

ON DISTRIBUTIONS ADMITTING A SUFFICIENT STATISTIC*

BY
B. O. KOOPMAN

Let x be a variate taking on values, determined by chance, on the axis of reals, R ; and let the frequency function $f(\theta_1, \dots, \theta_\nu, x)$ be known, but involve ν parameters $(\theta_1, \dots, \theta_\nu)$ which are unknown, but are confined to a known region Ω of real ν -space. That is, we are assuming that if ν values of $(\theta_1, \dots, \theta_\nu)$ are given,

$$(1) \quad f(\theta_1, \dots, \theta_\nu, x) = \lim_{\Delta x \rightarrow 0+} \frac{\text{prob. } [x \leq x < x + \Delta x]}{\Delta x},$$

$$\int_R f(\theta_1, \dots, \theta_\nu, x) dx = 1.$$

A problem of practical importance is the statistical estimation of the parameters $(\theta_1, \dots, \theta_\nu)$: Let the result of n independent observations of x made on the assumption that $(\theta_1, \dots, \theta_\nu)$ is fixed yield the numbers x_1, \dots, x_n ,—the “sample”; assuming nothing known a priori concerning the position of $(\theta_1, \dots, \theta_\nu)$ in Ω , so that Bayes’ formula is inapplicable, how can the sample x_1, \dots, x_n be used to secure information regarding $(\theta_1, \dots, \theta_\nu)$?

According to R. A. Fisher† this problem is to be solved by finding ν functions of n arguments, $\phi_j(x_1, \dots, x_n)$ ($j = 1, \dots, \nu$), such that when the arguments are replaced by the values in the sample, the resulting values are the appropriate values of θ_j . The question of how appropriateness is to be determined has its roots deep in the foundations of the subject, and will not be considered here. The function ϕ_j is called by R. A. Fisher an estimate of θ_j , or a *statistic*.

Assuming that the statistics ϕ_j ($j = 1, \dots, \nu$) exist, a question arises immediately which, stated intuitively, runs as follows: does the position of the single point $(\phi_1(x_1, \dots, x_n), \dots, \phi_\nu(x_1, \dots, x_n))$ in Ω “contain all the information” relative to the position of $(\theta_1, \dots, \theta_\nu)$ “contained in the sample” of n numbers (x_1, \dots, x_n) when (as is usual) $n > \nu$? Or is “relevant information” lost when x_1, \dots, x_n are discarded and only the numbers

* Presented to the Society, April 20, 1935; received by the editors March 16, 1935.

† *On the mathematical foundations of theoretical statistics*, Philosophical Transactions, Royal Society of London, (A), vol. 222, pp. 309–368.

$\phi_1(x_1, \dots, x_n), \dots, \phi_\nu(x_1, \dots, x_n)$ retained? In the former case, R. A. Fisher describes the set of functions ϕ_j as a set of *sufficient statistics*. The term "relevant information" is used by Fisher in two senses which, to our way of thinking, have never been shown to be fully equivalent: in the intuitive sense (suggested by the word but never defined) and in the sense of a certain definite integral.*

The first object of the present paper is to give a simple definition of the *sufficiency* of a statistic which, we feel, expresses the intuitive notion of R. A. Fisher and is mathematically equivalent to certain of his formulations of this concept, at least under suitable mathematical restrictions.† The second object is to prove that the only distributions (of the analytic nature met with in practice, at least) which possess a sufficient statistic are of the very special exponential type of formula (4) below.

DEFINITION. *The distribution $f(\theta_1, \dots, \theta_\nu, x)$ shall be said to admit the system of statistics $\phi_j(x_1, \dots, x_n)$ ($j=1, \dots, \nu$) as sufficient statistics if the equations*

$$(2) \quad \phi_j(x_1, \dots, x_n) = \phi_j(x'_1, \dots, x'_n) \quad (j = 1, \dots, \nu)$$

imply the following identity in $(\theta_1, \dots, \theta_\nu)$, $(\theta'_1, \dots, \theta'_\nu)$ on Ω :

$$(3) \quad \frac{\prod_{i=1}^n f(\theta_1, \dots, \theta_\nu, x_i)}{\prod_{i=1}^n f(\theta_1, \dots, \theta_\nu, x'_i)} = \frac{\prod_{i=1}^n f(\theta'_1, \dots, \theta'_\nu, x_i)}{\prod_{i=1}^n f(\theta'_1, \dots, \theta'_\nu, x'_i)},$$

this equation to be interpreted after formal multiplication wherever denominators are zero.

