

# ON THE DISTRIBUTION OF A CONVEX EVEN FUNCTION OF SEVERAL INDEPENDENT ROUNDING-OFF ERRORS

H. P. MULHOLLAND

**1. Introduction.** In the course of their recent investigations into the effect of rounding-off errors in long computations, Goldstine and von Neumann<sup>1</sup> have considered the probability distribution of the bound of a matrix, or the length of a vector, whose elements are independent random variables. In the case where each element is the sum of one or more (elementary) rounding-off errors they presume that it has a cumulative distribution function (c.d.f.) of the type  $V_k(x/b)$ , where  $V_1(t)$  corresponds to the uniform (or "rectangular") distribution on the interval  $-1/2 < t < 1/2$ , and  $V_k(t)$  is its  $k$ -fold convolution, that is,

$$(1.1) \quad V_k(t) = V_{k-1}(t) * V_1(t) \quad (k = 2, 3, \dots),$$

while  $b$  is a constant (positive) scale factor. The parallel case in which each element has the normal c.d.f.  $\Phi(x/\sigma)$ , where

$$(1.2) \quad \Phi(t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^t e^{-u^2/2} du,$$

is much more manageable analytically: Goldstine and von Neumann dispose of this case first and then reduce the former case to it by means of a "fundamental lemma,"<sup>2</sup> in which they compare the distributions in the two cases of a function  $\phi(w)$  of a vector  $w = (w_1, w_2, \dots, w_M)$ . The main object of the present paper is to simplify and strengthen this lemma, so as to yield Theorem 1 below, which enables us somewhat to improve the numerical estimates they obtained.

We shall say that a function  $\phi(w)$  of a vector  $w$  is *convex* when for any  $w$  and  $w'$  we have  $2\phi(w+w') \leq \phi(2w) + \phi(2w')$ , and that it is *even* when  $\phi(-w) \equiv \phi(w)$ .

**THEOREM 1.** *Let the real vectors  $h = (h_1, h_2, \dots, h_M)$  and  $a = (a_1, a_2, \dots, a_M)$  and the real-valued function  $\phi(w)$  satisfy the following:*

(1.3) *the elements  $h_m$  are independent random variables such that*

---

Presented to the Society, September 7, 1951; received by the editors July 27, 1951.

<sup>1</sup> [3, Chapter VIII]. Numbers in brackets relate to the references listed at the end.

<sup>2</sup> [3, §8.3, pp. 192, 193].

$h_m$  has the c.d.f.  $V_{l_m}(x/b)$  [cf. (1.1)], and  $l_m \leq l$ , for  $m=1, 2, \dots, M$ ;

(1.4) the elements  $a_m$  are independent random variables such that  $a_m$  has the normal c.d.f.  $\Phi(x/\sigma)$ , for  $m=1, 2, \dots, M$ ;

(1.5)  $\phi(w)$  is a continuous convex even function of the vector  $w = (w_1, w_2, \dots, w_M)$ , with  $\phi(0) = 0$ .

Then

$$(1.6) \quad \text{Prob} [\phi(h) > t] \leq \text{Prob} [\phi(a) > t] \quad (t \geq 0),$$

provided that

$$(1.7) \quad \sigma \geq \sigma_l = \rho_l \cdot b(l/12)^{1/2},$$

where  $\rho_l$  is a constant depending upon  $l$  for which we have<sup>3</sup>

$$(1.8) \quad \rho_1 = (6/\pi)^{1/2} < 1.382, \quad 1 < \rho_l \leq \rho_2 < 1.0707 \quad (l \geq 2).$$

For comparison we note that in their lemma Goldstine and von Neumann required  $\phi(w)$  to be positively homogeneous and bounded by 1,<sup>4</sup> as well as continuous and convex, and that they established (for any  $t > 0$ ) the inequality<sup>5</sup>

$$\begin{aligned} \text{Prob} [\phi(h) > t + [M + (M/l)]^{1/2}b] \\ \leq 2e^{3M/(4l)}(\log_2 2l) \text{Prob} [\phi(a) > t], \end{aligned}$$

which is clearly weaker than (1.6), under the proviso  $\sigma \geq (l+1)^{1/2}b/2$ , which is more restrictive than (1.7). The improvements in their application of the lemma that result when we replace it by Theorem 1 will be summarized in §§6 and 7 below.

Theorem 1 is not in general "best possible." However, it will appear at the end of §4 that in the special case where  $l_m = l > 2$  ( $m=1, 2, \dots, M$ ), and<sup>6</sup>

$$(1.9) \quad |w| \rightarrow 0 \quad \text{as} \quad \phi(w) \rightarrow 0,$$

we can only have (1.6) holding for small values of  $t$  if  $\sigma \geq b(l/12)^{1/2}$ , that is, if  $a_m$  has as great a variance as  $h_m$ ; in view of (1.8) this means that there is only a margin for possible improvement of about 7% when  $l > 2$ .

