THE ASYMPTOTICAL DISTRIBUTION OF RANGE IN SAMPLES FROM A NORMAL POPULATION

By G. ELFVING, *Helsingfors*

1. *Introductory*. Consider a sample of n observations, taken from an infinite normal population with the mean 0 and the standard deviation 1. Let $\mathbf{a}$ be the smallest and $\mathbf{b}$ the greatest of the observed values. Then $\mathbf{w} = \mathbf{b} - \mathbf{a}$ is the *range* of the sample.

For certain statistical purposes knowledge of the sampling distribution of range is needed. The distribution function, however, involves a rather complicated integral, whose exact calculation is, for $n > 2$, impossible. Tippett (1925), E. S. Pearson (1926, 1932) and McKay & Pearson (1933) have studied and calculated the mean, the standard deviation and the Pearson constants β_1, β_2 of the range. Fitting appropriate Pearson curves to the distribution by means of these parameters, Pearson (1932) has computed approximate percentage points for it. Later on, Hartley (1942) and Hartley & Pearson (1942) have, by numerical integration, tabulated the distribution function for $n = 2, \ldots, 20$.

As pointed out by Pearson, the distribution of range is very sensitive to departures from normality in the tails of the parental distribution. The effect of such departures becoming more perceptible for increasing n, the practical importance of the range distribution is, perhaps, small for large samples. Nevertheless, it seems to be at least of theoretical interest to investigate the *asymptotical* distribution of range for $n \to \infty$. This is the purpose of the present paper.* The results are summarized in a theorem at the end of the inquiry.

2. *The exact distribution. Transformations*. The joint-frequency function of the extremes $\mathbf{a}$, $\mathbf{b}$ reads, as well known,

$$f_{\mathbf{ab}}(a,b) = n(n-1)\,\phi(a)\,\phi(b)\,[\Phi(b)-\Phi(a)]^{n-2}\,\dagger \tag{2·1}$$

(cf. e.g. Cramér, 1945, p. 370). Let $\mathbf{u} = \frac{1}{2}(\mathbf{a}+\mathbf{b})$ denote the arithmetical mean of the extreme values of the sample. Making in (2·1) the transformation $\mathbf{a} = \mathbf{u} - \frac{1}{2}\mathbf{w}$, $\mathbf{b} = \mathbf{u} + \frac{1}{2}\mathbf{w}$ and integrating with respect to u, we find for the frequency function of the range the expression

$$f_{\mathbf{w}}(w) = n(n-1)\int_{-\infty}^{\infty} \phi(u-\tfrac{1}{2}w)\,\phi(u+\tfrac{1}{2}w)\,[\Phi(u+\tfrac{1}{2}w)-\Phi(u-\tfrac{1}{2}w)]^{n-2}\,du. \tag{2·2}$$

The object of our inquiry is the limiting form of the distribution (2·2). It proves, however, more advantageous to pass to the limit in the joint distribution of $\mathbf{a}$, $\mathbf{b}$ or $\mathbf{u}$, $\mathbf{w}$, before integrating with respect to u.

The asymptotical distribution of $\mathbf{a}$ and $\mathbf{b}$ has been investigated by Fisher & Tippett (1928), and Gumbel (1936) (cf. also Cramér, 1945, p. 376). According to these authors, we have

$$\left.\begin{aligned} &E(\mathbf{u}) = 0, & &D(\mathbf{u}) = O(\log^{-\frac{1}{2}} n),\\ &E(\mathbf{w}) = 2\sqrt{(2\log n)} + O\left(\frac{\log\log n}{\sqrt{(\log n)}}\right), & &D(\mathbf{w}) = O(\log^{-\frac{1}{2}} n).\end{aligned}\right\} \tag{2·3}$$

From the formulae quoted it is seen that $\mathbf{u} \to 0$, $\mathbf{w} \to \infty$ *in probability* as $n \to \infty$. Our first task must, consequently, be a transformation of the variables $\mathbf{a}$, $\mathbf{b}$—or $\mathbf{u}$, $\mathbf{w}$—depending on n and intended to stabilize the probability mass, in order to provide a limiting distribution.

* Prof. H. Wold has kindly directed my attention to this problem.