Here, of course, the letters θ, θ', x, x' , etc., denote variables in the sense of classical analysis.

This definition, like so many definitions of applied mathematics, seeks to throw into precise and explicit form an intuitive conception; and in the nature of things, its a priori justification is to be sought in an examination of its adequacy as a rendering of the intuition in question: in a sort of introspection. In the present case the intuitive starting point may be illustrated in the following example: Suppose that a "random trial" can lead, among others, to the mutually exclusive results A or A^1 of respective a priori proba-

* R. A. Fisher, *Theory of statistical estimation*, Proceedings of the Cambridge Philosophical Society, vol. 22, pp. 712-714.

† For a comparative study of these various definitions, cf. a paper by J. L. Doob, entitled *Statistical estimation*, in the present number of these Transactions, pp. 410-421.

bilities $p(\theta)$ and $p'(\theta)$ ($p(\theta) + p'(\theta) \leq 1$), dependent upon the unknown parameter θ , itself having no a priori distribution, as above. Suppose first that $p(\theta)/p'(\theta)$ is independent of θ : then we should feel that the datum: "A happened" tells us nothing about the value of θ which is not contained in the datum: "A or A^1 happened." And, secondly, if, e.g., $p(\theta_1)/p'(\theta_1) < p(\theta_2)/p'(\theta_2)$, the datum "A happened" will, as compared with the less complete datum "A or A^1 happened," make us feel that θ_2 is a better guess for the value of θ than θ_1 . This is extended in obvious fashion to the case of ν parameters $(\theta) = (\theta_1, \dots, \theta_\nu)$.

The present application is obvious: Select arbitrarily the points (x_1, \dots, x_n) and (x'_1, \dots, x'_n) in cartesian n -space and describe about them mutually exclusive regions $\Delta\tau$ and $\Delta\tau'$, respectively. To quantities of higher order in general, when the regions are small,

$$p(\theta) = \Delta\tau \prod_{i=1}^n f(\theta_1, \dots, \theta_\nu, x_i),$$

$$p'(\theta) = \Delta\tau' \prod_{i=1}^n f(\theta_1, \dots, \theta_\nu, x'_i)$$

are the a priori probabilities that n contemplated trials give a sample in $\Delta\tau$ or in $\Delta\tau'$, respectively. Now $p(\theta)/p'(\theta)$ is, apart from the constant factor $\Delta\tau/\Delta\tau'$, equal to the left-hand member of equation (3). Thus, (3) expresses the fact that no information is lost, in the above intuitive sense, when the datum "the sample was in $\Delta\tau$ (approximately, $= (x_1, \dots, x_n)$)" is replaced by the datum "the sample was in either $\Delta\tau$ or $\Delta\tau'$ ". For a sufficient statistic, this must be the case wherever (x_1, \dots, x_n) and (x'_1, \dots, x'_n) are connected by the equation (2).

In this interpretation we have left out of account values for which $f(\theta_1, \dots, \theta_\nu, x) = 0$, or for which the above expressions for the probability of a point in $\Delta\tau, \Delta\tau'$ are not applicable: these are exceptional in the sense that there is, at least for the corresponding values of $(\theta_1, \dots, \theta_\nu)$, a set containing them all, and such that the probability of having a point in this set is zero.

It is necessary to emphasize a point here: For example when $\nu = 2$, in saying that (θ_1, θ_2) are unknown parameters, it is simply meant that the point (θ_1, θ_2) has an unknown position in Ω , with no a priori distribution. But it is perfectly conceivable that when a value of one of the parameters, θ_1 , is given, the other, θ_2 , may be by no means unknown in this complete sense: it may either be determined, or have a known a priori distribution, etc., i.e., θ_1 and θ_2 may be statistically dependent. It is shown at once with the aid of equations (2) and (3) used in various combinations, that when θ_1 and θ_2 are

statistically independent, a sufficient condition for the sufficiency of the statistics ϕ_1, ϕ_2 is that (i) when θ_1 is given, ϕ_2 be a sufficient statistic for the unknown θ_2 , and (ii) when θ_2 is given, ϕ_1 be a sufficient statistic for θ_1 . But it is perfectly conceivable that $f(\theta_1, \theta_2, x)$ admit the pair of sufficient statistics ϕ_1, ϕ_2 without admitting a sufficient statistic at all for θ_2 when the value of θ_1 is given, etc. Finally, $f(\theta_1, \theta_2, x)$ may admit only a single sufficient statistic, for θ_1 , say, dependent or not on θ_2 : this means that the equation $\phi_1(\theta_2; x_1, \dots, x_n) = \phi_1(\theta_2; x'_1, \dots, x'_n)$ implies (3) with $\theta'_2 = \theta_2$ ($\nu = 2$).