<sup>3</sup> Actually,  $\rho_l < \rho_2$  ( $l > 2$ ), as was stated in the abstract of this paper, but the proof of this refinement will be merely indicated (in §5).

<sup>4</sup> That is,  $\phi(sw) = |s| \cdot \phi(w)$  for every real  $s$ , and  $\phi(w) \leq 1$  when  $|w| \leq 1$ ; see [3, (8.15), p. 192].

<sup>5</sup> [3, (8.16), p. 193]; in re-writing this we have put  $\kappa\beta^{-s}(l+1)^{1/2}/2 = t$  and  $\beta^{-s} = b$ .

<sup>6</sup> Condition (1.9) is satisfied if  $\phi(w)$  is either the length  $|w|$  of the vector  $w$  or the bound of the matrix with elements  $w_1, w_2, \dots, w_M$  ( $M = n^2$ ); cf. [4, Theorem 289, p. 209].

In order to prove Theorem 1 we observe that (1.6) asserts that the distribution of  $\phi(h)$  is "more peaked about 0" than that of  $\phi(a)$ , in a sense introduced by Birnbaum [1], and we generalize to functions  $\phi(w)$  satisfying (1.5) a comparison theorem established by him for  $\phi(w) \equiv |w_1 + w_2 + \cdots + w_M|$ . This leads us to Theorem 2 in §2, which is of interest in itself since it deals with distributions of a fairly general type and may thus be of service also in other connexions.

**2. Definitions and outline of the argument.** We shall say that a probability distribution with c.d.f.  $F(x)$  is *more peaked about 0* than one with c.d.f.<sup>7</sup>  $G(x)$  when

$$(2.1) \quad F(x) - F(-x) \geq G(x) - G(-x) \quad (x \geq 0).$$

This implies that when the densities at  $x = 0$  exist we have

$$(2.2) \quad F'(0) \geq G'(0).$$

When the distributions are either (i) symmetrical about 0, or (ii) confined to  $(0, \infty)$ , we may replace (2.1) by

$$(2.3) \quad F(x) \geq G(x) \quad (x \geq 0);$$

in these cases we shall say that  $F(x)$  *majorizes*<sup>8</sup>  $G(x)$ . We note that the distributions with c.d.f.'s of type  $V_k(x/b)$  and  $\Phi(x/\sigma)$ , concerned in the hypotheses of Theorem 1, fall under case (i), while the distributions of  $\phi(h)$  and  $\phi(a)$  in its conclusion fall under case (ii).

**THEOREM 2.** *Let  $u_1, u_2, \cdots, u_M$  be independent random variables with unimodal absolutely continuous distributions symmetrical about 0, and similarly for  $v_1, v_2, \cdots, v_M$ . Then a necessary and sufficient condition for the distribution of  $\phi(u)$  to be more peaked about 0 than that of  $\phi(v)$ , that is,*

$$(2.4) \quad \text{Prob} [\phi(u) \leq t] \geq \text{Prob} [\phi(v) \leq t] \quad (t \geq 0),$$

*for every function  $\phi(w)$  satisfying (1.5), is that the distribution of  $u_m$  should be more peaked about 0 than that of  $v_m$  ( $m = 1, 2, \cdots, M$ ).*

The sufficiency part of this theorem is covered by Birnbaum's work<sup>9</sup> when  $\phi(w) \equiv |w_1 + w_2 + \cdots + w_M|$ ; we shall refer to this case as Birnbaum's theorem. Theorem 2 will be proved in §4 below.

<sup>7</sup> Or, alternatively, when  $F(x)$  and  $G(x)$  are absolutely continuous, that the density  $F'(x)$  is more peaked about 0 than  $G'(x)$ .

<sup>8</sup> For a discrete analogue see [4, p. 45].

<sup>9</sup> [1, Lemma, p. 77, and Theorem 1, p. 79].

In virtue of Theorem 2 we can establish Theorem 1 by proving that, subject to a suitable choice of  $\sigma$ , the distribution with c.d.f.  $V_{l_m}(x/b)$  is more peaked about 0 than that with c.d.f.  $\Phi(x/\sigma)$ , or in other words that  $V_{l_m}(x/b)$  majorizes  $\Phi(x/\sigma)$ . Now  $l_m \leq l$ , and hence  $V_{l_m}(x/b)$  majorizes  $V_l(x/b)$ , by a limiting case of Birnbaum's theorem. In fact, when  $0 < \epsilon < b$ , and  $k = l - l_m \geq 1$ , it is easy to see that  $V_k(x/\epsilon)$  majorizes  $V_k(x/b)$ . Therefore, by Birnbaum's theorem,  $V_{l_m}(x/b) * V_k(x/\epsilon)$  majorizes  $V_{l_m}(x/b) * V_k(x/b) = V_{l_m+k}(x/b) = V_l(x/b)$  [cf. (1.1)]. On letting  $\epsilon \rightarrow 0$  we see that since  $V_{l_m}(x/b)$  is continuous it must majorize  $V_l(x/b)$ ,<sup>10</sup> as required.

