Ann. Inst. Statist. Math.
32 (1980), Part A, 247-253

DISTRIBUTION OF A DISTANCE FUNCTION

A. H. KHAN AND MOHD. YAQUB

(Received June 7, 1977; revised Mar. 13, 1980)

Summary

The distributions of the square of the distance between a random point and a fixed point on an n-dimensional unit sphere when (i) the two points lie on the surface of whole sphere and (ii) the two points lie on the surface in the positive quadrant, have been derived and their moments obtained. Some test statistics are also proposed.

1. Introduction

Let $(p_1, p_2, \cdots, p_n)$ and $(q_1, q_2, \cdots, q_n)$ be two probability functions, such that

$$p_i, q_i \geqq 0 \qquad \text{for all } i \text{ and } \sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i .$$

Matusita [5] has defined a distance function

$$\left\{ \sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2 \right\}^{1/2}$$

for measuring the divergence between $(p_1, p_2, \cdots, p_n)$ and $(q_1, q_2, \cdots, q_n)$.

The quantities $(\sqrt{p_1}, \sqrt{p_2}, \cdots, \sqrt{p_n})$ and $(\sqrt{q_1}, \sqrt{q_2}, \cdots, \sqrt{q_n})$ may be interpreted as two different points on an n-dimensional sphere of unit radius in the positive quadrant. With this in mind, we have obtained the distribution of the square of the distance between two points on the unit sphere in the positive quadrant and have denoted it by M, in honour of Prof. Kameo Matusita. For application of these distributions, one may refer to the survey articles of Moran ([6], [7]).

The probability density function (p.d.f.) of X will be denoted by $f(x)$ in this article.

2. Distribution of the distance function when points are on n-dimensional sphere of unit radius

We note that the surface area of the unit sphere in n-dimension is (Anderson [1], p. 176)

$$\frac{2\pi^{n/2}}{\Gamma(n/2)} . \tag{2.1}$$

Let $(Y_1, Y_2, \cdots, Y_n)$ be a point on the surface of unit sphere, then it can be seen that the uniform distribution on the unit sphere with respect to $Y_1, Y_2, \cdots, Y_{n-1}$ is given by

$$\frac{\Gamma(n/2)}{2\pi^{n/2}}(1-y_1^2-\cdots-y_{n-1}^2)^{-1/2} . \tag{2.2}$$

A point on the surface of the unit sphere will have polar coordinates

$$\begin{aligned} y_1 &= \sin\theta_1 \\ y_2 &= \cos\theta_1 \sin\theta_2 \\ &\vdots \\ y_{n-1} &= \cos\theta_1 \cos\theta_2 \cdots \cos\theta_{n-2} \sin\theta_{n-1} \\ y_n &= \cos\theta_1 \cos\theta_2 \cdots \cos\theta_{n-2} \cos\theta_{n-1} . \end{aligned} \tag{2.3}$$

From the definition of the transformation (2.3), we have

$$(1-y_1^2-\cdots-y_{n-1}^2)^{-1/2} = y_n^{-1} = (\cos\theta_1 \cos\theta_2 \cdots \cos\theta_{n-2} \cos\theta_{n-1})^{-1} . \tag{2.4}$$

The Jacobian is given by

$$\frac{\partial(y_1, \cdots, y_{n-1})}{\partial(\theta_1, \cdots, \theta_{n-1})} = \cos^{n-1}\theta_1 \cos^{n-2}\theta_2 \cdots \cos^2\theta_{n-2} \cos\theta_{n-1} . \tag{2.5}$$

Thus in view of equations (2.2) to (2.5), we have

$$f(\theta_1, \theta_2, \cdots, \theta_{n-1}) = \frac{\Gamma(n/2)}{2\pi^{n/2}} \cos^{n-2}\theta_1 \cos^{n-1}\theta_2 \cdots \cos\theta_{n-2} . \tag{2.6}$$

