PROBABILITY DISTRIBUTIONS AND ORTHOGONAL PARAMETERS

By V. S. HUZURBAZAR

Communicated by H. Jeffreys

Received 7 March 1949; *and in revised form* 4 July 1949

1. *Introduction.* Let $f(x, \alpha_i)$ be the probability density function of a distribution depending on n parameters $\alpha_i (i = 1, 2, \ldots, n)$. Then following Jeffreys (1) we shall say that the parameters α_i are orthogonal if

$$E\left(\frac{\partial^2}{\partial\alpha_i \partial\alpha_j} \log f\right) \equiv 0 \text{ for all } i, j \quad (i \neq j).$$

In general, we have a choice in choosing the form of the parameters of a distribution. If the β_j are any n independent functions of the α_i's, we may take the β_j as parameters in place of the α_i's. In a problem of estimation it is often convenient to estimate some suitable functions of the parameters rather than the parameters themselves. The importance of orthogonal parameters, as pointed out by Jeffreys, lies in the fact that the maximum likelihood equations to determine them are nearly orthogonal, and hence a practical solution is obtained with great ease by the method of iteration. This is particularly important in view of the fact that often the maximum likelihood equations do not admit explicit solutions and an approximation is necessary.

In this paper the problem of finding orthogonal parameters for a probability distribution is considered. It is pointed out that an arbitrary distribution with more than three parameters does not, in general, admit orthogonal parameters. Even the general case of three parameters does not look very hopeful owing to the complexity of the differential equations. For a distribution with two parameters, however, the problem is usually solvable. It is shown that parameters of location and scale can be made orthogonal by keeping either of them fixed and by a *linear* transformation of the other.

2. *Equations for finding orthogonal parameters.* At the outset we observe that the problem of finding orthogonal parameters is equivalent to finding n functions satisfying $\frac{1}{2}n(n-1)$ differential equations. It may be expected that there will be no solution in general for $n > 3$ unless the equations are mutually consistent, which may be so only for some distributions of a special character. This point appears to have been overlooked by Jeffreys. However, it is important to know what can be done for distributions with two or three parameters in general, and for some special distributions with more than three parameters. Jeffreys (1, 2) has obtained orthogonal parameters for Pearson's Types II and VII (which depend on three parameters), and for Type I (four parameters) when the asymmetry is small.

If the parameters α_i in $f(x, \alpha_i)$ are transformed into the β_j we find by actual differentiation

$$\frac{\partial^2}{\partial\beta_k \partial\beta_l} \log f = \frac{\partial\alpha_i}{\partial\beta_k} \frac{\partial\alpha_j}{\partial\beta_l} \frac{\partial^2}{\partial\alpha_i \partial\alpha_j} \log f + \frac{\partial^2\alpha_i}{\partial\beta_k \partial\beta_l} \frac{\partial}{\partial\alpha_i} \log f,$$

https://doi.org/10.1017/S0305004100025743 Published online by Cambridge University Press

where the summation convention is used with respect to i and j. Write

$$E_{ij} = E\left(-\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}\log f\right), \quad E'_{kl} = E\left(-\frac{\partial^2}{\partial\beta_k\partial\beta_l}\log f\right).$$

Since $E\left(\frac{\partial}{\partial\alpha_i}\log f\right) \equiv 0$, we have

$$E'_{kl} = E_{ij}\frac{\partial\alpha_i}{\partial\beta_k}\frac{\partial\alpha_j}{\partial\beta_l}. \tag{1}$$

If the β_j are orthogonal parameters we must have

$$E_{ij}\frac{\partial\alpha_i}{\partial\beta_k}\frac{\partial\alpha_j}{\partial\beta_l} = 0 \quad (k \neq l). \tag{2}$$

Thus we have $\frac{1}{2}n(n-1)$ simultaneous partial differential equations of the first order with the α_i as dependent variables. The equations are non-linear since the E_{ij} are functions of the α_i's. As already observed there will be no solution in general when $n > 3$. Comparatively little is known of simultaneous non-linear partial differential equations with more than one dependent variable. Hence even in the case $n = 3$ not many problems may admit consistent solutions.

3. *Geometrical significance.* The problem of finding orthogonal parameters is equivalent to reducing the quadratic differential form $J = E_{ij}d\alpha_i d\alpha_j$ to the diagonal form $E'_{ii}d\beta_i^2$ by a transformation of the α_i's. The differential form J, being positive definite and invariant (1) for all parametric transformations, satisfies the requisites of the square of an element of distance in Riemannian geometry. J may be regarded as providing a metric of the parametric space generated by the aggregate of all points $(\alpha_1, \alpha_2, \ldots, \alpha_n)$. When the metric of a Riemannian n-space is of the form $g_{ii}dx_i^2$ the co-ordinate hyper-surfaces form an n-tuply orthogonal system. A Riemannian n-space with an arbitrary metric, however, does not admit an n-tuply orthogonal system of hyper-surfaces when $n > 3$ (cf. Eisenhart (3)).