It remains only to show that we can choose  $\sigma$  so as to make  $V_l(x/b)$  majorize  $\Phi(x/\sigma)$ , that is,

$$(2.5) \quad V_l(x/b) \geq \Phi(x/\sigma) \quad (x \geq 0),$$

since then we shall have  $V_{l_m}(x/b)$  majorizing  $\Phi(x/\sigma)$  a fortiori when  $l_m \leq l$ . Both sides of (2.5) increase strictly and continuously, from  $1/2$  at  $x=0$ , up to 1; hence this inequality is equivalent to

$$(2.6) \quad \sigma \geq \sigma_l = b \sup_{1/2 < y < 1} V_l^{-1}(y)/\Phi^{-1}(y).$$

Now  $V_l^{-1}(y)/\Phi^{-1}(y)$  is clearly continuous for  $1/2 < y < 1$ , and it tends to 0 as  $y \rightarrow 1-0$  and to  $\Phi'(0)/V_l'(0)$  as  $y \rightarrow 1/2$ ; thus the supremum in (2.6) is always finite and<sup>11</sup>

$$(2.7) \quad \sigma_l \geq \frac{b\Phi'(0)}{V_l'(0)}.$$

When  $l=1$ , we have, for  $1 > y > 1/2$  and so for  $t = \Phi^{-1}(y) > 0$ ,

$$\frac{V_1^{-1}(y)}{\Phi^{-1}(y)} = \frac{y - (1/2)}{\Phi^{-1}(y)} = \frac{\Phi(t) - \Phi(0)}{t} \leq \Phi'(0),$$

since  $\Phi(t)$  is concave for  $t \geq 0$  [cf. (1.2)]. Thus we have equality in (2.7) when  $l=1$ ; hence  $\sigma_1 = b/(2\pi)^{1/2}$ . Since  $V_1(x/b)$  majorizes  $\Phi(x/\sigma_1)$  and  $V_l(x/b)$  is the  $l$ -fold convolution of  $V_1(x/b)$ , Birnbaum's theorem shows that  $V_l(x/b)$  majorizes the  $l$ -fold convolution of the normal c.d.f.  $\Phi(x/\sigma_1)$ , which is  $\Phi[x/(\sigma_1 l^{1/2})]$ ; thus  $\sigma_l \leq \sigma_1 l^{1/2} = b \cdot [l/(2\pi)]^{1/2}$ . Noting that the variance of  $V_l(x/b)$  is  $lb^2/12$  we introduce the ratio

<sup>10</sup> More generally, the convolution of two symmetrical unimodal absolutely continuous distributions is less peaked about 0 than either of them. For a discrete analogue of this see [4, Theorem 46, p. 49].

<sup>11</sup> We could also have obtained (2.7) as a special case of (2.2).

$$(2.8) \quad \rho_l = \sigma_l / (lb^2/12)^{1/2} = (\sigma_l/b)(12/l)^{1/2}.$$

We have thus obtained

$$(2.9) \quad \rho_l \leq \rho_1 = (6/\pi)^{1/2} < 1.382 \quad (l \geq 1).$$

This crude upper bound for  $\rho_l$  ( $l > 1$ ) will be improved in §5 by means of a similar argument based on evaluations of  $\rho_2$  and  $\rho_3$ ; we shall then have completed (1.8) in Theorem 1.

**3. Lemmas on cross-sections of convex regions.** We shall say that a function of  $t$  is *symmetrically decreasing* when it is an even function of  $t$  and decreases as  $|t|$  increases; we note that a unimodal probability distribution symmetrical about 0 has a symmetrically decreasing density.

**LEMMA 1.** *Let  $C$  be a bounded convex region symmetrical about the origin in  $(x_1, x_2, \dots, x_M)$ -space and  $A(t)$  be the hyperarea  $[(M-1)$ -dimensional content] of its cross-section by the hyperplane  $x_M = t$ . Then  $A(t)$  is symmetrically decreasing.*

By a theorem due to H. Brunn (cf. [2, pp. 124-129])  $[A(t)]^{1/(M-1)}$  is a concave function of  $t$  for all values of  $t$  that give a nonempty section. Since  $C$  is symmetrical,  $A(t)$  must be even, and it follows that  $A(t)$  is symmetrically decreasing for all values of  $t$ . (We take  $A(t) = 0$  when  $t$  gives an empty section.)