In the following we shall use the notation $A \times B$ for the product set of two sets A and B , consisting of all pairs of elements (a, b) where a is an element of A , b one of B . And we shall write $A^2 = A \times A$, the set of pairs (a, a') , a and a' both in A .

THEOREM I. *Let $f(\theta_1, \dots, \theta_\nu, x)$ be analytic and not zero at each point of a subset $\Omega \times R - T$ of $\Omega \times R$, and let $\phi_j(x_1, \dots, x_n)$ ($j = 1, \dots, \nu$) be continuous throughout R^n ; finally, suppose that $n > \nu$.*

Then a necessary condition that $(\phi_1, \dots, \phi_\nu)$ form a set of sufficient statistics for this distribution is that at each given point (a_1, \dots, a_ν, b) of $\Omega \times R - T$ a neighborhood $N_{ab} = \omega_{ab} \times r_{ab} \subset \Omega \times R - T$ exist, where

$$\begin{aligned} \omega_{ab}: & \quad |\theta_j - a_j| < h_j & (j = 1, \dots, \nu), \\ r_{ab}: & \quad |x - b| < h, \end{aligned}$$

such that

$$(4) \quad f(\theta_1, \dots, \theta_\nu, x) = \exp \left[\sum_{k=1}^{\mu} \Theta_k X_k + \Theta + X \right],$$

where Θ_k, Θ are real, single-valued, analytic functions of $(\theta_1, \dots, \theta_\nu)$ in ω_{ab} , and X_k, X are real, single-valued, analytic functions of x in r_{ab} , and where, finally, $\mu \leq \nu$ ($\mu = 0$ means that all the functions Θ_k, X_k are lacking).

Furthermore, if μ has the smallest value for which the identity (4) is valid, a circumstance which can always be brought about, it must follow that

$$\sum_{i=1}^n X_k(x_i) = V_k[\phi_1(x_1, \dots, x_n), \dots, \phi_\nu(x_1, \dots, x_n)] \quad (k = 1, \dots, \mu),$$

where V_k is a single-valued function of its ν arguments.

We shall give the proof in the case $\nu = 2$, which is sufficiently illustrative. We shall assume, further, that $n = 3$: For if $n > 3$ we have but to write $\phi_j(x_1, x_2, x_3) = \phi_j(x_1, x_2, x_3, b, \dots, b)$ and, in (3), to take $x_4 = \dots = x_n = x'_4 = \dots = x'_n = b$, to have (3) a consequence of (2) with $\nu = 2, n = 3$.

Since N_{ab} is in $\Omega \times R - T$, $f(\theta_1, \theta_2, x)$ is real, single-valued, analytic, and

$\neq 0$, throughout N_{ab} ; and hence (with the real determination of the log) the function

$$\log \frac{f(\theta_1, \theta_2, x)}{f(\theta'_1, \theta'_2, x)}$$

is real, single-valued, and analytic for all $(\theta_1, \theta_2, \theta'_1, \theta'_2, x)$ in the neighborhood $\omega_{ab}^2 \times r_{ab}$ of (a_1, a_2, a_1, a_2, b) . It follows that on setting successively $(\theta'_1, \theta'_2) = (a_{11}, a_{12}), (a_{21}, a_{22}), (a_{31}, a_{32})$ (all in ω_{ab}) in equation (3) and then taking logarithms, it may be made to yield the system

$$(5) \quad \sum_{i=1}^n \log \frac{f(\theta_1, \theta_2, x_i)}{f(a_{j1}, a_{j2}, x_i)} = \sum_{i=1}^n \log \frac{f(\theta_1, \theta_2, x'_i)}{f(a_{j1}, a_{j2}, x'_i)} \quad (j = 1, 2, 3),$$

and each member will be a real, single-valued, and analytic function for all points $(\theta_1, \theta_2, x_1, x_2, x_3)$ etc. on $\omega_{ab} \times r_{ab}^3$.

