case. Convergence is, however, often very slow and, as Howe [1955] has pointed out, it is possible for differences between successive iterates to be extremely small and yet to be far from the exact solution. . . . It is possible to construct A-matrices for which the maximum likelihood method breaks down, either because there is no solution in terms of real numbers or because one or more of the $v_i$ are zero" (p. 15). "Experience has shown that in the absence of good initial approximations to the loadings, some difficulties can arise. A common occurrence, when three or more factors are postulated at the outset, is for one of the loadings to approach unity and then the method breaks down" (p. 20). Lord [1956] used the same procedure and had trouble finding initial estimates. "The original hypothesis of the author suggested that $m$ should be at least 9 for the 33-variable matrix analyzed. However, application of Lawley's method to the initial set of trial values for the factor loadings failed because the computations generated imaginary numbers. Extremely close initial approximations to the solution were necessary whenever $m$ was at all large" (p. 39). For one set of data Maxwell [1961] found that "After more than 1,100 iterations the loadings had still not converged and that for test 8 on the first factor was approaching unity. Under these conditions the estimate of the residual variance of this test was virtually zero, and since the iterative process involves weighting the trial loadings by the inverse of the residual variances the process now breaks down" (p. 55).

The procedure proposed by Rao was used by Browne [1965], who says "This computational procedure was found to converge far too slowly for use even on a computer the size of the I.B.M. 7094 which was used for all computation in the present study. Furthermore, it was found that it was possible for differences between corresponding values of the uniquenesses $u_i$ on successive iterations to be very small while these values were far from the correct solution" (p. 76). Rao's procedure was recently investigated by Hemmerle [1965]. With one of his test matrices he found that Rao's method "attained a residual sum of squares in 5 1/2 minutes that was less than Lawley's procedure achieved after 25 1/2 minutes" and, with the same test matrix using Rao's procedure, he found that "The communalities obtained after 870 iterations differed by less than .065 from those obtained after 470 iterations but by more than .315 from those obtained after 70 iterations."