**LEMMA 2.** *Let  $x_1, x_2, \dots, x_{M-1}$  be independent random variables each of which has a unimodal absolutely continuous distribution symmetrical about 0. Let  $C$  be any convex region symmetrical about the origin in  $(x_1, x_2, \dots, x_M)$ -space and let*

$$(3.1) \quad s(t) = \text{Prob} [x \in C \mid x_M = t].$$

*Then  $s(t)$  is symmetrically decreasing.*

We write  $p_m(x_m) = -f_m(x_m)$  for the probability density of  $x_m$ ; then  $p_m(x_m)$  is symmetrically decreasing and clearly tends to 0 as  $|x_m| \rightarrow \infty$ . Hence

$$p_m(x_m) = \int_{|x_m|}^{\infty} df_m(u_m) \quad (m = 1, 2, \dots, M-1),$$

where  $f_m(u_m)$  is increasing for  $0 \leq u_m < \infty$ . Let  $\chi(x_1, x_2, \dots, x_M)$  denote the characteristic function of the region  $C$ . Then<sup>12</sup>

<sup>12</sup> The inversion of the order of integration is covered by Fubini's Theorem; cf. [5, p. 77].

$$\begin{aligned}
s(t) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \chi(x_1, \cdots, x_{M-1}, t) p_1(x_1) \\
&\quad \cdots p_{M-1}(x_{M-1}) dx_1 \cdots dx_{M-1} \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \chi(x_1, \cdots, x_{M-1}, t) dx_1 \cdots dx_{M-1} \int_{|x_1|}^{\infty} df_1(u_1) \\
&\quad \cdots \int_{|x_{M-1}|}^{\infty} df_{M-1}(u_{M-1}) \\
&= \int_0^{\infty} \cdots \int_0^{\infty} df_1(u_1) \cdots df_{M-1}(u_{M-1}) \int_{-u_1}^{u_1} dx_1 \\
&\quad \cdots \int_{-u_{M-1}}^{u_{M-1}} \chi(x_1, \cdots, x_{M-1}, t) dx_{M-1}.
\end{aligned}$$

But when  $|t| \leq T$ , the inner  $(M-1)$ -fold integral here is the cross-sectional hyperarea by the hyperplane  $x_M = t$  of the region  $CU_T$ , where  $U_T$  denotes the rectangular region  $|x_1| \leq u_1, \cdots, |x_{M-1}| \leq u_{M-1}, |x_M| \leq T$ . Now  $CU_T$  is clearly convex, symmetrical about the origin, and *bounded*. Hence, by Lemma 1, the inner integral, and therefore also  $s(t)$ , is symmetrically decreasing for  $|t| \leq T$ . Moreover, since  $T$  is arbitrary, this restriction upon  $t$  may be dropped.

4. **Generalization of Birnbaum's theorem.** We start with a preliminary theorem.

**THEOREM 3.** *Let  $u_1, u_2, \cdots, u_M$  and  $v_1, v_2, \cdots, v_M$  be as in Theorem 2. Then a necessary and sufficient condition for*

$$(4.1) \quad \text{Prob} [u \in C] \geq \text{Prob} [v \in C]$$

*to hold for every convex region  $C$  symmetrical about the origin is that the density  $p_m(t)$  of  $u_m$  should be more peaked about 0 than the density  $q_m(t)$  of  $v_m$ , that is,*

$$(4.2) \quad P_m(t) = \int_{-\infty}^t p_m(r) dr \geq \int_{-\infty}^t q_m(r) dr = Q_m(t) \quad (t \geq 0)$$

for  $m = 1, 2, \cdots, M$ .

In proving sufficiency we need only deal with the case when  $q_m$  is distinct from  $p_m$  for only one value of  $m$ , since the proof can then be completed by induction. Suppose, for definiteness, that  $p_m(t) \equiv q_m(t)$  ( $m = 1, 2, \cdots, M-1$ ), while  $P_M(t) \geq Q_M(t)$  for  $t \geq 0$ . We have, using (3.1) above and integrating by parts,

$$\begin{aligned}
\text{Prob } [u \in C] &= \int_{-\infty}^{\infty} \text{Prob } [u \in C \mid u_M = t] p_M(t) dt \\
&= \int_{-\infty}^{\infty} s(t) p_M(t) dt = 2 \int_0^{\infty} p_M(t) s(t) dt \\
&= s(0) + 2 \int_0^{\infty} [P_M(t) - 1] \cdot d[-s(t)] \\
&\geq s(0) + 2 \int_0^{\infty} [Q_M(t) - 1] \cdot d[-s(t)] = \text{Prob } [v \in C].
\end{aligned}$$

We have used here the evenness of  $p_m(t)$  and  $q_m(t)$  and the symmetrically decreasing character of  $s(t)$ , the latter resulting from Lemma 2 and implying that  $-s(t)$  is increasing for  $t \geq 0$ ; we have used also the hypothesis (4.2) for  $m = M$ .