Result (2.6) could have also been obtained as:

Let $\boldsymbol{X} = \text{Col}(X_1, X_2, \cdots, X_n)$ has normal distribution with mean vector $\boldsymbol{\mu} = \text{col}(\mu_1, \mu_2, \cdots, \mu_n)$ and covariance matrix Σ, to be denoted as $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$. Now, if we put $\boldsymbol{\mu} = \boldsymbol{0}$ and $\Sigma = I$ and consider a point on the n-dimensional sphere of radius r, by making a transformation as given in (2.3), result (2.6) will be obtained. This in turn also implies that $\boldsymbol{Y} = \boldsymbol{X}/S$, is uniformly distributed over the unit n-dimensional sphere, where $\boldsymbol{X} \sim N(\boldsymbol{0}, I)$ and $S^2 = \|X\|^2 = (X_1^2 + \cdots + X_n^2)$ (Kariya and Eaton [4]) and (Yaqub and Khan [9]).

Now if the coordinates of the other point on this sphere be $(1, 0, \cdots, 0)$, then the square of the distance between these two points will be

$$M^* = 2[1-\sin\theta_1] .$$

Thus, making the transformation from θ_1 to M^* and integrating (2.6) w.r.t. $\theta_2, \theta_3, \cdots, \theta_{n-1}$, we get

$$(2.7)\qquad f(M^*)=\frac{1}{2^{n-2}B\left(\frac{1}{2}, \frac{n-1}{2}\right)}[M^*(4-M^*)]^{(n-3)/2}, \qquad 0\leqq M^*\leqq 4 .$$

In the case of circle, if we put $M^*=4V$, it reduces to the arcsine density (Rao [8], p. 165).

An alternative approach

Let the two points on the unit sphere be $\boldsymbol{Y}=\mathrm{col}(Y_1, Y_2, \cdots, Y_n)$ and $\boldsymbol{a}=\mathrm{col}(a_1, a_2, \cdots, a_n)$, where $\boldsymbol{Y}=\boldsymbol{X}/S$. The square of the distance between these two points will be

$$(2.8)\qquad M^*=2(1-W), \qquad \text{where } W=\boldsymbol{a}'\boldsymbol{X}/S .$$

Kariya and Eaton [4] has shown that if $\|\boldsymbol{a}\|=(a_1^2+\cdots+a_n^2)^{1/2}=1$ and $\boldsymbol{X}$ has $N(\boldsymbol{0}, I)$, then $t=\sqrt{n-1}\,(W/\sqrt{1-W^2})$ has Students' t distribution with $(n-1)$ d.f.

Thus making the transformation

$$(2.9)\qquad t=\sqrt{n-1}\,\frac{2-M^*}{[M^*(4-M^*)]^{1/2}}$$

in the p.d.f. of t, the result (2.7) is obtained.

It may further be noted that if $\boldsymbol{a}=\mathrm{col}(1, 0, \cdots, 0)$, then

$$W=X_1/S=Y_1$$

where W^2 has $B(1/2, (n-1)/2)$ (Rao [8], p. 165) in this way also the p.d.f. of M^* can be obtained.

3. Distribution of the distance function, on the unit sphere in the positive quadrant

The surface area in the positive quadrant of the n-dimensional sphere of unit radius is one 2^nth part of the whole surface area (Balakrishnan et al. [2]). Thus, in the positive quadrant

$$(3.1)\qquad f(\theta_1, \theta_2, \cdots, \theta_{n-1})=\frac{2^{n-1}\Gamma(n/2)}{\pi^{n/2}}\cos^{n-2}\theta_1\cos^{n-3}\theta_2\cdots\cos\theta_{n-2}$$

$$0\leqq\theta_i\leqq\pi/2,\ i=1, 2, \cdots, n-1 .$$

Now, the square of the distance from $(1, 0, \cdots, 0)$ is $M=2(1-\sin\theta_1)$. Here again making the transformation from θ_1 to M in (3.1) and then integrating it w.r.t. $\theta_2, \theta_3, \cdots, \theta_{n-1}$, we get

$$(3.2) \qquad f(M)=\frac{1}{2^{n-3}B\left(\frac{1}{2}, \frac{n-1}{2}\right)}[M(4-M)]^{(n-3)/2}\,, \qquad 0\leqq M\leqq 2\,.$$

In this case also the alternative approach of Section 2 can be utilized.