Since by hypothesis ϕ_1, ϕ_2 form a system of sufficient statistics, equations (5) are a consequence of

$$(6) \quad \begin{aligned} \phi_1(x_1, x_2, x_3) &= \phi_1(x'_1, x'_2, x'_3), \\ \phi_2(x_1, x_2, x_3) &= \phi_2(x'_1, x'_2, x'_3). \end{aligned}$$

Geometrically, this means that, given any (θ_1, θ_2) of ω_{ab} , the locus (6) on r_{ab}^3 through each given point (x'_1, x'_2, x'_3) of this region must be a subset of the locus (5) through this same point. Now this circumstance implies the identical vanishing of the jacobian of the left-hand members of (5) with respect to x_1, x_2, x_3 throughout $\omega_{ab} \times r_{ab}^3$: For suppose that a point $(\theta_1^0, \theta_2^0, x_1^0, x_2^0, x_3^0)$ of this neighborhood existed at which the jacobian failed to vanish; a closed cubical neighborhood S_0 of (x_1^0, x_2^0, x_3^0) would exist, lying wholly within r_{ab}^3 , at no point of which the jacobian vanishes. It would then follow by the Implicit Function Theorem that equations (5), when $\theta_j = \theta_j^0, (x_1, x_2, x_3)$ and (x'_1, x'_2, x'_3) are on S_0 , could only be satisfied when $(x_1, x_2, x_3) = (x'_1, x'_2, x'_3)$. Hence, (6) would imply $(x_1, x_2, x_3) = (x'_1, x'_2, x'_3)$, provided we remain confined to S_0 . Hence the equations

$$u_1 = \phi_1(x_1, x_2, x_3), \quad u_2 = \phi_2(x_1, x_2, x_3)$$

define a one-one correspondence between S_0 and U_0 , where U_0 is the range of the point (u_1, u_2) defined by these equations as (x_1, x_2, x_3) traces out S_0 : by the continuity of ϕ_i and the nature of S_0 , U_0 will be bounded and closed in the u_1u_2 -plane. But it follows at once under these circumstances that the above correspondence is continuous both ways, and thus contradicts the preservation of dimensionality under homeomorphism.

Consider the matrix of the derivatives of the left-hand members of (5) with respect to x_1, x_2, x_3 . It is found at once to have the form

$$\mathbf{M}_3 = \| \| M_{ij} \|_{(i, j=1,2,3)} = \left\| \left\| \frac{\partial}{\partial x_i} \log \frac{f(\theta_1, \theta_2, x_i)}{f(a_{j1}, a_{j2}, x_i)} \right\|_{(i, j=1,2,3)} \right\|$$

Let us write, further, for the upper left-hand minors,

$$\mathbf{M}_2 = \| \| M_{ij} \|_{(i, j=1,2)} \quad \mathbf{M}_1 = \| M_{11} \|,$$

and let ρ be the identical rank of \mathbf{M}_3 (the order of the non-identically vanishing determinant of highest order in \mathbf{M}_3). We have seen that under the hypothesis of our theorem, we are confined to the possibilities $\rho=0, 1, 2$. It is obviously permissible to assume that it is $\det \mathbf{M}_\rho$ which does not vanish identically.

Case 1; $\rho=0$. The equation

$$(7) \quad \det \mathbf{M}_1 = \frac{\partial}{\partial x_1} \log \frac{f(\theta_1, \theta_2, x_1)}{f(a_{11}, a_{12}, x_1)} = 0$$

can be integrated with respect to x_1 , from b to x in r_{ab} , so that

$$\log \frac{f(\theta_1, \theta_2, x)}{f(a_{11}, a_{12}, x)} = \log \frac{f(\theta_1, \theta_2, b)}{f(a_{11}, a_{12}, b)},$$

and on solving this for $f(\theta_1, \theta_2, x)$, the latter is seen to be of the form (4) with $\mu=0$ and

$$\Theta(\theta_1, \theta_2) = \log \frac{f(\theta_1, \theta_2, b)}{f(a_{11}, a_{12}, b)},$$

$$X(x) = \log f(a_{11}, a_{12}, x),$$

and these functions are obviously of the required analytic nature.

