[7] D.R. Fulkerson and O.A. Gross, *Incidence matrices and interval graphs*, Pacific J. Math. 15 (1965), pp. 835–855.
[8] F.R. Hodson, D.G. Kendall and P. Tautu (eds.), *Proc. of the Anglo-Romanian Conference of Mathematics in the Archaeological and Historical Sciences*, Edinburgh University Press, Edinburgh 1971.
[9] W. Kaluscha, *Zur Chronologie der Platonischen Dialoge*, Wiener Studien 25, 26 and 27 (1904).
[10] D.G. Kendall, *A statistical approach to Flinders Petrie's sequence dating*, Bull. Int. Statist. Inst. 40 (1963), pp. 657–680.
[11] —, *Incidence matrices, interval graphs, and seriation in archaeology*, Pacific J. Math. 28 (1969), pp. 565–570.
[12] —, *Some problems and methods in statistical archaeology*, World Archaeology 1 (1969), pp. 68–76.
[13] —, *Abundance matrices and seriation in archaeology*, Zeitschrift f. Wahrscheinlichkeitstheorie 17 (1971), pp. 104–112.
[14] —, *A mathematical approach to seriation*, Philos. Trans. Roy. Soc. London (A) 269 (1971), pp. 125–134.
[15] —, *Seriation from abundance matrices*, in [8] (1971), pp. 213–254.
[16] —, *Some data-analytic problems in archaeology and history*, Sci. and Hum. North Holland Publ. Comp. (1972), pp. 1371–1376.
[17] J.B. Kruskal, *Multidimensional scaling, I and II*, Psychometrica 29 (1964), pp. 1–27 and 28–42.
[18] —, *Multidimensional scaling in archaeology*, in [8] (1971), pp. 119–132.
[19] W.M.F. Petrie, *Sequences in prehistoric remains*, J. Anthropol. Inst. 29 (1899), pp. 259–301.
[20] W.S. Robinson, *A method for chronologically ordering archaeological deposits*, American Antiquity 16 (1951), pp. 293–301.
[21] R. Sibson, *Some thoughts on sequencing methods*, in [8] (1971), pp. 263–266.
[22] —, *Local order multidimensional scaling*, Proc. 39 Session of I.S.I. (1973) (inv. paper).
[23] R.N. Shepard, *The analysis of proximities: Multidimensional scaling with an unknown distance function, I and II*, Psychometrica 27 (1962), pp. 125–139 and 219–246.
[24] E.M. Wilkinson, *Archaeological seriation and the travelling salesman problem*, in [8] (1971), pp. 276–283.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

USE OF MATRIX APPROXIMATION IN STATISTICS

L. C. A. CORSTEN

Department of Mathematics, Agricultural University, Wageningen, The Netherlands

In trying to approximate an n by p array Y of data by a matrix C of rank k, one may want to minimize the approximation error matrix E in some sense. Minimization of $|Ex|/|x|$ for all x, or of $|E'y|/|y|$ for all y suggests minimization of all eigenvalues of $E'E$ or EE' simultaneously. This minimization can be attained by the canonical decomposition $V\Lambda U'$ of Y, where $U'U = V'V = I_r$, r is the rank of Y, and λ_j is the positive square root of the jth largest characteristic value of $Y'Y$ or YY' $(j = 1, \ldots, r)$.

The required approximation C of $Y = V\Lambda U' = \sum_{j=1}^{r} \lambda_j v_{*j} u'_{*j}$ obtained by suppressing the last $r-k$ terms in this sum equals $V_k \Lambda_k U'_k$. In this way, $Y'Y = U\Lambda^2 U'$ will be approximated by $(U_k\Lambda_k)(U_k\Lambda_k)'$, and any symmetric matrix S by $U_k \Lambda_k^* U'_k$ where Λ_k^* contains k characteristic values of S in non-increasing order of their absolute value. The approximation C of Y may be written as AB' where $A = V_k = YU_k\Lambda_k^{-1}$ and $B = U_k\Lambda_k$, and the rank k approximation of $Y'Y$ equals BB'. When each row of Y contains a multivariate observation at a corresponding individual and each column corresponds to a component of such a multivariate vector, each column a_{*j} of A may be conceived of as the set of n values of a new characteristic (a factor), each row a_{i*} of A as a set of factor scores for the ith individual, and each row b_{j*} of B as the set of factor loadings for the jth component of the multivariate observations.

As the columns of A are orthonormal the structure of the columns of C approximating those of Y may be visualized by means of their coordinates b_{j*} in k-space, the inner products between those columns approximating those between y_{*j}. Each row a_{i*} of factor scores may, likewise in k-space, visualize the mutual position of the individuals, and the inner product between a_{i*} and b_{j*} is the approximation of c_{ij}.

Approximation of Y by a rank k matrix C plus $1\beta'_1$ and (or) $\beta_2 1'$ where β_1 is a p-vector and β_2 an n-vector is a useful modification of the first situation. Not only the situation that $n^{-1}Y'Y$ is a covariance matrix Σ is covered now, but it also leads to an exact test on the presence of a multiplicative term in a two-way analysis of variance table in addition possibly to row effect and (or) a column effect.