† $\Phi(x)$ denotes the distribution function and $\phi(x) = \Phi'(x)$ the frequency function of the normal distribution with mean at $x = 0$ and unit standard deviation.

Following the example of the authors mentioned above, we should have to introduce the new variables

$$\mathbf{a}' = n\Phi(\mathbf{a}), \quad \mathbf{b}' = n\Phi(-\mathbf{b}).$$

For our purpose it proves, however, advantageous to subject $\mathbf{a}'$ and $\mathbf{b}'$ to a new transformation, independent of n, taking

$$\left.\begin{aligned} \mathbf{x}e^{\mathbf{y}} &= 2n\Phi(\mathbf{a}) \quad\ = 2n\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u}), \\ \mathbf{x}e^{-\mathbf{y}} &= 2n\Phi(-\mathbf{b}) = 2n\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u}). \end{aligned}\right\} \tag{2·4}$$

Conversely,

$$\left.\begin{aligned} \mathbf{x} &= 2n\sqrt{[\Phi(\mathbf{a})\,\Phi(-\mathbf{b})]} = 2n\sqrt{[\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u})\,\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u})]}, \\ \mathbf{y} &= \tfrac{1}{2}\log\frac{\Phi(\mathbf{a})}{\Phi(-\mathbf{b})} = \tfrac{1}{2}\log\frac{\Phi(-\frac{1}{2}\mathbf{w}+\mathbf{u})}{\Phi(-\frac{1}{2}\mathbf{w}-\mathbf{u})}. \end{aligned}\right\} \tag{2·5}$$

As $\mathbf{a} \leqq \mathbf{b}$ and thus $\Phi(\mathbf{a})+\Phi(-\mathbf{b}) \leqq 1$, it follows from (2·4), that $\mathbf{x}$, $\mathbf{y}$ are subjected to the restrictions

$$\mathbf{x} \geqq 0, \quad \mathbf{x}\cosh\mathbf{y} \leqq n. \tag{2·6}$$

Performing the transformation, we find

$$\left|\frac{\partial(a,b)}{\partial(x,y)}\right| = \frac{x}{2n^2\phi(a)\,\phi(b)}, \tag{2·7}$$

and thus, letting $f_n(x,y)$ denote the joint-frequency function of $\mathbf{x}, \mathbf{y}$,

$$f_n(x,y) = \frac{n-1}{2n}x\left(1-\frac{x\cosh y}{n}\right)^{n-2}. \tag{2·8}$$

This formula is valid in the region (2·6); outside of it, we have to put $f_n(x,y) = 0$.

The new variables $\mathbf{x}, \mathbf{y}$ depend, of course, on $\mathbf{u}$ as well as $\mathbf{w}$. It will, however, be shown later, that $\mathbf{x}$, for large n, tends to coincide with the variable

$$\mathbf{x}^* = 2n\Phi(-\tfrac{1}{2}\mathbf{w}),$$

which depends exclusively on $\mathbf{w}$. For testing purposes, the former variable may thus, in large samples, be used as a substitute for the range. These considerations justify the transformation (2·4) as well as a closer study of the distribution of $\mathbf{x}$ and its limiting form.

3. *Limit passage and remainder term.* The limiting form of the joint-frequency function (2·8) is immediately seen to be

$$f(x,y) = \tfrac{1}{2}xe^{-x\cosh y} \quad (x \geqq 0). \tag{3·1}$$

The integral of this function, taken over the whole half-plane $x \geqq 0$, is easily seen to equal 1; (3·1) is, consequently, the frequency function of a well-determined two-dimensional distribution.

Let the marginal distribution functions in x, corresponding to (2·8) and (3·1), be denoted by $F_n(x)$ and $F(x)$ respectively. Our next task will be to estimate the remainder $|F_n(x)-F(x)|$, which is, obviously, at most equal to the integral

$$\Delta_n = \int_0^x \int_0^\infty 2\,|f_n(\xi,\eta)-f(\xi,\eta)|\,d\xi\,d\eta. \tag{3·2}$$

To begin with, we estimate the quotient f_n/f upwards. By differentiation with respect to the variable $z = x\cosh y$, this quotient is found to attain the maximum value

