To find such a k we consider

$$I(t) = \int_{\sum_{i=1}^{n} \rho_i Y_i \geq t} \cdots \int Q(Y_1, \cdots, Y_n) \left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma} \right) - \epsilon \right] dY_1 \cdots dY_n .$$

As t tends to $-\infty$, $I(t)$ tends to $L(1)$, where L was defined by (4.3). Since the ϵ-quantile in Π was less than X_0 it follows that $I(-\infty) = L(1) > 0$. Since $I(t) < 0$ for large t, there exists t_0 such that $I(t_0) = 0$, and clearly,

$$\psi\left(\frac{X_0 - t_0}{\sigma}\right) - \epsilon > 0.$$

Setting in (4.5) $k = [\psi((X_0 - t_0)/\sigma) - \epsilon]^{-1}$, one obtains a φ_0 such that

$$L(\varphi_0) = I(t_0) = 0.$$

The selection φ_0 is the linear truncation to the set $\sum_{i=1}^{n} \rho_i Y_i \geq t_0$.

By a similar and somewhat simpler argument one proves the following theorem.

THEOREM 2. *A selection such that*

1° *in* Π^* *the mean of* X *has a value greater than or equal to a pre-assigned number* $m > 0$,

2° *the fraction retained is maximum,*

is a linear truncation to a set $\sum_{i=1}^{n} \rho_i Y_i \geq t_0$.

An immediate consequence of Theorems 1 and 2 is that a linear truncation, using a properly determined weighted score $\sum_{i=1}^{n} \rho_i Y_i$ and cutting score t_0, is more economical than any truncation to a set $Y_i \geq t_i$, $i = 1, 2, \cdots, n$, that is than any truncation performed on each admission score separately.

REFERENCES

[1] Z. W. BIRNBAUM, "Effect of linear truncation on a multinormal population," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 272–279.
[2] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. I (1936), pp. 1–37, particularly pp. 10–11.

THE DISTRIBUTION OF DISTANCE IN A HYPERSPHERE

BY J. M. HAMMERSLEY

University of Oxford

1. Summary. Deltheil ([1], pp. 114–120) has considered the distribution of distance in an n-dimensional hypersphere. In this paper I put his results (17) in a more compact form (16); and I investigate in greater detail the asymptotic form of the distribution for large n, for which the rather surprising result emerges that this distance is almost always nearly equal to the distance between the

extremities of two orthogonal radii. I came to study this distribution by the need to compute a doubly-threefold integral, which measures the damage caused to plants by the presence of radioactive tracers in their fertilizers; for the distribution affords a method of evaluating numerically certain multiple integrals. I hope to describe elsewhere this application of the theory.

2. Derivation of the frequency function. Let T_1 and T_2 be vector spaces of n and $2n$ dimensions respectively. Let P and Q be any pair of points in T_1. Denote by (PQ) the point in T_2, whose first n coordinates are the coordinates of P in T_1 and whose last n coordinates are the coordinates of Q in T_1. Let $\{P\}$ and $\{Q\}$ be point sets in T_1, and let $\{PQ\}$ be the point set in T_2 such that $(PQ) \in \{PQ\}$ if and only if both $P \in \{P\}$ and $Q \in \{Q\}$. Let $M_1\{P\}$ denote the n-dimensional measure of the point set $\{P\}$ in T_1, and let $M_2\{PQ\}$ denote the $2n$-dimensional measure of the point set $\{PQ\}$ in T_2. Then

$$M_2\{PQ\} = \int_{\{P\}} M_1\{Q\}\, dM_1\{P\}. \tag{1}$$

Let R be a fixed point in T_1; and let $S_n(a)$ be the n-dimensional hypersphere in T_1 with centre R and radius a. Let A and B be any two points chosen at random in $S_n(a)$, the distributions of A and B being independent and uniform over the interior of $S_n(a)$. Denote the distance AB by r; and let $\lambda = r/2a$, so that λ may take any value in the interval $0 \leq \lambda \leq 1$. We require the frequency function of λ, which we shall denote by $f_n(\lambda)$.