If the points considered lie in the positive quadrant only, then Y_i and a_i are non-negative and consequently W hence t is non-negative. But since the t-distribution is symmetric about $t=0$, with $(n-1)$ d.f., the p.d.f. for t non-negative is

$$(3.3) \qquad f(t)=\frac{2}{\sqrt{n-1}\,B\left(\frac{1}{2}, \frac{n-1}{2}\right)}\left(1+\frac{t^2}{n-1}\right)^{-n/2}, \qquad 0<t<\infty$$

or when $\boldsymbol{a}=\mathrm{col}\,(1, 0, \cdots, 0)$

$$(3.4) \qquad f(w)=f(y_1)=\frac{2}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)}(1-y_1^2)^{(n-3)/2}\,, \qquad 0\leqq y_1\leqq 1\,.$$

Thus on making the transformation $t=\sqrt{n-1}\,((2-M)/[M(4-M)]^{1/2})$ in (3.3) or $w=(1-M/2)$ in (3.4), we can get the p.d.f. of M, as obtained in (3.2).

4. Moments of the distribution

4.1. *Moment when the points lie on the whole sphere*

The rth moment in this case is

$$\mu_r'=\frac{1}{2^{n-2}B\left(\frac{1}{2}, \frac{n-1}{2}\right)}\int_0^4 M^{*r}[M^*(4-M^*)]^{(n-3)/2}dM^*\,.$$

Put $M^*=4u$, then

$$\mu_r'=\frac{2^{2r+n-2}}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)}\int_0^1 u^{(2r+n-1)/2-1}(1-u)^{(n-1)/2-1}du$$

$$=\frac{2^{2r+n-2}}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)}\cdot B\left(\frac{2r+n-1}{2}, \frac{n-1}{2}\right)\,.$$

But since $B(1/2, (n-1)/2)=2^{n-2}B((n-1)/2, (n-1)/2)$,

$$\mu_r'=4^r\,\frac{\Gamma((2r+n-1)/2)\Gamma(n-1)}{\Gamma(r+n-1)\Gamma((n-1)/2)}\,.$$

In particular,

$$\mu_1'=2\ , \qquad \mu_2'=4\frac{n+1}{n}\ .$$

Thus,

$$\text{Mean}=2\ , \qquad \text{Variance}=\frac{4}{n}\ .$$

4.2. *Moment when points lie on the surface in positive quadrant*

The rth moment in the case of Section 3 is given by

$$\mu_r'=\frac{1}{2^{n-3}B\left(\frac{1}{2},\frac{n-1}{2}\right)}\int_0^2 M^r[M(4-M)]^{(n-3)/2}dM\ .$$

Putting $M=4\sin^2\alpha/2$, it reduces to

$$\begin{aligned}\mu_r'&=\frac{2^{r+1}}{B\left(\frac{1}{2},\frac{n-1}{2}\right)}\int_0^{\pi/2}(1-\cos\alpha)^r\sin^{n-2}\alpha d\alpha\\ &=\frac{2^r}{B\left(\frac{1}{2},\frac{n-1}{2}\right)}\sum_{i=0}^{r}\left\{(-1)^i\binom{r}{i}2\int_0^{\pi/2}\cos^i\alpha\sin^{n-2}\alpha d\alpha\right\}\\ &=\frac{2^r}{B\left(\frac{1}{2},\frac{n-1}{2}\right)}\sum_{i=0}^{r}(-1)^i\binom{r}{i}B\left(\frac{i+1}{2},\frac{n-1}{2}\right)\ .\end{aligned}$$

