has approximately the same power for testing $\xi > 0$ (and $\xi < 0$) as the corresponding Student t-test based on

(5)
$$t = \frac{(\bar{r} - \xi)\sqrt{n(n-1)}}{\sqrt{\sum_1^n (r_i - \bar{r})^2}}$$

for $n \leq 10$.

Using the notation of section 4 let

$$r_u = \frac{\sqrt{s}}{K_1}\left[K_1 y_u - \sum_1^n y_i + K_2 \sqrt{\frac{r}{s}}\, x\right], \quad (u = 1, \cdots, n),$$

where $\dfrac{K_1}{K_2} > 0$. Then from consideration of (4) with $C = 0$ it is seen that the r_u are independently distributed according to $N(\xi, \sigma^2)$, where ξ equals a positive constant times $(\nu - \mu)$. Following the derivations in section 4 with $C = 0$, it is seen that the test of $\xi > 0$ with this particular choice of the r_u is identical with the test of $\nu > \mu$ given in (B) of section 3. Similarly the test of $\xi < 0$ is identical with the test (B) of $\nu < \mu$. Thus the test (B) has approximately the same power for testing $\nu > \mu$ (and $\nu < \mu$) as the Student t-test based on the value of t given in (5) if $n \geq 10$. Replacing the r_u in (5) by their values in terms of $x, y_1, \cdots, y_n, n, r$, and s, it is found that (5) becomes

$$t = \frac{[x - \bar{y} - (\nu - \mu)]}{\sqrt{\sum_1^n (y_i - \bar{y})^2}} \cdot \sqrt{\frac{n-1}{s\left(\dfrac{1}{r} + \dfrac{1}{ns}\right)}}.$$

This proves that test (B) is approximately as powerful for testing $\nu > \mu$ and $\nu < \mu$ as the most powerful test based on the quantities $x, y_1, \cdots, y_n$ if $n \leq 10$. As test (A) is a particular case of test (B), these results also apply to test (A).

REFERENCES

[1] J. F. Daly, "On the use of the sample range in an a analogue of Student's t-test". *Annals of Math. Stat.*, Vol. 17 (1946), pp. 71-74.
[2] E. Lord, "The use of range in place of standard deviation in t-test," *Biometrika*, Vol. 34 (1947), pp. 41–67.

ON THE NORM OF A MATRIX

By Albert H. Bowker
University of North Carolina

In studying the convergence of iterative procedures in matrix computation and in setting limits of error after a finite number of steps, Hotelling [1] used the square root of the sum of squares of the elements of a matrix as its norm. A wide class of functions exists which may be employed as norms in matrix calculation and substituted directly in the expressions derived by Hotelling. The purpose of this note is to make a few general remarks about this class of functions and to propose a new norm which appears to have some value in computation.

A function $\phi(A)$ of the elements of a real matrix A may be termed a legitimate norm if it has the following four properties:

$$\phi(cA) = |c| \phi(A), \; c \text{ a scalar;} \tag{1}$$

$$\phi(A + B) \leqq \phi(A) + \phi(B), \text{ if } A + B \text{ is defined;} \tag{2}$$

$$\phi(AB) \leqq \phi(A)\phi(B), \text{ if } AB \text{ is defined;} \tag{3}$$

$$\phi(e_{ij}) = 1, \text{ where } e_{ij} \text{ is a fundamental unit matrix} \tag{4}$$

whose elements are all zero except the one in the ith row and jth column, whose value is unity. These four conditions are identical with the first four axioms of Rella [2], who has shown them to be independent. Properties (1), (2), and (3) are used directly in investigations of convergence and error, but the importance of property (4) is indicated by some of its immediate consequences. Clearly $e_i' A e_j = a_{ij}$, where e_i is a fundamental unit vector. From (3) and (4) it follows that $|a_{ij}| \leqq \phi(A)$ for all i and j and we have that

$$\max_{(ij)} |a_{ij}| \leqq \phi(A). \tag{5}$$

Thus $\phi(A)$ has the useful property that the norm of a matrix of errors exceeds or equals the maximum possible error. Since $\phi(A^m) \leqq \phi^m(A)$, it follows from (5) that the elements of A^m will tend to zero as m increases if $\phi(A) < 1$, a result which is useful in establishing convergence. Also $\phi(A) \geqq 0$.

One further consequence of (1) to (4) is of interest. Suppose A is a square matrix and let λ be any of its roots. Then there exists a non-null vector x such that $Ax = \lambda x$. Now $\phi(\lambda x) = \lambda\phi(x) \leqq \phi(A)\phi(x)$ and we have

$$\lambda \leqq \phi(A). \tag{6}$$

Thus, every legitimate norm is an upper bound to the characteristic roots.

Clearly many functions exist which satisfy (1) to (4). The norm used by Hotelling is $N(A) = \sqrt{\sum_{i,j} a_{ij}^2}$. A new norm which may have some value is obtained as follows:

$$R(A) = \max_{(i)} R_i(A) \tag{7}$$

where

$$R_i(A) = \sum_j |a_{ij}| .$$

Clearly $R(cA) = |c| R(A)$. To show that R satisfies (2), consider

$$R_i(A + B) = \sum_j |a_{ij} + b_{ij}| \leqq \sum_j |a_{ij}| + \sum_j |b_{ij}| \leqq R(A) + R(B).$$

Since the above inequality holds for all i,

$$R(A + B) \leqq R(A) + R(B).$$

Now $AB = || \sum_{\alpha} a_{i\alpha} b_{\alpha j} ||$

and

$$R_i(AB) = \sum_j | \sum_\alpha a_{i\alpha} b_{\alpha j} | \leqq \sum_j \sum_\alpha | a_{i\alpha} | \cdot | b_{\alpha j} |$$
$$\leqq \sum_\alpha | a_{i\alpha} | R_\alpha(B) \leqq R(B)R(A).$$

Hence $R(AB) \leqq R(A)R(B)$. Clearly $R(e_{ij}) = 1$. Similarly it may be shown that $C(A) = \max_{(j)} \sum_i | a_{ij} |$ also satisfies the conditions of a norm.