The volume content of $S_n(a)$ is

$$V_n(a) = \pi^{n/2}a^n/\Gamma(\tfrac{1}{2}n + 1); \tag{2}$$

and the content of the segment of the surface of $S_n(a)$ bounded by a right hyperspherical cone, whose vertex is at R and whose line generators make a fixed semi-vertical angle θ with a fixed radius of $S_n(a)$, is

$$U_n(a, \theta) = \frac{2\pi^{(n-1)/2}a^{n-1}}{\Gamma(\frac{1}{2}n - \frac{1}{2})}\int_0^\theta \sin^{n-2}\phi\, d\phi. \tag{3}$$

As a particular case of (2), the whole surface of $S_n(a)$ has content

$$U_n(a, \pi) = 2\pi^{n/2}a^{n-1}/\Gamma(\tfrac{1}{2}n). \tag{4}$$

Let $\{AB\}$ be the point set in T_2 such that $(AB) \in \{AB\}$ if and only if the corresponding points A and B satisfy all the inequalities

$$0 \leq RA \leq a, \qquad 0 \leq RB \leq a, \qquad r \leq AB \leq r + dr. \tag{5}$$

Then, by the definition of $f_n(\lambda)$,

$$M_2\{AB\} \propto f_n(r/2a)\, dr;$$

but since

$$\int_0^{2a} M_2\{AB\}\, dr = V_n^2, \qquad \int_0^{2a} f_n(r/2a)\, dr/2a = 1,$$

we have

(6) $$M_2\{AB\} = V_n^2 f_n(r/2a)\,dr/2a \equiv p_n(r, a)\,dr, \qquad \text{say.}$$

Consider also the point set $\{CD\}$ in T_2 such that $(CD) \epsilon \{CD\}$ if and only if the corresponding points C and D satisfy all the inequalities

(7) $$0 \leq RC \leq a + da, \qquad a \leq RD \leq a + da, \qquad r \leq CD \leq r + dr.$$

For each fixed D of $\{D\}$, C is constrained to lie on the segment of the hyperspherical shell of thickness dr, radius r, and centre D, bounded by the intersection of this shell with $S_n(a + da)$. The hyperspherical cone, with vertex D, whose line generators all pass through this intersection, has a semi-vertical angle θ given by

(8) $$\cos\theta = r/2a = \lambda;$$

and so, from (3), the M_1 of all C which satisfy (7) for each fixed D is $U_n(r, \arccos\lambda)\,dr$. On the other hand the M_1 of all D which satisfy (7) is the content of the hyperspherical shell of thickness da, radius a, and centre R, and is thus $U_n(a, \pi)\,da$ by virtue of (4). Consequently, from (1)

(9) $$M_2\{CD\} = U_n(r, \arccos\lambda)U_n(a, \pi)\,da\,dr.$$

On the other hand, by symmetry, $M_2\{CD\} = \frac{1}{2}M_2\{EF\}$, where $(EF) \epsilon \{EF\}$ if and only if the corresponding points E and F satisfy either all the inequalities

$$0 \leq RE \leq a + da, \qquad a \leq RF \leq a + da, \qquad r \leq EF \leq r + dr,$$

or all the inequalities

$$0 \leq RF \leq a + da, \qquad a \leq RE \leq a + da, \qquad r \leq EF \leq r + dr.$$

We can express this in another way by saying that $(EF) \epsilon \{EF\}$ if and only if the corresponding points E and F satisfy all the inequalities

$$0 \leq RE \leq a + da, \qquad 0 \leq RF \leq a + da, \qquad r \leq EF \leq r + dr,$$

but do not satisfy all the inequalities

$$0 \leq RE \leq a, \qquad 0 \leq RF \leq a, \qquad r \leq EF \leq r + dr.$$

From this second point of view we see that

$$M_2\{EF\} = p_n(r, a + da)\,dr - p_n(r, a)\,dr = \frac{\partial}{\partial a} p_n(r, a)\,dr\,da;$$

and so

(10) $$M_2\{CD\} = \frac{1}{2}\frac{\partial}{\partial a} p_n(r, a)\,dr\,da.$$

Then from (2), (3), (4), (6), (9), and (10).

(11) $$\frac{1}{2}\frac{\partial}{\partial a}\left\{\frac{\pi^n a^{2n}}{[\Gamma(\frac{1}{2}n + 1)]^2} f_n\left(\frac{r}{2a}\right)\cdot\frac{1}{2a}\right\} = \left\{\frac{2\pi^{(n-1)/2} r^{n-1}}{\Gamma(\frac{1}{2}n - \frac{1}{2})}\int_0^{\arccos\lambda} \sin^{n-2}\phi\,d\phi\right\}\left\{\frac{2\pi^{n/2} a^{n-1}}{\Gamma(\frac{1}{2}n)}\right\}.$$

By performing the partial differentiation on the left-hand side, then substituting $z = \cos\phi$ and $r = 2a\lambda$, and using the relations

$$\Gamma(\tfrac{1}{2}n + 1) = \tfrac{1}{2}n\Gamma(\tfrac{1}{2}n), \qquad \pi^{1/2}\Gamma(n+1) = 2^n\Gamma(\tfrac{1}{2}n + \tfrac{1}{2})\Gamma(\tfrac{1}{2}n + 1),$$

$$B(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}n + \tfrac{1}{2}) = \{\Gamma(\tfrac{1}{2}n + \tfrac{1}{2})\}^2/\Gamma(n+1),$$

we reduce (11) to the form

$$(12)\quad (2n-1)f_n(\lambda) - \lambda f_n'(\lambda) = \frac{2n(n-1)\lambda^{n-1}}{B(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}n + \tfrac{1}{2})}\int_\lambda^1 (1 - z^2)^{(n-3)/2}\,dz.$$