Case 2; $\rho=1$. Here $\det \mathbf{M}_2 \equiv 0$, but $\det \mathbf{M}_1 \neq 0$. This continues to be true if the first column of \mathbf{M}_2 is subtracted from the second. After this has been done, the symbol $\partial/\partial x_1$ may be taken outside the determinant, and the resulting equation integrated with respect to x_1 from b to x in ω_{ab} , obtaining

$$\begin{aligned} & \left| \begin{array}{cc} \log \frac{f(\theta_1, \theta_2, x)}{f(a_{11}, a_{12}, x)} & \log \frac{f(a_{11}, a_{12}, x)}{f(a_{21}, a_{22}, x)} \\ \frac{\partial}{\partial x_2} \log \frac{f(\theta_1, \theta_2, x_2)}{f(a_{11}, a_{12}, x_2)} & \frac{\partial}{\partial x_2} \log \frac{f(a_{11}, a_{12}, x_2)}{f(a_{21}, a_{22}, x_2)} \end{array} \right| \\ &= \left| \begin{array}{cc} \log \frac{f(\theta_1, \theta_2, b)}{f(a_{11}, a_{12}, b)} & \log \frac{f(a_{11}, a_{12}, b)}{f(a_{21}, a_{22}, b)} \\ \frac{\partial}{\partial x_2} \log \frac{f(\theta_1, \theta_2, x_2)}{f(a_{11}, a_{12}, x_2)} & \frac{\partial}{\partial x_2} \log \frac{f(a_{11}, a_{12}, x_2)}{f(a_{21}, a_{22}, x_2)} \end{array} \right|. \end{aligned}$$

If, now, we can find a value for x_2 in r_{ab} such that

$$(8) \quad \frac{\partial}{\partial x_2} \log \frac{f(a_{11}, a_{12}, x_2)}{f(a_{21}, a_{22}, x_2)} \neq 0,$$

we can solve the above equation for $f(\theta_1, \theta_2, x)$ explicitly, and establish the form (4) for it, with $\mu=1$, and explicit expressions of the required analytic character for $\Theta_1(\theta_1, \theta_2)$, $\Theta(\theta_1, \theta_2)$, $X_1(x)$, $X(x)$. But to assert that no a_{21} , a_{22} , x exist for which (8) is true is to require that (7) be an identity, as is seen by multiplying (8) through by -1 , and then replacing a_{21} , a_{22} , x by θ_1 , θ_2 , x , respectively. But that is Case 1, which we are at present excluding.

Case 3; $\rho=2$. The proof is similar to the above. The first column of \mathbf{M}_3 is subtracted from the second and the third, and the symbol $\partial/\partial x_1$ taken out; then the determinant equation is integrated with respect to x_1 from b to x in r_{ab} , and the resulting equation solved for $f(\theta_1, \theta_2, x)$. The explicit form of the functions in (4) shows that they have the required analytic character. The only case where the equation might fail to be solvable for $f(\theta_1, \theta_2, x)$ is when a certain 2-rowed determinant vanishes identically; but an obvious transformation would show that this would imply that $\det \mathbf{M}_2 \equiv 0$, which would be Case 1 or 2, both at present excluded.

In order to prove the last paragraph of our theorem, we substitute expressions (4) (μ assumed to be minimal) into (3), take logarithms, etc., thus obtaining as the necessary consequence of (2)

$$\begin{aligned} & \sum_{k=1}^{\mu} \left\{ [\Theta_k(\theta_1, \dots, \theta_\nu) - \Theta_k(\theta'_1, \dots, \theta'_\nu)] \cdot \sum_{i=1}^n X_k(x_i) \right\} \\ &= \sum_{k=1}^{\mu} \left\{ [\Theta_k(\theta_1, \dots, \theta_\nu) - \Theta_k(\theta'_1, \dots, \theta'_\nu)] \sum_{i=1}^n X_k(x'_i) \right\}, \end{aligned}$$

which, being an identity in the θ , θ' 's, yields, on setting successively $(\theta'_1, \dots, \theta'_\nu) = (a_{j1}, \dots, a_{j\nu})$ ($j=1, \dots, \mu$), all on ω_{ab}^ν , μ equations, which establish the identity

$$\sum_{i=1}^n X_k(x_i) = \sum_{i=1}^n X_k(x'_i) \quad (k = 1, \dots, \mu)$$

as a necessary consequence of (2) (in other words, the remaining conclusion of our theorem), provided the determinant

$$\Delta = |\Delta_{kj}| = |\Theta_k(\theta_1, \dots, \theta_\nu) - \Theta_k(a_{j1}, \dots, a_{j\nu})| \quad (k, j = 1, \dots, \mu)$$

is not identically zero.