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)^{n-2}e^2 = 1+\frac{1}{n}+O\left(\frac{1}{n^2}\right)$$

for $z = 2$. We thus find, for example,

$$\frac{f_n}{f} < 1+\frac{3}{2n} \quad (n \geqq 5). \tag{3·3}$$

For the further estimations, it proves necessary to divide the domain of integration in (3·2) into an interior and an exterior part by means of a convenient abscissa $\eta = y$. In order to secure the Maclaurin expansion of $\log\left(1-\frac{1}{n}\xi\cosh\eta\right)$ within the interior region, we have to choose y so as to satisfy the inequality $\frac{x\cosh y}{n} \leqq k$ with an appropriate $k<1$. Taking, for simplicity, $k = 1-\sqrt{\frac{1}{2}}$ and observing that $\cosh y \leqq e^y$, we see that the condition mentioned is fulfilled if

$$e^y \leqq \frac{n}{x}(1-\sqrt{\tfrac{1}{2}}). \tag{3·4}$$

Now we may estimate f_n/f downwards in the interior domain of integration. Expanding $\log\left(1-\frac{1}{n}\xi\cosh\eta\right)$, we find

$$\log\frac{f_n}{f} = \log\left(1-\frac{1}{n}\right)+\frac{2}{n}\xi\cosh\eta-\frac{n-2}{2n^2}\xi^2\cosh^2\eta\left(1-\vartheta\frac{\xi\cosh\eta}{n}\right)^{-2} \quad (0<\vartheta<1). \tag{3·5}$$

According to the determination of y, the remainder factor is seen to be <2 for $\xi \leqq x$, $\eta \leqq y$. For $n \geqq 3$, we have $\log\left(1-\frac{1}{n}\right) > -\frac{3}{2n}$. Omitting, further, the positive term in (3·5) and replacing $n-2$ by n, we find

$$\frac{f_n(\xi,\eta)}{f(\xi,\eta)}-1 > \log\frac{f_n(\xi,\eta)}{f(\xi,\eta)} > -\frac{\frac{3}{2}+\xi^2\cosh^2\eta}{n};$$

hence, combining with (3·3),

$$\left|\frac{f_n(\xi,\eta)}{f(\xi,\eta)}-1\right| < \frac{\frac{3}{2}+\xi^2\cosh^2\eta}{n} \quad (\xi\leqq x, \eta\leqq y;\ n\geqq 5). \tag{3·6}$$

In the exterior domain of integration, (3·3) directly yields

$$|f_n(\xi,\eta)-f(\xi,\eta)| < f(\xi,\eta) \quad (\xi\leqq x, \eta \geqq y). \tag{3·7}$$

We proceed to the estimation of the integral (3·2), denoting its interior and exterior part by I_1 and I_2 respectively. For the former we have, according to (3·6), the inequality

$$I_1 = \int_0^x\int_0^y\left|\frac{f_n}{f}-1\right|2f\,d\xi\,d\eta < \frac{1}{n}\int_0^x\int_0^y(\tfrac{3}{2}\xi+\xi^3\cosh^2\eta)\,e^{-\xi\cosh\eta}\,d\xi\,d\eta, \tag{3·8}$$

for the latter, according to (3·7),

$$I_2 = \int_0^x\int_y^\infty 2\,|f_n-f|\,d\xi\,d\eta < \int_0^x\int_y^\infty \xi e^{-\xi\cosh\eta}\,d\xi\,d\eta. \tag{3·9}$$

The integration with respect to ξ may be explicitly performed. We have, in fact, putting for brevity $\cosh\eta = a$,

$$\int_0^x \xi e^{-a\xi}\,d\xi = \frac{1}{a^2}\{1-e^{-ax}[1+ax]\}, \tag{3·10}$$

$$\int_0^x \xi^3 e^{-a\xi}\,d\xi = \frac{6}{a^4}\left\{1-e^{-ax}\left[1+ax+\frac{(ax)^2}{2}+\frac{(ax)^3}{6}\right]\right\}. \tag{3·11}$$