We multiply (12) by $-\lambda^{-2n}$ and use the reduction formula

$$(13)\quad (n-1)\int_\lambda^1 (1-z^2)^{(n-3)/2}\,dz = n\int_\lambda^1 (1-z^2)^{(n-1)/2}\,dz + \lambda(1-\lambda^2)^{(n-1)/2}.$$

Each side of the resulting equation is a perfect differential coefficient, and upon integration we obtain

$$(14)\qquad f_n(\lambda) = \frac{2n\lambda^{n-1}}{B(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}n + \tfrac{1}{2})}\int_\lambda^1 (1-z^2)^{(n-1)/2}\,dz + c\lambda^{2n-1},$$

where C is the constant of integration. We obtain the cumulative distribution function by integrating (14) over 0 to λ,

$$(15)\quad F_n(\lambda) = (2\lambda)^n I_{1-\lambda^2}(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}) + I_{\lambda^2}(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}n + \tfrac{1}{2}) + C\lambda^{2n}/2n,$$

where $I_x(p, q)$ is the incomplete beta-function ratio

$$I_x(p,q) = \int_0^x z^{p-1}(1-z)^{q-1}\,dz/B(p,q)$$

tabulated by Pearson [2]. Putting $\lambda = 1$ in (15) we get

$$1 = F_n(1) = 1 + C/2n;$$

so $C = 0$, and we have the final result

$$(16)\qquad f_n(\lambda) = 2^n n\lambda^{n-1} I_{1-\lambda^2}(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}).$$

This compact form may be compared with Deltheil's expression [1] for the frequency function of r, namely

$$(17)\qquad g_n(r) = \frac{n^2 r^{n-1}}{a^{2n}}\int_{r/2}^a \rho^{n-1} h_n\left(\frac{r}{\rho}\right) d\rho,$$

where

$$h_n(2\sin\theta) = \int_0^{\frac{1}{2}\pi-\theta} \sin^{n-2}\phi\,d\phi \Big/ \int_0^{\frac{1}{2}\pi} \sin^{n-2}\phi\,d\phi,$$

expressions which he evaluates only for the particular cases $n = 3, 5, 7, 9$.

Interesting particular cases of (16) are

$$(18)\qquad f_1(\lambda) = 2(1-\lambda), \qquad f_2(\lambda) = \frac{16}{\pi}\lambda\{\arccos\lambda - \lambda(1-\lambda^2)^{1/2}\},$$
$$f_3(\lambda) = 12\lambda^2(1-\lambda)^2(2+\lambda),$$

which give the appropriate frequency functions for a line, a circle, and a sphere respectively.

3. Recurrence relations and moments of the distribution. From (13) and (14) we have a recurrence relation for penadjacent values of n,

$$(19)\qquad \frac{f_n(\lambda)}{n} = 4\lambda^2\frac{f_{n-2}(\lambda)}{n-2} - \frac{2\Gamma(n)}{\{\Gamma(\frac{1}{2}n+\frac{1}{2})\}^2}\lambda^n(1-\lambda^2)^{(n-1)/2}.$$

In connection with (18) this shows that

$$(20)\quad f_{2n+1}(\lambda) = P_{4n+1}(\lambda), \qquad f_{2n}(\lambda) = P_{2n-1}(\lambda)\arccos\lambda + P_{4n-2}(\lambda)(1-\lambda^2)^{1/2},$$

where $P_N(\lambda)$ denotes an unspecified polynomial in λ of degree N or less.

From (16) the rth moment of $f_n(\lambda)$ about $\lambda = 0$ is

$$(21)\qquad \mu'_{nr} = \left\{\frac{n\Gamma(n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})}\right\}\left\{\frac{\Gamma(\frac{1}{2}n+\frac{1}{2}r+\frac{1}{2})}{(n+r)\,\Gamma(n+\frac{1}{2}r+1)}\right\}.$$

I have not been able to obtain the characteristic function of $f_n(\lambda)$ explicitly from (21) it appears to be of a higher type than the hypergeometric function.