Suppose that $\Delta \equiv 0$, and let Δ' be the determinant obtained from Δ by

subtracting the first column from each of the others: $\Delta' \equiv 0$. Now the complement of the last element in the first column of Δ' is

$$|\Theta_k(a_{11}, \dots, a_{1\mu}) - \Theta_k(a_{11}, \dots, a_{1l})|$$

$$(k = 1, \dots, \mu - 1; l = 2, \dots, \mu);$$

if this is identically zero in the a 's, we should have

$$(9) \quad |\Delta_{kj}|_{(k, j=1, \dots, \mu-1)} \equiv 0.$$

Leaving this case to one side, suppose the a 's such that this complement is not equal to zero. Then the first column of Δ' is linearly dependent upon the other columns, and since these latter are independent of the θ 's, we have $\mu - 1$ functions T_s , whose explicit form shows them to be real, single-valued, and analytic on ω'_{ab} , and $\mu(\mu - 1)$ real constants C_{ks} , such that

$$\Theta_k(\theta_1, \dots, \theta_\nu) = \sum_{s=1}^{\mu-1} C_{ks} T_s(\theta_1, \dots, \theta_\nu) + \Theta_k(a_{11}, \dots, a_{1\mu}) \quad (k = 1, \dots, \mu).$$

On substituting these expressions into (4), it becomes evident that it can be written with a number less than μ of non-identically vanishing functions $\Theta_k(\theta_1, \dots, \theta_\nu) X_k(x)$, contrary to our assumption.

Finally, in the case that (9) is true, we replace Δ by the determinant in (9), and after a discussion precisely like the one above, obtain the same result, viz., that the value of μ can be reduced.

This completes the proof of Theorem I.

COROLLARY. *If in the hypothesis of Theorem I the requirement that ϕ_1, \dots, ϕ_ν be sufficient statistics is replaced by the requirement that, when values of $\theta_{\lambda+1}, \dots, \theta_\nu$ are given, $\theta_1, \dots, \theta_\lambda$ are wholly unknown and have the sufficient statistics $\dot{p}_j(\theta_{\lambda+1}, \dots, \theta_\nu; x_1, \dots, x_n)$ ($j=1, \dots, \lambda$), where $0 < \lambda < \nu$, $\lambda < n$, then the conclusion remains in force, except that $\mu \leq \lambda$, and that the functions X_k, X, V_k involve $\theta_{\lambda+1}, \dots, \theta_\nu$.*

The proof is obtained just as the proof of Theorem I, except that λ is used in place of ν , and $\theta_1, \dots, \theta_\lambda, \theta'_1, \dots, \theta'_\nu$ treated as were $\theta_1, \dots, \theta_\nu, \theta'_1, \dots, \theta'_\nu$, the residual variables $\theta_{\lambda+1}, \dots, \theta_\nu$ being carried, unaltered, in all the expressions.

THEOREM II. *A sufficient condition for the sufficiency of the system of statistics $\phi_j(x_1, \dots, x_n)$ ($j=1, \dots, \nu$) of the distribution $f(\theta_1, \dots, \theta_\nu, x)$ is that, if $R = R^* + R^{**}$ ($R^* R^{**} = 0$),*

$$f(\theta_1, \dots, \theta_\nu, x) = 0 \text{ for all } (\theta_1, \dots, \theta_\nu, x) \text{ on } \Omega \times R^{**},$$

$$(10) \quad f(\theta_1, \dots, \theta_\nu, x) = \exp \left[\sum_{k=1}^{\mu} \Theta_k X_k + \Theta + X \right]$$

$$\text{for all } (\theta_1, \dots, \theta_\nu, x) \text{ on } \Omega \times R^*;$$

$$(11) \quad \sum_{i=1}^n X_k(x_i) = V_k[\phi_1(x_1, \dots, x_n), \dots, \phi_\nu(x_1, \dots, x_n)]$$

($k = 1, \dots, \mu$)

for all $(\theta_1, \dots, \theta_\nu, x)$ on $\Omega \times R^*$;

where Θ_k, Θ are real single-valued functions of $(\theta_1, \dots, \theta_\nu)$ at all points of Ω , and X_k, X , real single-valued functions of x defined almost everywhere on R^* , such that, for each $(\theta_1, \dots, \theta_\nu)$ on Ω , equation (1) is valid in the sense of Lebesgue, and, finally, V_k is real and single-valued for all values of its ν arguments on the range of $(\phi_1, \dots, \phi_\nu)$ as (x_1, \dots, x_n) describes $(R^*)^n$.