4. *Distributions depending on two parameters.* Suppose that we take $\alpha_1 = \beta_1$ and try to adjust $\alpha_2 = \alpha_2(\beta_1, \beta_2)$ so that β_1 and β_2 are orthogonal. We have to solve

$$E'_{12} = E_{12}\frac{\partial\alpha_2}{\partial\beta_2} + E_{22}\frac{\partial\alpha_2}{\partial\beta_1}\frac{\partial\alpha_2}{\partial\beta_2} = 0.$$

Since α_2 must necessarily depend on β_2 (for otherwise the parameters α_1 and α_2 would not be independent) we have

$$E_{12} + E_{22}\frac{\partial\alpha_2}{\partial\beta_1} = 0. \tag{3}$$

Now $$E_{12} = E_{12}(\beta_1, \alpha_2), \quad E_{22} = E_{22}(\beta_1, \alpha_2).$$

Since β_2 does not appear directly in the equation (3), the equation is really equivalent to an ordinary differential equation of the first order and first degree. It will be, however, non-linear in general. For most of the well-known distributions with two parameters the variables in the equation can be separated and the equation is solvable.

As an example, consider the negative binomial distribution

$$f(x, \alpha_1, \alpha_2) = (1-\alpha_2)^{\alpha_1+1}\alpha_2^x\frac{(\alpha_1+x)!}{\alpha_1!\,x!}.$$

https://doi.org/10.1017/S0305004100025743 Published online by Cambridge University Press

Here $$E_{12} = \frac{1}{1-\alpha_2}, \quad E_{22} = \frac{\alpha_1+1}{\alpha_2(1-\alpha_2)^2}.$$

With $\alpha_1 = \beta_1$, the equation for α_2 is

$$\frac{1}{\beta_1+1} + \frac{1}{\alpha_2(1-\alpha_2)}\frac{\partial\alpha_2}{\partial\beta_1} = 0,$$

giving $$\frac{\alpha_2}{1-\alpha_2}(\beta_1+1) = \psi(\beta_2),$$

where ψ is an arbitrary function. Take $\psi(\beta_2) = \beta_2$ for simplicity. β_2 is the mean of the distribution. We then have

$$\beta_1 = \alpha_1, \quad \beta_2 = \frac{\alpha_2}{1-\alpha_2}(\alpha_1+1).$$

4.1. *Parameters of location and scale.* An interesting case occurs when the two parameters are of location and scale. Let

$$f(x, \alpha_1, \alpha_2) = \frac{1}{\alpha_2}\exp\left\{\phi\left(\frac{x-\alpha_1}{\alpha_2}\right)\right\},$$

so that α_1 and α_2 are respectively the location and scale parameters. Write $\xi = \frac{x-\alpha_1}{\alpha_2}$. Then

$$E_{11} = \frac{k_{11}}{\alpha_2^2}, \quad E_{12} = \frac{k_{12}}{\alpha_2^2}, \quad E_{22} = \frac{k_{22}}{\alpha_2^2},$$

where $k_{11} = -E\{\phi''(\xi)\}$, $k_{12} = -E\{\xi\phi''(\xi)\}$, $k_{22} = -E\{1+2\xi\phi'(\xi)+\xi^2\phi''(\xi)\}$.

The k's are independent of α_1 and α_2. With $\alpha_2 = \beta_2$, the equation for α_1 is

$$\frac{k_{12}}{\alpha_2^2} + \frac{k_{11}}{\alpha_2^2}\frac{\partial\alpha_1}{\partial\beta_2} = 0 \quad \text{or} \quad k_{12} + k_{11}\frac{\partial\alpha_1}{\partial\beta_2} = 0,$$

giving $$\alpha_1 = \beta_1 + k\beta_2, \quad \text{where} \quad k = -\frac{k_{12}}{k_{11}}.$$

If, instead, we keep α_1 fixed and adjust α_2, we have

$$\alpha_2 = k'\beta_1 + \beta_2 \quad \left(k' = -\frac{k_{12}}{k_{22}}\right).$$

The parameters of location and scale can therefore be made orthogonal by keeping either of them fixed and by a *linear* transformation of the other.