4. The asymptotic form of the distribution for large n**.** The distribution function is, by (15),

$$(22)\qquad F_n(\lambda) = (2\lambda)^n I_{1-\lambda^2}(\tfrac{1}{2}n+\tfrac{1}{2},\tfrac{1}{2}) + I_{\lambda^2}(\tfrac{1}{2}n+\tfrac{1}{2},\tfrac{1}{2}n+\tfrac{1}{2}).$$

We show firstly that as $n \to \infty$ the first term of this expression tends to zero. This term is clearly zero if $\lambda = 0$. If $\lambda > 0$

$$\int_0^{1-\lambda^2} z^{(n-1)/2}(1-z)^{-1/2}\,dz \le \lambda^{-1}\int_0^{1-\lambda^2} z^{(n-1)/2}\,dz = (1-\lambda^2)^{(n+1)/2}/\tfrac{1}{2}(n+1)\,\lambda.$$

Hence

$$(2\lambda)^n I_{1-\lambda^2}(\tfrac{1}{2}n+\tfrac{1}{2},\tfrac{1}{2}) \le \frac{(2\lambda)^n\,\Gamma(\frac{1}{2}n+1)}{\pi^{1/2}\Gamma(\frac{1}{2}n+\frac{1}{2})}\cdot\frac{(1-\lambda^2)^{(n+1)/2}}{(\frac{1}{2}n+\frac{1}{2})\,\lambda}$$
$$\le \frac{2\Gamma(\frac{1}{2}n+1)}{\pi^{1/2}\Gamma(\frac{1}{2}n+\frac{3}{2})}(1-\lambda^2)\{4\lambda^2(1-\lambda^2)\}^{(n-1)/2} \le \frac{2\Gamma(\frac{1}{2}n+1)}{\pi^{1/2}\Gamma(\frac{1}{2}n+\frac{3}{2})} \to 0$$

as $n \to \infty$. Secondly, as $n \to \infty$

$$I_{\lambda^2}(\tfrac{1}{2}n+\tfrac{1}{2},\tfrac{1}{2}n+\tfrac{1}{2}) \sim N_{\lambda^2}(\tfrac{1}{2},1/4(n+1)) \sim N_{\lambda^2}(\tfrac{1}{2},1/4n),$$

(see Cramér [3] p. 252 with $p = q = \frac{1}{2}$), where $N_x(\mu,\sigma^2)$ is the normal cumulative distribution function of x for mean μ and variance σ^2. Hence λ is asymptotically distributed as $N_\lambda(1/\sqrt{2}, 1/8n)$; and the asymptotic distribution of r is $N_r(a\sqrt{2}, a^2/2n)$. This establishes the result stated in the summary.

It can also be proved, by considering the limiting form of the recurrence relation (19), that the frequency function f_n is asymptotically normal. The main difficulty of proving this fact lies in showing that the frequency function actually possesses a limiting form; and the proof is rather too long to be given here.

REFERENCES

[1] R. Deltheil, *Probabilités Géométriques. Traité du Calcul des Probabilités et de ses Applications: Tome II, Fasciscule II*, Gauthier-Villars, 1926.
[2] K. Pearson, *Tables of the Incomplete Beta-Function*, Cambridge University Press, 1934.
[3] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

A NOTE ON THE ASYMPTOTIC SIMULTANEOUS DISTRIBUTION OF THE SAMPLE MEDIAN AND THE MEAN DEVIATION FROM THE SAMPLE MEDIAN

By R. K. Zeigler

Bradley University

Consider a random sample of $2k + 1$ values from a one-dimensional distribution of the continuous type with cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x) = F'(x)$. Let the mean, standard deviation and median of the distribution be denoted by m, σ and θ respectively (θ assumed to be unique). We shall suppose that in some neighborhood of $x = \theta$, $f(x)$ has a continuous derivative $f'(x)$.

If we arrange the sample values in ascending order of magnitude:

$$x_1 < x_2 < \cdots < x_{2k+1},$$

there is a unique sample median x_{k+1} which we shall denote by ξ. The mean deviation from the sample median is then defined by

$$M = \frac{1}{2k} \sum_{i=1}^{2k+1} | x_i - \xi |.$$

In the material that follows we shall assume that the sample items have been ordered only to the extent that k of them are less than ξ and k of them are greater than ξ.

We then have the following

Theorem. *Let $f(x)$ be a pdf with finite second moment, continuous at $x = \theta$ with $f(\theta) \neq 0$. Then the simultaneous distribution of ξ and M is asymptotically normal. The means of the limiting distribution are θ, the population median, and u', the mean deviation from the population median, while the asymptotic variances are $1/4f^2(\theta)2k$ and $((m - \theta)^2 + \sigma^2 - u'^2)/2k$. The asymptotic expression for the correlation coefficient is $(m - \theta)/\sqrt{(m - \theta)^2 + \sigma^2 - u'^2}$.*

Proof: Let $u = (M - u')\sqrt{2k}$ and $v = (\xi - \theta)\sqrt{2k}$, where $u' = E \mid x - \theta \mid$. Then the simultaneous characteristic function of the two random variables u