In order to deduce remainder formulas for (a) moderate, (b) small x, we omit in (3·10) and (3·11), (a) all the negative terms, (b) the terms with x^2 and x^3. According to the Maclaurin expansion

$$e^{ax} = 1+ax+e^{\vartheta ax}\frac{(ax)^2}{2} \quad (0<\vartheta<1),$$

the expression in curled brackets in (3·10) is at most equal to $\frac{1}{2}a^2x^2$. Inserting these estimations in (3·8), we obtain for the interior integral the inequalities

$$I_1 < \frac{15}{2n}\int_0^y \frac{d\eta}{\cosh^2\eta} = \frac{15}{2n}\operatorname{tgh} y < \frac{15}{2n}, \tag{3·12a}$$

$$I_1 < \frac{15}{4n}x^2\int_0^y d\eta \leqslant \frac{4x^2}{n}y. \tag{3·12b}$$

For the exterior integral, (3·10) yields

$$I_2 < \int_y^\infty \frac{d\eta}{\cosh^2\eta} = 1 - \operatorname{tgh} y < 2e^{-2y}. \tag{3·13}$$

Finally, we have to join the results (3·12) and (3·13). Combining, first, (3·12a) with (3·13) and determining e^{-y} from (3·4) (taken with the equality sign), we obtain, after some slight simplifications in the numerical coefficients,

$$\Delta_n < \frac{8}{n}\left(1 + \frac{3x^2}{n}\right) \quad (n \geqq 5). \tag{3·14a}$$

Combining, on the other hand, (3·12b) with (3·13), we find

$$\Delta_n < \frac{4x^2}{n}y + 2e^{-2y}.$$

This expression attains, for fixed x and n, its minimum when $y = \log\frac{\sqrt{n}}{x}$. For $n \geqq 12$, this value of y also satisfies (3·4), and we obtain, as a parallel estimate to (3·14a),

$$\Delta_n < \frac{4x^2}{n}\left(\log\frac{\sqrt{n}}{x} + \frac{1}{2}\right) \quad (n \geqq 12). \tag{3·14b}$$

The formulas (3·14a, b) are both valid for all positive x and all $n \geqq 12$.

4. *The asymptotical distribution.* Having established the limiting distribution of the variable $\mathbf{x}$ defined in (2·5), we are going to examine its properties.

The frequency function of the distribution considered reads, according to (3·1),

$$f(x) = x\int_0^\infty e^{-x\cosh y}\,dy = x\int_1^\infty \frac{e^{-xt}}{\sqrt{(t^2-1)}}\,dt. \tag{4·1}$$

Changing the order of integration, we easily find the distribution function, the mean and the variance of (4·1) to be

$$F(x) = 1 - \int_0^\infty \frac{1 + x\cosh y}{\cosh^2 y}e^{-x\cosh y}\,dy = 1 - \int_1^\infty \frac{1+xt}{t^2\sqrt{(t^2-1)}}e^{-xt}\,dt, \tag{4·1'}$$

$$E(\mathbf{x}) = \tfrac{1}{2}\pi, \quad D^2(\mathbf{x}) = 4 - \tfrac{1}{4}\pi^2. \tag{4·2}$$

The numerical evaluation of the distribution is much simplified by the fact that $f(x)$ as well as $F(x)$ is closely connected with certain *Bessel functions*. Denote

$$\phi(x) = \int_0^\infty e^{-x\cosh y}\,dy = \int_1^\infty \frac{e^{-xt}}{\sqrt{(t^2-1)}}\,dt. \tag{4·3}$$

By differentiation and partial integration, this function is found to satisfy the differential equation

$$\phi''(x) + \frac{1}{x}\phi'(x) - \phi(x) = 0. \tag{4·4}$$

Changing x into $-ix$, we obtain for the function $\psi(x) = \phi(-ix)$ the equation

$$\psi''(x) + \frac{1}{x}\psi'(x) + \psi(x) = 0; \tag{4·4'}$$

hence, $\psi(x)$ is a Bessel function of order zero.