So far we have considered the location and scale parameters together. Now suppose that only one of the two parameters is of location or scale and the other is a general one. Then

$$f(x, \alpha_1, \alpha_2) = \exp\{\phi(x-\alpha_1, \alpha_2)\} \quad \text{or} \quad \frac{1}{\alpha_1}\exp\left\{\phi\left(\frac{x}{\alpha_1}, \alpha_2\right)\right\},$$

according as α_1 is a location or a scale parameter. It can be shown by actual differentiation that each E_{ij} is of the form $A_{ij}(\alpha_1)\,B_{ij}(\alpha_2)$, the A_{ij} and B_{ij} being respectively functions of α_1 and α_2 only. If we keep either of the parameters fixed, the equation to determine the other is solvable since the variable can be separated. The problem of orthogonality is therefore always solvable when one of the parameters is of location or scale.

5. *Distributions with three parameters.* We discuss below a case where we can find orthogonal parameters.

https://doi.org/10.1017/S0305004100025743 Published online by Cambridge University Press

If $E_{12} = E_{13} = 0$ and E_{23}/E_{33} is independent of α_1, then the parameters can be made orthogonal by taking $\alpha_1 = \beta_1$, $\alpha_2 = \beta_2$, and adjusting $\alpha_3 = \alpha_3(\beta_1, \beta_2, \beta_3)$. For to do this α_3 must satisfy the three equations

$$E_{32}\frac{\partial \alpha_3}{\partial \beta_1} + E_{33}\frac{\partial \alpha_3}{\partial \beta_1}\frac{\partial \alpha_3}{\partial \beta_2} = 0, \tag{4}$$

$$\frac{\partial \alpha_3}{\partial \beta_1} = 0, \tag{5}$$

and

$$E_{23} + E_{33}\frac{\partial \alpha_3}{\partial \beta_2} = 0. \tag{6}$$

Since (5) implies (4), we have to solve only (5) and (6). From (5), α_3 is a function of β_2 and β_3 only. From (6),

$$\frac{\partial \alpha_3}{\partial \beta_2} = -\frac{E_{23}}{E_{33}} = \phi(\alpha_2, \alpha_3), \quad \text{by hypothesis.}$$

Hence

$$\frac{\partial \alpha_3}{\partial \beta_2} = \phi(\beta_2, \alpha_3). \tag{7}$$

The parameter β_1 does not appear in (7). If, then, in the solution of (7) we take care not to introduce any function of β_1 as an arbitrary constant, α_3 will not contain β_1 and then (5) will also be satisfied. Thus we have to take α_3 as a function of β_2 and β_3 only, and need solve a single equation (7). The equation (7) will be often solvable, being really equivalent to an ordinary differential equation of the first order and first degree.

Pearson's Types II and VII are illustrations of the case in point. As a matter of fact it was Jeffreys's derivation of orthogonal parameters for these types that suggested the case.

5.1. *Pearson's Type III distribution.* Here

$$f(x, \alpha_1, \alpha_2, \alpha_3) = \frac{1}{\alpha_3 ! \alpha_2} e^{-(x-\alpha_1)/\alpha_2}\left(\frac{x-\alpha_1}{\alpha_2}\right)^{\alpha_3} \quad (\alpha_1 \leqslant x < \infty).$$

Consideration of dimensions suggests the following transformation

$$\alpha_1 = \beta_1 + \beta_2 u(\beta_2), \quad \alpha_2 = \beta_2 v(\beta_3), \quad \alpha_3 = w(\beta_3), \tag{8}$$

where u, v, w are functions of β_3 only, and are to be adjusted so that the β's are orthogonal. It is found, however, that the equations are inconsistent. If, however, α_3 is so large that we can neglect α_3^{-2}, the equations become consistent. One solution can be obtained in the form

$$u = -(\beta_3 - 1)\beta_3^{-\frac{1}{2}}, \quad v = \beta_3^{-\frac{1}{2}}, \quad w = \beta_3,$$

leading to the transformation

$$\alpha_1 = \beta_1 - (\beta_3 - 1)\beta_3^{-\frac{1}{2}}\beta_2, \quad \alpha_2 = \beta_3^{-\frac{1}{2}}\beta_2, \quad \alpha_3 = \beta_3. \tag{9}$$

In conclusion I wish to thank Prof. Jeffreys for his encouragement and criticism.

REFERENCES

(1) JEFFREYS, H. *Theory of probability*, 2nd ed. (Oxford, 1948), pp. 158, 184.
(2) JEFFREYS, H. *Phil. Trans.* A, 237 (1938), 231–71.
(3) EISENHART, L. P. *Riemannian geometry* (Princeton 1926), p. 43.

FITZWILLIAM HOUSE
CAMBRIDGE

https://doi.org/10.1017/S0305004100025743 Published online by Cambridge University Press