For the necessity we have merely to take for  $C$  the region where the  $m$ th coordinate does not exceed  $t$  in absolute value, and we get  $P_m(t) - P_m(-t) \geq Q_m(t) - Q_m(-t)$  for  $t \geq 0$ , that is, the distribution of  $u_m$  is more peaked about 0 than that of  $v_m$ , as required.

PROOF OF THEOREM 2. We note first that the conditions (1.5) upon  $\phi$  imply

(4.3) for every  $t$  the region  $C_t$  where  $\phi(w) \leq t$  is convex and symmetrical about the origin;

(4.4)  $C_t$  is empty if and only if  $t < 0$ .

In fact, the convexity of  $C_t$  clearly results (cf. [4, Theorem 98, p. 80]) from the convexity and continuity of  $\phi(w)$ , and the symmetry of  $C_t$  from the evenness of  $\phi(w)$ ; further, (4.3) implies that  $C_t$  either contains the origin or is empty, and since  $\phi(0, 0, \dots, 0) = 0$ , the origin is in  $C_t$  if and only if  $t \geq 0$ , whence (4.4), which shows, in particular, that  $\phi(w)$  is nowhere negative.

Now suppose, firstly, that (4.2) holds. Then, by (4.3) and (4.4), we may take any  $t \geq 0$  and apply Theorem 3 to the region  $C_t$ : we get  $\text{Prob } [u \in C_t] \geq \text{Prob } [v \in C_t]$ . By the definition of  $C_t$  this is equivalent to the inequality in (2.4). Secondly, suppose that (2.4) holds: then, taking  $\phi(w) \equiv |w_m|$  we have the required relation (4.2) between the distributions of  $u_m$  and  $v_m$ .

REMARK. For a fixed  $\phi$  satisfying the additional condition (1.9) the relation (2.4) implies the following relation between densities at the origin:

$$(4.5) \quad p_1(0)p_2(0) \cdots p_M(0) \geq q_1(0)q_2(0) \cdots q_M(0);$$

and, in particular, in the "case of equal components" when  $p_m(x)$

$\equiv p(x)$ ,  $q_m(x) \equiv q(x)$  ( $m = 1, 2, \dots, M$ ), we get

$$(4.6) \quad p(0) \geq q(0),$$

which may be compared with (2.2). To deduce (4.5) we have only to note that, in virtue of (1.5), the region  $\phi(w) \leq t$  must contain the origin as an interior point when  $t > 0$  and, by (1.9), can have an arbitrarily small diameter by choice of  $t$ . In the special case of Theorem 1 where  $l_m = l$  ( $m = 1, 2, \dots, M$ ) and  $\phi$  satisfies (1.9) we see from (4.6) that the inequality (1.6) implies  $V'_i(0)/b \geq \Phi'(0)/\sigma$ , and so

$$(4.7) \quad \sigma/b \geq \Phi'(0)/V'_i(0) = (2\pi)^{-1/2}/V'_i(0).$$

It can be shown<sup>13</sup> that the right-hand side does not exceed  $(l/12)^{1/2}$ , except when  $l = 2$ , so that for  $l \neq 2$  we must have  $\sigma/b \geq (l/12)^{1/2}$  if (1.6) is to hold in the special case considered.

**5. Normal distributions less peaked about 0 than the  $l$ -fold rectangular distribution.** In virtue of (2.5), (2.6), and (2.8) the normal distributions we require are those with variances  $\sigma^2$  not less than  $\sigma_i^2 = \rho_i^2 b^2 l/12$ . The following lemma gives some numerical information about  $\rho_l$ , especially when  $l$  is small, and it covers (1.8) in Theorem 1. When  $l$  is large, the inequality  $2l(\rho_l^2 - 1) \leq (\ln l) + 0.4212 - (2l)^{-1}$  gives better results, but its proof is beyond the scope of the present paper.

**LEMMA 3.** *For the ratio  $\rho_l$  defined in (2.8) above we have*

$$(5.1) \quad \rho_l > 1 \quad (l = 1, 2, \dots);$$

$$(5.2) \quad \rho_1 = (6/\pi)^{1/2} < 1.382, \quad 1.0703 < \rho_2 < 1.0707, \\ \rho_3 = [32/(9\pi)]^{1/2} < 1.0639;$$

$$(5.3) \quad (3m + 2k)\rho_{3m+2k}^2 \leq 3m\rho_3^2 + 2k\rho_2^2 \quad (m, k = 0, 1, \dots);$$

$$(5.4) \quad \rho_l \leq \rho_2 < 1.0707 \quad (l \geq 2);$$

$$(5.5) \quad \frac{1}{\rho_l} \leq \left(\frac{l}{12}\right)^{1/2} \frac{V'_i(0)}{\Phi'(0)} \\ = \left(\frac{\pi l}{6}\right)^{1/2} \frac{1}{(l-1)!} \sum_{j=0}^{[l/2]} (-1)^j \binom{l}{j} \left(\frac{l}{2} - j\right)^{l-1},$$

where  $[l/2]$  denotes the greatest integer contained in  $l/2$ .