In order to specify this function, we will deduce an asymptotical expression for the function (4·3), valid for large x. For this purpose, we make in the latter integral (4·3) the substitution $t = 1 + u/x$ and write

$$\left(1 + \frac{u}{2x}\right)^{-\frac{1}{2}} = 1 - \vartheta\frac{u}{4x} \quad (0 < \vartheta < 1).$$

Performing the integration, we obtain

$$\phi(x) = \sqrt{\left(\frac{\pi}{2}\right)}\, x^{-\frac{1}{2}}\, e^{-x}\left(1 + O\left(\frac{1}{x}\right)\right), \tag{4·5}$$

which shows that the Bessel function $\psi(x) = \phi(-ix)$ tends to zero for $x \to +i\infty$. This function is, consequently, proportional to the *Hankel function* $H_0^{(1)}(x)$ (cf. Jahnke-Emde, 1909, p. 94). Comparing the asymptotical expressions of $\phi(x)$ and $iH_0^{(1)}(ix)$, we find the proportional factor to be $\frac{1}{2}\pi$, whence

$$f(x) = x\frac{\pi i}{2}H_0^{(1)}(ix). \tag{4·6}$$

We proceed to the calculation of $F(x)$. Every integral of $xH_0^{(1)}(x)$ is (cf. Jahnke-Emde, p. 165) of the form $xH_1^{(1)}(x) + \text{Const.}$, where $H_1^{(1)}(x)$ is the *first* order Hankel function corresponding to $H_0^{(1)}(x)$; consequently,

$$F(x) = \frac{\pi x}{2}H_1^{(1)}(ix) + C.$$

Now $\frac{\pi x}{2}H_1^{(1)}(ix)$ tends to zero as $-(\frac{1}{2}\pi x)^{\frac{1}{2}}\, e^{-x}$ for $x \to \infty$ (cf. Jahnke-Emde, 1909, p. 101); hence $C = 1$ and

$$F(x) = 1 - x\left[-\frac{\pi}{2}H_1^{(1)}(ix)\right]. \tag{4·7}$$

For small x, $F(x)$ has the expansion

$$F(x) = \left(\log\frac{2}{\gamma x} + \frac{1}{2}\right)\frac{x^2}{2} + \left(\log\frac{2}{\gamma x} + \frac{5}{4}\right)\frac{x^4}{16} + \dots, \tag{4·8}$$

where

$$\log\frac{2}{\gamma} = 0{\cdot}11593\dots. \tag{4·9}$$

The factors of x in (4·6) and (4·7) are tabulated in Jahnke-Emde (1909, pp. 135–6). Below, we give a short table of $f(x)$ and $F(x)$. The corresponding curves are seen in Fig. 1.

5. *Connexion between the variable* $\mathbf{x}$ *and the range.* We now turn back to the original object of our inquiry: the asymptotical distribution of the range.

Consider the variable

$$\mathbf{x} = 2n\sqrt{[\Phi(-\tfrac{1}{2}\mathbf{w} + \mathbf{u})\,\Phi(-\tfrac{1}{2}\mathbf{w} - \mathbf{u})]} \tag{5·1}$$

introduced in (2·4). As mentioned earlier,

$$\mathbf{w} \to \infty, \quad \mathbf{u} \to 0 \text{ in probability} \quad (n \to \infty). \tag{5·2}$$

Under such circumstances, for large n, $\mathbf{x}$ may be expected to behave substantially as the variable

$$\mathbf{x}^* = 2n\Phi(-\tfrac{1}{2}\mathbf{w}), \tag{5·3}$$

which depends *exclusively on the range.*

We shall now prove that $\mathbf{x}^*/\mathbf{x} \to 1$ in probability as $n \to \infty$. According to the well-known asymptotic formula

$$\Phi(-x) = \frac{1}{x\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\left(1 - \frac{\vartheta}{x^2}\right) \quad (x > 0); \; 0\,(<\vartheta < 1),$$

we may, for $|\mathbf{u}| < \frac{1}{2}\mathbf{w}$, write

$$\frac{\mathbf{x}^*}{\mathbf{x}} = e^{\frac{1}{2}\mathbf{u}^2}\left(1 - \frac{4\mathbf{u}^2}{\mathbf{w}^2}\right)^{\frac{1}{2}} \{1 + O[(\tfrac{1}{2}\mathbf{w} - |\mathbf{u}|)^{-2}]\}.$$

x	$f(x)$	$F(x)$	x	$f(x)$	$F(x)$
0·0	0·0000	0·0000	1·5	0·3207	0·5839
0·1	0·2427	0·0146	2·0	0·2278	0·7202
0·2	0·3505	0·0448	2·5	0·1559	0·8153
0·3	0·4118	0·0832	3·0	0·1042	0·8795
0·4	0·4458	0·1262	4·0	0·0446	0·9501
0·5	0·4622	0·1718	5·0	0·0185	0·9798
0·6	0·4665	0·2183	6·0	0·0075	0·9919
0·7	0·4624	0·2648	7·0	0·0030	0·9968
0·8	0·4522	0·3106	8·0	0·0012	0·9988
0·9	0·4380	0·3552	9·0	0·0005	0·9995
1·0	0·4210	0·3981	10·0	0·0002	0·9998