Since the approximation is scale dependent, the criterion $(Ex, Ex)/(x, x)$ may be replaced with $(Ex, Ex)/(x, S_2 x)$ where S_2 is a relevant positive definite matrix, e.g. an error covariance matrix in a multivariate regression situation or a covariance matrix to which the matrix of interest has to be compared. This modified criterion is equivalent to $z'L'E'ELz/z'z$ where $L'L = S_2^{-1}$ and L an upper triangular matrix. Now the approximation C of Y will be AB' with $A = YL'U_k\Lambda_k^{-1}$ as before, and $B = M'U_k\Lambda_k$ where $L'M = I_p$ and M is a lower triangular matrix. Then $Y'Y$ will be approximated simultaneously by BB' again.

When a relevant matrix S_2 is not available one may find, given the covariance matrix $\Sigma = n^{-1}Y'Y$, a diagonal scaling matrix K such that an approximation of $K\Sigma K - I$ induced by the approximation $U_k\Lambda_k^2U_k'$ of $K\Sigma K$, namely $U_k(\Lambda_k^2 - I)U_k'$, will be perfect in the diagonal elements. This idea borrowed from factor analysis is directed towards equalizing by a suitable rescaling, the variance approximation errors, and so the rescaled specific variances are set equal to one beforehand. Finding such a K is exactly what happens in maximum likelihood factor analysis, where K^{-2} is the required matrix of specific variances.

Now with $\Lambda_k^2 - I = \tilde{\Lambda}_k^2$ one may choose $C = AB'$ with $A = n^{-1/2}YKU_k\tilde{\Lambda}_k^{-1}$ and $B = K^{-1}U_k\tilde{\Lambda}_k$, where A contains factor scores in agreement with Bartlett's recommendation.

In the case where Y is a contingency table N, it is preferable to rescale N to $R^{-1/2}NK^{-1/2}$ where R is a diagonal matrix of row totals of N, and K similarly of column totals. In the canonical decomposition $\sum_{j=1}^{r} \lambda_j v_{*j}u_{*j}'$ all λ_j are at most 1, while λ_1 equals one, $\lambda_1 v_{*1}u_{*1}'$ representing square roots of expected frequencies under independence. The statistic $n\lambda_2^2$ may be used for testing independence, its asymptotic null distribution being known. The rank k approximation of $N - R^{1/2}v_{*1}u_{*1}'K^{1/2}$, i.e. $R^{1/2}V_k\Lambda_kU_kK^{1/2}$ may serve the study of dependence. The present paper has been published recently:

Reference

[1] L. C. A. Corsten, *Matrix approximation, a key to application of multivariate methods*, Proceedings of the 9th Biometric Conf. Boston 1976, Vol. 1, pp. 61–77.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

ON BASIC CONCEPTS OF MATHEMATICAL STATISTICS

N. N. ČENCOV

Institute of Applied Mathematics, Moscow, U.S.S.R.

The basic essential concepts and structures have been formalized rather intensively for the last ten years. In this paper(1) the geometric approaches, which naturally arise in the analysis of statistical concepts, are considered.

Let (Ω, S) be a measurable space of elementary events, let $\{P_\theta\}$ be a family of a probability distribution, a priori possible, over (Ω, S), and let $(\mathscr{E}, B)$ be a measurable space of decisions.

Any of Wald's statistical decision rules [9], both determinated and randomized, can be written as a transition probability distribution $\Pi(\omega; d\varepsilon)$ from Ω onto $(\mathscr{E}, B)$. Thus if we use the rule Π, our decision will be distributed according to the law

(1) $$Q_\theta = P_\theta\Pi: \quad Q_\theta(\cdot) = \int_\Omega P_\theta(d\omega)\Pi(\omega; \cdot).$$

The value of the parameter θ at which the observations occur is unknown to the observer; he only knows that the observed P belongs to $\{P_\theta\}$. Therefore, all a priori conclusions about the quality of the decision rule Π are based on the properties of the families $\{P_\theta\Pi\}$.

It is natural to say that the families $\{P_\theta^{(i)}\}$ on $(\Omega^{(i)}, S^{(i)})$, $i = 1, 2$, parametrized by the same parameter $\theta \in \Theta$ are *equivalent in the theory of statistical inference* if, for any space of decision $(\mathscr{E}, B)$ and for any rule $\Pi^{(i)}(\omega^{(i)}; d\varepsilon)$, $i = 1, 2$, which leads to the family of laws $P_\theta^{(i)}\Pi^{(i)} = Q_\theta$ there exists a rule $\Pi^{(j)}(\omega^{(j)}; d\varepsilon)$, $j = 2, 1$, which leads to the same family $\{Q_\theta\}$:

(2) $$P_\theta^{(j)}\Pi^{(j)} = Q_\theta = P_\theta^{(i)}\Pi^{(i)}, \quad \forall\theta \in \Theta.$$

THEOREM 1. *The families $\{P_\theta^{(1)}\}$ and $\{P_\theta^{(2)}\}$ are equivalent in the theory of statistical inference iff there exist decision rules $\mathrm{III}^{(21)}$ and $\mathrm{III}^{(12)}$ such that*

(3) $$P_\theta^{(1)} = P_\theta^{(2)}\mathrm{III}^{(21)}, \quad P_\theta^{(2)} = P_\theta^{(1)}\mathrm{III}^{(12)}, \quad \forall\theta \in \Theta.$$

(1) The text following below combines two lectures of the author: "On basic concepts of mathematical statistics" and "On testing hypotheses".