<sup>13</sup> The proof can be effected easily when  $l$  is odd by using the characteristic functions (for the relations between which cf. [6, p. 305, Ex. 6]), and with somewhat more trouble when  $l$  is even but greater than 2.



We note that for a c.d.f.  $F(x)$  the variance, if finite, equals

$$\int_0^\infty x^2 d[F(x) - F(-x)] = \int_0^\infty [1 - F(x) + F(-x)] 2x \cdot dx.$$

Now, by (2.5) and (2.6),  $V_i(x/b) \geq \Phi(x/\sigma_i)$  ( $x \geq 0$ ); thus

$$\begin{aligned} \frac{lb^2}{12} &= \int_0^{lb/2} [2 - 2V_i(x/b)] 2x \cdot dx \\ &\leq \int_0^{lb/2} [2 - 2\Phi(x/\sigma_i)] 2x \cdot dx < \sigma_i^2, \end{aligned}$$

and hence, by (2.8),  $\rho_i > 1$ , as required for (5.1).

The first part of (5.2) was proved in §2 [cf. (2.9)]. Next, from (2.6) and (2.8) we get

$$\begin{aligned} (5.6) \quad \left(\frac{l}{12}\right)^{1/2} \rho_i &= \sup_{1/2 < y < 1} \frac{V_i^{-1}(y)}{\Phi^{-1}(y)} = \sup_{z > 0} \frac{V_i^{-1}[\Phi(z)]}{z} \\ &= \sup_{t > 0} \frac{t}{\Phi^{-1}[V_i(t)]}. \end{aligned}$$

We note the known formula (cf. [6, pp. 277, 278])

$$(5.7) \quad V_i(x) = \frac{1}{l!} \sum_{j=0}^{[(l/2)+x]} (-1)^j \binom{l}{j} \left(\frac{l}{2} + x - j\right)^l.$$

When  $l=2$ , we have  $V_2(x) = 1 - (x-1)^2/2$  for  $0 \leq x \leq 1$  and so  $V_2^{-1}(y) = 1 - (2-2y)^{1/2}$  for  $1/2 \leq y \leq 1$ . Thus, from (5.6),

$$(5.8) \quad \rho_2/6^{1/2} = \sup \{1 - [2 - 2\Phi(z)]\}/z \quad (z > 0).$$

Substitution here of 0.919 for  $z$  yields  $\rho_2 > 1.0703$ . Again, choosing  $\sigma/b$  so as to make  $V_2(x/b) - \Phi(x/\sigma)$  stationary at  $x=0.919$ , we can verify by elementary calculations that the only minimum of this difference (for  $x \geq 0$ ) occurs at  $x=0.919$  and is positive, so that  $V_2(x/b)$  majorizes  $\Phi(x/\sigma)$ ; we find  $(\sigma/b)(12/2)^{1/2} < 1.0707$ , and the second part of (5.2) follows.

For the third part of (5.2) we have to show that  $V_3(y) - \Phi[y(9\pi/8)^{1/2}]$  is non-negative for  $y \geq 0$ . Since it vanishes for  $y=0$  and  $y \rightarrow \infty$ , we need only establish that its derivative

$$\delta(y) = V_3'(y) - (9\pi/8)^{1/2} \Phi'[y(9\pi/8)^{1/2}] = V_3'(y) - (3/4)e^{-9\pi y^2/16}$$

changes sign only once, from positive to negative, as  $y$  increases from

0 to  $\infty$ . But this may be verified by elementary calculations<sup>14</sup> from the formulae [cf. (5.7)]

$$(4/3)\delta(y) = \begin{cases} 1 - (4y^2/3) - e^{-9\pi y^2/16} & (0 \leq y \leq 1/2), \\ (3 - 2y)^2/6 - e^{-9\pi y^2/16} & (1/2 \leq y \leq 3/2), \\ -e^{-9\pi y^2/16} & (3/2 \leq y). \end{cases}$$