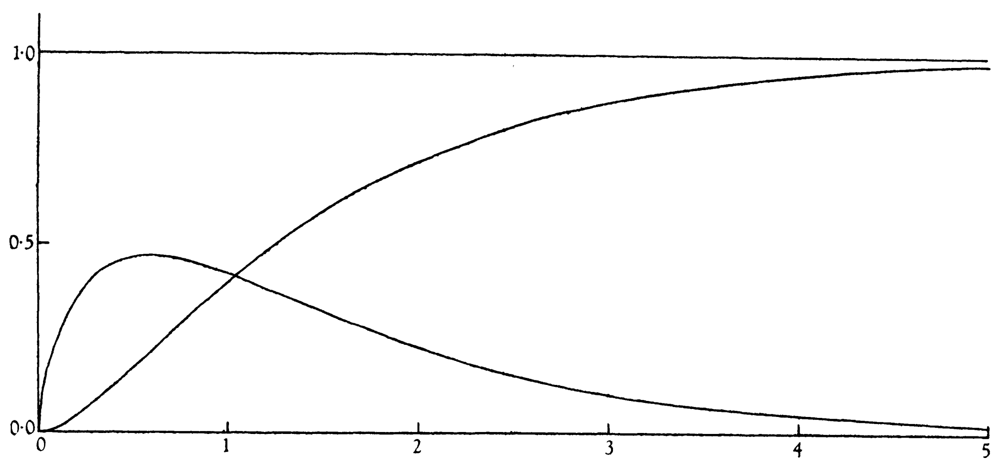

Fig. 1

Given an arbitrary $\epsilon > 0$, we obviously may find two positive numbers u_ϵ and w_ϵ $(> u_\epsilon)$ such that

$$\left|\frac{\mathbf{x}^*}{\mathbf{x}} - 1\right| < \epsilon \quad \text{if} \quad \mathbf{w} \geqq w_\epsilon, \quad |\mathbf{u}| \leqq u_\epsilon. \tag{5·4}$$

On account of (5·2), we may, on the other hand, choose n_ϵ so that the probability of the simultaneous validity of the latter inequalities in (5·4) exceeds $1 - \epsilon$ if $n \geqq n_\epsilon$. Consequently,

$$P\left\{\left|\frac{\mathbf{x}^*}{\mathbf{x}} - 1\right| < \epsilon\right\} > 1 - \epsilon \quad (n \geqq n_\epsilon), \tag{5·5}$$

which proves our statement.

As shown in section 3, the distribution function $F_n(x)$ of $\mathbf{x}$ converges to $F(x)$ as $n \to \infty$. Since $F(0) = 0$, it follows from (5·5), by a well-known method of argument, that the distribution function $F_n^*(x)$ of $\mathbf{x}^*$ converges to the same limiting function. The asymptotical distribution of the range, suitably transformed, is hereby established.

For practical purposes, it would, of course, be desirable to possess a reasonably accurate estimate of the remainder $F_n^*(x) - F(x)$, or at least an estimate of the difference $F_n^*(x) - F_n(x)$, to be combined with the results (3·14).

For $n = 20$, the accuracy of $F(x)$ as substitute for $F_n^*(x)$ may be checked by means of Hartley's (1942) tables. The discrepancy amounts to about 0·004 for $x = 0{\cdot}1$, 0·025 for $x = 1$ and 0·010 for $x = 4$.

The theoretical evaluation of $F_n^*(x) - F(x)$ seems to be somewhat complicated and, besides, of little use since $\mathbf{x}^*$, for most purposes, may be replaced by $\mathbf{x}$. A few remarks concerning the relations between $\mathbf{x}$, $\mathbf{x}^*$ and their distribution functions will, however, be added below.