In particular,

$$\mu_1'=2\left[1-\frac{\Gamma(n/2)}{\sqrt{\pi}\,\Gamma((n+1)/2)}\right]$$

$$\mu_2'=4\left[1-\frac{2\Gamma(n/2)}{\sqrt{\pi}\,\Gamma((n+1)/2)}+\frac{1}{n}\right]\ .$$

5. Test statistic

1. In Section 2, we have shown that the p.d.f. of M^* is

$$f(M^*)=\frac{1}{2^{n-2}B\left(\frac{1}{2},\frac{n-1}{2}\right)}[M^*(4-M^*)]^{(n-3)/2}\ , \qquad 0\leqq M^*\leqq 4\ .$$

In the case of 3-dimensional sphere if we put $M^*=4v$, it reduces to

$$f(v)=1\ , \qquad 0\leqq v\leqq 1\ .$$

It is now easy to show that $z=-2\log_e v$ has the χ^2 distribution with 2 d.f.

Thus, if there are s chords on the unit sphere of 3-dimensions, then $\sum_{i=1}^{s} z_i$ will be distributed as χ^2 with $2s$ d.f.

Such type of statistics are useful in Genetics (Heuch [3], p. 686, Balakrishnan and Sanghvi [2], p. 860).

2. As is shown in Section 2, under the alternative approach

$$t=\sqrt{n-1}\frac{W}{\sqrt{1-W^2}}=\sqrt{n-1}\frac{2-M^*}{[M^*(4-M^*)]^{1/2}}$$

has Students' t-distribution with $(n-1)$ d.f.

The Statistic $M^*/4$ has beta distribution (see (2.7)) therefore table of incomplete beta function may be used in testing the hypotheses of interest regarding M^*. Also, because of the relationship between t and M^*, t-table may also be used.

3. Kariya and Eaton ([4], p. 211) has considered a testing problem and has shown that

$$\phi(x)=\begin{cases} 1 & \text{if } \dfrac{\boldsymbol{a}'\boldsymbol{X}}{\|\boldsymbol{X}\|}>k \\ 0 & \text{if } \dfrac{\boldsymbol{a}'\boldsymbol{X}}{\|\boldsymbol{X}\|}<k \end{cases}$$

is UMP test of its size. Thus the distribution of $W=\boldsymbol{a}'\boldsymbol{X}/\|\boldsymbol{X}\|$ may be used to obtain the value of k.

Acknowledgement

The authors are very much grateful to Prof. K. Matusita and Mr. S. Aki of the Institute of Statistical Mathematics and the referee whose comments led to the improvement of this paper.

ALIGARH MUSLIM UNIVERSITY

REFERENCES

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons.

[2] Balakrishnan, V. and Sanghvi, L. D. (1968). Distance between populations on the basis of attribute data, *Biometrics*, **24**, 859-865.

[3] Heuch, I. (1975). The relationship between separation time and genetic distance based on angular transformations of gene frequencies, *Biometrics*, **31**, 685-700.

[4] Kariya, T. and Eaton, M. L. (1977). Robust tests for spherical symmetry, *Ann. Statist.*, **5**, 206-215.

[5] Matusita, K. (1955). Decision rules based on distance for problems of fit, two samples and estimation, *Ann. Math. Statist.*, **26**, 631-641.
[6] Moran, P. A. P. (1966). A note on recent research in geometrical probability, *J. Appl. Prob.*, **3**, 453-463.
[7] Moran, P. A. P. (1969). A second note on recent research in geometrical probability, *Adv. Appl. Prob.*, **1**, 73-89.
[8] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, Second Edition, John Wiley and Sons.
[9] Yaqub, M. and Khan, A. H. (1979). The offset multivariate normal distribution (to be published).