To get (5.3) we note that, by Birnbaum's theorem (cf. §2), if  $V_l(x/b)$  majorizes  $\Phi(x/\sigma)$  and  $V_m(x/b)$  majorizes  $\Phi(x/\sigma')$ , then  $V_{l+m}(x/b) = V_l(x/b) * V_m(x/b)$  majorizes  $\Phi(x/\sigma) * \Phi(x/\sigma') = \Phi(x/\sigma'')$ , where  $\sigma''^2 = \sigma^2 + \sigma'^2$ . It follows that  $\sigma_{l+m}^2 \leq \sigma_l^2 + \sigma_m^2$ , that is,

$$(5.9) \quad (l + m)\rho_{l+m}^2 \leq l\rho_l^2 + m\rho_m^2.$$

Repeated application of this yields  $\rho_{km} \leq \min(\rho_k, \rho_m)$  for  $k, m = 1, 2, \dots$ . In particular, these inequalities give

$$(3k + 2m)\rho_{3k+2m}^2 \leq 3k\rho_{3k}^2 + 2m\rho_{2m}^2 \leq 3k\rho_3^2 + 2m\rho_2^2,$$

which is (5.3). Since, by (5.2),  $\rho_3 < \rho_2$ , this gives  $\rho_l \leq \rho_2$  for  $l = 3k + 2m \geq 2$ ; indeed  $\rho_l < \rho_2$  for  $l > 2$  except<sup>15</sup> possibly for  $l = 4$ .

Lastly, the inequality in (5.5) is equivalent to (2.7) above, and the equality results from (5.7). We remark that, by (5.2), we have equality throughout (5.5) when  $l = 1$  and  $l = 3$ , but<sup>16</sup> not when  $l = 2$ .

**6. Applications of Theorem 1.** We consider two cases: (a) where  $M = n^2$  and  $\phi(w)$  is the bound of a matrix of order  $n$  with elements  $w_1, w_2, \dots, w_n$  (in some order), (b) where  $M = n$  and  $\phi(w)$  is the length  $|w|$  of the vector with elements  $w_1, w_2, \dots, w_n$ . In both cases  $\phi(w)$  clearly fulfils the conditions (1.5) of Theorem 1. The following theorem is obtained when we bring in also the approximations used by Goldstine and von Neumann for the case of normally distributed elements.

**THEOREM 4.** (a) Let  $H = (h_{ij})$  be a matrix whose elements are independent random variables having the c.d.f.'s  $V_{l_{ij}}(h_{ij}/\beta^{-s})$  where  $l_{ij} \leq l$ . Then for the bound  $|H|$  of  $H$  we have

$$(6.1) \quad \text{Prob} [|H| > \rho_l \beta^{-s} (\ln r/6)^{1/2}] \leq (2re^{1-r})^n (4r - 4)^{-1} (r\pi n)^{-1/2}.$$

<sup>14</sup> Note that, for  $1/2 \leq y \leq 3/2$ ,  $\delta(y)$  has the same sign as  $[(3 - 2y)/6]^{1/2} - \exp(-9\pi y^2/32)$ , which is easily shown to be decreasing and thus can change sign only once.

<sup>15</sup> It can be verified that  $\rho_4 \leq 3/8^{1/2} < 1.07 < \rho_2$  by arguments like those used in evaluating  $\rho_3$ . Hence  $\rho_l < \rho_2$  ( $l > 2$ ).

<sup>16</sup> It can be verified that  $l = 4$  is also a case of inequality.

(b) Let  $x = (x_1, x_2, \dots, x_n)$  be a vector whose elements  $x_i$  are independent random variables with c.d.f.'s  $V_{m_i}(x_i/\beta^{-s})$ , where  $m_i \leq l$ . Then

$$(6.2) \quad \text{Prob} [ |x| > \rho_l \beta^{-s} (l n r / 12)^{1/2} ] \leq (r e^{1-r})^{n/2} (r - 1)^{-1} (\pi n)^{-1/2}.$$

In both (6.1) and (6.2)  $\rho_l$  is a constant depending on  $l$  and satisfying (1.8) above.

(a) From (1.6) of Theorem 1 we get, with  $b = \beta^{-s}$  and  $t = \sigma_l (2rn)^{1/2}$ ,

$$\text{Prob} [ |H| > \sigma_l (2rn)^{1/2} ] \leq \text{Prob} [ |A| > \sigma_l (2rn)^{1/2} ],$$

where  $\sigma_l = \rho_l \beta^{-s} (l/12)^{1/2}$  and  $A$  is a matrix  $(a_{ij})$  whose elements are independent random variables each with c.d.f.  $\Phi(x/\sigma_l)$ . But<sup>17</sup>

$$\text{Prob} [ |A| > \sigma_l (2rn)^{1/2} ] \leq (2r e^{1-r})^n (4r - 4)^{-1} (r\pi n)^{-1/2}.$$

Combining these inequalities and inserting the value of  $\sigma_l$  we have (6.1).