To begin with, we note that always $\mathbf{x} \leqq \mathbf{x}^*$, the equality sign being valid only if $\mathbf{u} = 0$. Consider, in fact, the function $x(u)$, defined by (5·1) for a fixed w. Inserting for Φ its analytical expression, we easily find that $D^{(2)} \log x(u) \leqq 0$ for all u. Hence, $x(u)$ has no minimum and at most one maximum, and the latter is, by symmetry, seen to be attained for $u = 0$, being thus equal to $\mathbf{x}^*$.

From $\mathbf{x} \leqq \mathbf{x}^*$, it follows that $F_n^*(x) \leqq F_n(x)$ for all x. We will show that the difference $F_n(x) - F_n^*(x)$ may be expressed as a double integral.

The variables $\mathbf{u}$ and $\mathbf{w}$ are, according to (2·4), well-determined functions of $\mathbf{x}$ and $\mathbf{y}$ in the region (2·6); and so is the variable $\mathbf{x}^*$, on account of (5·3).

On the *level curve* $\mathbf{x}^* = x_0$, $\mathbf{w}$ has a constant value w_0, determined by

$$2n\Phi(-\tfrac{1}{2}w_0) = x_0, \tag{5·6}$$

and this curve is, consequently, given in parametric form by the equations

$$x = 2n\sqrt{[\Phi(-\tfrac{1}{2}w_0 + u)\,\Phi(-\tfrac{1}{2}w_0 - u)]}, \quad y = \tfrac{1}{2}\log\frac{\Phi(-\tfrac{1}{2}w_0 + u)}{\Phi(-\tfrac{1}{2}w_0 - u)}, \tag{5·7}$$

where u runs through all values from $-\infty$ to $+\infty$. The latter function (5·7) being, obviously, monotonously increasing, we may imagine u eliminated, writing (5·7) in the form

$$x = \xi_n(x_0, y) \quad (-\infty < y < \infty). \tag{5·7′}$$

From the proof of the inequality $\mathbf{x} \leqq \mathbf{x}^*$ given above, it follows that the function (5·7′) has a single maximum for $y = 0$. When $y \to \pm\infty$, the function obviously tends to zero.

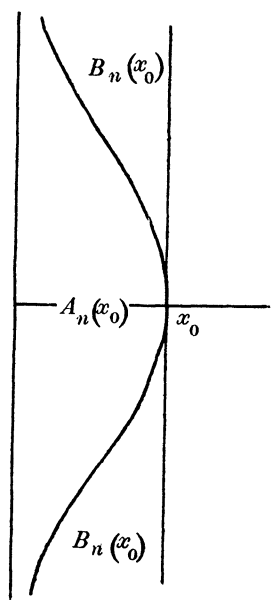

Fig. 2

The inequality $\mathbf{x}^* \leqq x_0$ is fulfilled on the left side of the curve (5·7′), the inequality $\mathbf{x} \leqq x_0$ on the left side of the straight line $\mathbf{x} = x_0$. Let us for brevity denote the regions (cf. fig. 2)

$$0 \leqq \mathbf{x} \leqq \xi_n(x_0, \mathbf{y}), \quad \xi_n(x_0, \mathbf{y}), < \mathbf{x} \leqq x_0 \tag{5·8}$$

by $A_n(x_0)$ and $B_n(x_0)$ respectively. The difference $F_n(x_0) - F_n^*(x_0)$ is, then, the probability of the points $\mathbf{x}$, $\mathbf{y}$ falling within the region $B_n(x_0)$. Dropping the indices 0, we thus obtain the expression sought for

$$F_n(x) - F_n^*(x) = \iint_{B_n(x)} f_n(\xi, \eta)\, d\xi\, d\eta. \tag{5·9}$$

Comparing, finally, the transformed range distribution function $F_n^*(x)$ directly with its limiting form $F(x)$, we find

$$\begin{aligned} F_n^*(x) - F(x) &= [F_n(x) - F(x)] - [F_n(x) - F_n^*(x)] \\ &= \iint_{\xi \leqq x} (f_n - f)\, d\xi\, d\eta - \iint_{B_n(x)} f_n\, d\xi\, d\eta \\ &= \iint_{A_n(x)} (f_n - f)\, d\xi\, d\eta - \iint_{B_n(x)} f\, d\xi\, d\eta. \end{aligned} \tag{5·10}$$

The former integral is, obviously, at most equal to the remainder expression Δ_n in (3·2), estimated in (3·14).