Since the convergence of an iterative procedure is often proved by the norm being less than one, since the norm appears in the upper bound for the error after a finite number of iterations, and since the norm of a matrix of errors is taken to indicate the magnitude of the errors, a reasonable method of choosing among several available legitimate norms is to select the smallest. It is natural to inquire whether an optimum norm in this sense exists; that is, is there a function $\phi^*(A)$ such that $\phi^*(A)$ possesses properties (1) through (4) and such that $\phi^*(A) \leqq \phi(A)$ for all other $\phi(A)$ satisfying these conditions. Assume such a $\phi^*(A)$ does exist. Clearly $\phi^*(A) = \phi^*(A')$, as, if either exceeded the other, the smaller could be taken as $\phi^*(A)$. Let Λ^2 be the largest root of AA'. Then by (6)

$$\Lambda^2 \leqq \phi^*(AA') \leqq \phi^{*2}(A) \text{ and } \Lambda \leqq \phi^*(A).$$

But Rella [2] has shown that Λ possesses (1) to (4). Thus

$$\phi^*(A) = \Lambda.$$

But, for a row vector, $C(A) \leqq \Lambda$. Consequently, no minimal norm exists. It is interesting to note that a worst norm does exist, namely $P(A) = \sum_{i,j} | a_{ij} |$. Since $A = \sum_{i,j} e_{ij} a_{ij}$, $\phi(A) \leqq P(A)$. Clearly $P(A)$ satisfies (1) to (4) and hence is the worst possible legitimate norm.

In practical computation, the choice so far is between $N(A)$ and $R(A)$ (or $C(A)$). No general inequalities exist and it would probably be advisable to compute both. $R(A)$ may be less than $N(A)$ and indicate convergence when $N(A)$ fails to do so. Often $R(A)$ may be computed visually and convergence proved without computing the sum of squares of the elements.

The functions $N(A)$ and $R(A)$ may also be useful in finding a simple first approximation to A^{-1}. A sufficient condition that Hotelling's iterative method for finding the inverse of a matrix A will converge is that the roots of $D = 1 - AC_0$ be less than one in absolute value where C_0 is a first approximation to A^{-1}. If the iterative procedure is to be carried out by a fully automatic computing machine such as the one described by Alt [3] it may be advisable to start with a rather poor first approximation which is easy to construct. If A has positive roots and if M is any upper bound to these roots and if C_0 is a matrix with diagonal elements equal to $1/M$ and zeros elsewhere, the iterative procedure will converge but the norm of D will not necessarily be less than one. From (6), any legitimate norm may be taken as M.

Finally, it is interesting to point out the relation of this note to some work on the problem of finding upper bounds to the roots. In fact, the inequalities $\lambda \leqq N(A)$ and $\lambda \leqq R(A)$, which are consequences of (6), are Theorem 2 of Farnell [4] and Theorem 3 of Barankin [5] respectively.

REFERENCES

[1] HAROLD HOTELLING, "Some new methods in matrix calculation," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 1–34.

[2] T. RELLA, "Über den absoluten Betrag von Matrizen," International Congress of Mathematicians at Oslo (1936).

[3] FRANZ L. ALT, "Multiplication of matrices," *Math. Tables and Aids to Compution*, Vol. 2 (1946), pp. 12–13.

[4] A. B. FARNELL, "Limits for the characteristic roots of a matrix," *Bull. Amer. Math. Soc.*, Vol. 50 (1944), pp. 789–794.

[5] EDWARD W. BARANKIN, "Bounds for the characteristic roots of a matrix," *Bull. Amer. Math. Soc.*, Vol. 51 (1945), pp. 767–770.

DEFINITION OF THE PROBABLE DEVIATION

BY M. FRÉCHET

Faculty of Science, University of Paris

The probable deviation has recently been defined by E. J. Gumbel [1], [2] as the smallest of the intervals corresponding to the probability $\frac{1}{2}$. It so happened that the author was led to an equivalent definition starting from a general idea which may be applied to absolutely general cases and which, for this reason, might be of interest.

In recent years, the author has been occupied with a study of random elements of any nature (curves, surfaces, functions, qualitative elements), a study whose future seems promising, [3]. I gave a definition of the mean of such an element expressed by an abstract integral which, however, is only defined if the random element is situated in a metric vectorial (Wiener-Banach) space.[1] But[2] a still more general definition is valid if the random element is placed in any metric space. It consists of taking, as mean position of the random element X, a fixed (non-statistical) element $b = \bar{X}$ such that the function of a which represents the mean $M(X, a)^2$ of the squared distance of X to the fixed element a, is minimum for $a = b$. (In the case where X and a are numbers, and where $M(X)^2$ is finite, we know that this minimum is reached and that there is one, and only one, determination b of a). This definition has the advantage of also defining the equiprobable position of X. This is a fixed element $c = \bar{\bar{X}}$ such that $M(X, a)$ is minimum for $c = a$. (If X and a are numbers, we know that this minimum is still reached, but may be so reached by several values of $\bar{\bar{X}}$).

Since reading Gumbel's paper, a still more general definition suggested itself.

[1] For the definition of metric vectorial spaces see [4].

[2] See Note 2, p. 503 of [4].