(b) From (1.6) of Theorem 1 we get, with  $b = \beta^{-s}$ ,  $t = \sigma_l (rn)^{1/2}$ , and  $x$  for  $h$ ,

$$\text{Prob} [ |x| > \sigma_l (rn)^{1/2} ] \leq \text{Prob} [ |a| > \sigma_l (rn)^{1/2} ],$$

where  $\sigma_l$  is as in (a). But (cf. [3, (8.13.a), p. 192])

$$\text{Prob} [ |a| > \sigma_l (rn)^{1/2} ] \leq (r e^{1-r})^{n/2} (r - 1)^{-1} (\pi n)^{-1/2};$$

combining these inequalities and inserting the value of  $\sigma_l$  we have (6.2).

**COROLLARY.** *With the notation and hypotheses of the theorem and also  $l = un$ ,  $u \geq 1$ ,  $n \geq 2$ , we have*

$$(6.3) \quad \text{Prob} [ |H| > 0.841 u^{1/2} n \beta^{-s} ] < 0.027 \times 2^{-n} n^{-1/2} < 0.02 \times 2^{-n},$$

$$(6.4) \quad \text{Prob} [ |x| > 0.595 u^{1/2} n \beta^{-s} ] < 0.21 \times 2^{-n} n^{-1/2} < 0.15 \times 2^{-n}.$$

The inequalities (6.3) and (6.4) follow from (6.1) and (6.2), respectively, when we put<sup>18</sup>  $r = 3.7$  and observe that, by (1.8),  $\rho_l \leq 1.0707$  for  $l \geq 2$ . For comparison we note the corresponding inequalities given by Goldstine and von Neumann:<sup>19</sup> for  $4 \geq u \geq 1$ ,  $n \geq 10$ ,

$$(6.5) \quad \text{Prob} [ |H| > (u + 0.1)^{1/2} (1.53 + u^{-1/2}) n \beta^{-s} ] < 0.07 \times 2^{-n},$$

$$(6.6) \quad \text{Prob} [ |x| > (u + 0.1)^{1/2} (1.34 n \beta^{-s}) ] < 0.03 \times 2^{-n/2}.$$

<sup>17</sup> [3, (8.3), p. 190].

<sup>18</sup> Cf. [3, (8.9), p. 191, and (8.13.b), p. 192].

<sup>19</sup> [3, (8.32), (c) and (d), p. 198].

It is easily verified that, for  $n \geq 10$ , (6.3) is stronger than (6.5), and (6.4) than (6.6).

7. **Numerical results connected with the inverting of a matrix.** We proceed to summarize numerical improvements obtainable in the results of Goldstine and von Neumann's Chapter IX [3, pp. 199–202] when (6.3) and (6.4) above are used instead of their (8.32) (see (6.5) and (6.6) above). The major change is a reduction in the probabilities of the neglected cases from at worst about 0.1% to at worst about 0.01%; also, in the remaining cases the estimates of error are reduced to at most 7/16 of their former values. In fact, subject to the standing assumption that  $n \geq 10$ , we have from (6.3) and (6.4), with at least 99.99% probability in each case,

$$(7.1) \quad |H| \leq 0.841u^{1/2}n\beta^{-s}, \quad |x| \leq 0.595u^{1/2}n\beta^{-s}.$$

Substitution of  $u = 1, 2$ , and 4 in the first of these, and of  $u = 4$  in the second, yields coefficients of  $n\beta^{-s}$  not exceeding 0.841, 1.19, 1.682, and 1.19, respectively. These correspond to the coefficients 2.66, 3.25, 4.12, and 2.72 in [3, (9.13), (9.1), (9.2) or (9.9), and (9.3)], respectively. It may be verified that the greatest ratio of corresponding coefficients is  $1.19/2.72 = 7/16$ . The coefficients in [3, (9.4), (9.12), second parts of (9.15) and (9.15<sub>I</sub>), (9.16), (9.16<sub>I</sub>), (9.17), (9.17<sub>I</sub>), and (9.19)] can therefore all be reduced to 7/16 of their values, while those in [3, (9.18) and (9.18')] can be multiplied by  $(16/7)^{1/2}$  and those in [3, (9.18''), (9.19.a), (9.19.b), and (9.19.c)] by  $(16/7)^{1/3}$ . In particular, the coefficients 114 and 293 in the main error estimates, [3, (9.16) and (9.16<sub>I</sub>)], can be replaced by 50 and 129, respectively.

#### REFERENCES

1. Z. W. Birnbaum, *On random variables with comparable peakedness*, Ann. Math. Statist. vol. 19 (1948) pp. 76–81.
2. T. Bonnesen, *Les problèmes des isopérimètres et des isépiphanes*, Paris, 1929.
3. H. H. Goldstine and J. von Neumann, *Numerical inverting of matrices of high order*. II, Proceedings of the American Mathematical Society vol. 2 (1951) pp. 188–202.
4. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge, 1934.
5. S. Saks, *Theory of the integral*, Warsaw, 1937.
6. J. V. Uspensky, *Introduction to mathematical probability*, New York, 1937.

UNIVERSITY OF BIRMINGHAM