6. *Conclusion.* Our main results may be summarized in the following theorem:

THEOREM. Consider a sample of n observations from an infinite normal population with mean 0 and standard deviation 1. Let **a** be the smallest, **b** the greatest of the observed values, and put

$$\mathbf{x} = 2n\sqrt{[\Phi(\mathbf{a})\,\Phi(-\mathbf{b})]}, \quad \mathbf{x}^* = 2n\Phi\left(-\frac{\mathbf{b}-\mathbf{a}}{2}\right),$$

the latter variable being evidently a simple transformation of the range of the sample. Then

(1) $\mathbf{x} \leqq \mathbf{x}^*$; $\mathbf{x}^*/\mathbf{x} \to 1$ in probability $(n \to \infty)$.

(2) The distribution functions $F_n(x)$ and $F_n^*(x)$ of $\mathbf{x}$ and $\mathbf{x}^*$ tend, for $n \to \infty$, to the common limit

$$F(x) = 1 - \int_1^\infty \frac{1+xt}{t^2\sqrt{(t^2-1)}}\, e^{-xt}\, dt = 1 + \frac{\pi x}{2} H_1^{(1)}(ix),$$

where $H_1^{(1)}(z)$ is the first order Bessel function, which vanishes as $-\left(\frac{\pi z}{2i}\right)^{-\frac{1}{2}} e^{iz}$ for $z \to +i\infty$.

(3) For $n \geqq 12$, $F_n(x)$ satisfies the inequalities

$$|F_n(x) - F(x)| < \frac{8}{n}\left(1 + \frac{3x^2}{n}\right), \quad |F_n(x) - F(x)| < \frac{4x^2}{n}\left(\log\frac{\sqrt{n}}{x} + \frac{1}{2}\right).$$

7. *Generalization.* A great part of our conclusions does not presuppose the normality of the parental population. Thus, the distribution (2·8) of the variables **x**, **y** defined by (2·5) is the same for any continuous probability law and so, consequently, is its limiting form; however, if the parental distribution is non-symmetrical, with distribution function $G(x)$, say, the factor $\Phi(-\mathbf{b})$ in (2·5) must, of course, be replaced by $1 - G(\mathbf{b})$ instead of $G(-\mathbf{b})$, and the variable $\mathbf{x}^*$ is to be defined by

$$\mathbf{x}^* = 2n\sqrt{\{G(-\tfrac{1}{2}\mathbf{w})\,[1 - G(\tfrac{1}{2}\mathbf{w})]\}}.$$

The proof of the statement $\mathbf{x}^*/\mathbf{x} \to 1$ requires, however, convenient assumptions concerning the parental distribution. It can be proved that the assertion mentioned—and, consequently, the theorem stated above—are valid if the frequency function of this distribution is of the form

$$g(x) = C\exp\left[-\frac{1}{p}|x|^p\right],$$

where $1 < p \leqq 2$.

REFERENCES

CRAMÉR, H. (1945). *Mathematical Methods of Statistics.* Uppsala.

FISHER, R. A. & TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24**, 180.

GUMBEL, E. J. (1936). Les valeurs extrêmes des distribution statistiques. *Ann. Inst. Poincaré*, **5**, 115–58.

HARTLEY, H. O. (1942). The range in random samples. *Biometrika*, **32**, 334–48.

HARTLEY, H. O. & PEARSON, E. S. (1942). The probability integral of the range in samples of *n* observations from a normal population. *Biometrika*, **32**, 301–10.

JAHNKE, E. & EMDE, F. (1909). *Funktionentafeln.* Leipzig and Berlin.

MCKAY, A. T. & PEARSON, E. S. (1933). A note on the distribution of range in samples of *n*. *Biometrika*, **25**, 415–20.

PEARSON, E. S. (1926). A further note on the distribution of range in samples, taken from a normal population. *Biometrika*, **18**, 173–94.

PEARSON, E. S. (1932). The percentage limits for the distribution of range in samples from a normal population. *Biometrika*, **24**, 404–17.

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, **17**, 364–87.