It is shown immediately, upon substitution and reduction, that equation (2) implies (3), and this formal work is valid for each $(\theta_1, \dots, \theta_\nu)$, $(\theta'_1, \dots, \theta'_\nu)$ on Ω , and almost every (x_1, \dots, x_n) , (x'_1, \dots, x'_n) on $(R^*)^n$. This furnishes our proof.

It may be remarked that a sufficient condition constituting a partial converse of the corollary to Theorem I in the same sense that Theorem II is a partial converse to Theorem I is readily formulated. Furthermore, theorems of the above nature are easily obtained in the case where x denotes a point of N -space R^N : the multivariate distributions.

A final remark is that all our results have a form invariant under change of parameter, and also under change of variable x , as indeed they should have.

THEOREM III. *Let equations (10) of the hypothesis of Theorem II hold with the same conditions imposed on Θ_k, Θ, X_k, X as in that theorem; suppose further that, for each (x_1, \dots, x_n) of $(R^*)^n$, the expression $\prod_{i=1}^n f(\theta_1, \dots, \theta_\nu, x_i)$, regarded as a function of $(\theta_1, \dots, \theta_\nu)$, have a unique maximum $(\hat{\theta}_1, \dots, \hat{\theta}_\nu)$ interior to Ω , the above function being differentiable with respect to $\theta_1, \dots, \theta_\nu$ at $(\hat{\theta}_1, \dots, \hat{\theta}_\nu)$ and taking on a positive value at that point; and let the determinant $|\partial \Theta_k / \partial \theta_j| \neq 0$ at $(\hat{\theta}_1, \dots, \hat{\theta}_\nu)$; and suppose, lastly, that the functions $\hat{\theta}_j = \hat{\theta}_j(x_1, \dots, x_n)$ ($j = 1, \dots, \nu$) are admitted to constitute a system of statistics for the estimation of θ_j ($j = 1, \dots, \nu$); then this system of statistics is a sufficient one.*

In view of Theorem II, all that it is necessary for us to show is that, under the hypothesis of the present theorem, equations (11) are valid with $\phi_j(x_1, \dots, x_n)$ defined as $\hat{\theta}_j(x_1, \dots, x_n)$. Now on account of the differentiability at $(\hat{\theta}_1, \dots, \hat{\theta}_\nu)$, etc., the conditions for a maximum

$$(12) \quad \frac{\partial}{\partial \theta_j} \log \prod_{i=1}^n f(\theta_1, \dots, \theta_\nu, x_i) = 0 \quad (j = 1, \dots, \nu)$$

must be satisfied when $(\theta_1, \dots, \theta_\nu) = (\hat{\theta}_1, \dots, \hat{\theta}_\nu)$; these become, by use of (10),

$$(12') \quad \sum_{k=1}^{\nu} \left\{ \frac{\partial \Theta_k}{\partial \theta_j} \left[\sum_{i=1}^n X_k(x_i) \right] \right\} = -n \frac{\partial \Theta}{\partial \theta_j} \quad (j = 1, \dots, \nu)$$

and our task will be accomplished when it is shown that, for a given set of values $(\hat{\theta}_1, \dots, \hat{\theta}_\nu)$, these equations determine the ν quantities $[\sum_{i=1}^n X_k(x_i)]$ uniquely. (The existence of at least one set of values for the [] expressions is a direct consequence of the definition of $\hat{\theta}_j$.) Now the uniqueness is an immediate consequence of the assumption

$$\left| \frac{\partial \Theta_k}{\partial \theta_j} \right|_{(\hat{\theta})} \neq 0$$

of our hypothesis.

The statistic $(\hat{\theta}_1, \dots, \hat{\theta}_\nu)$ in this theorem is called by R. A. Fisher the "maximum likelihood" statistic, and Theorem III is our rendering of Fisher's theorem that if a sufficient statistic exist, the maximum likelihood gives such a statistic.

As examples, consider the normal distribution

$$f(\theta_1, \theta_2, x) = \frac{1}{\theta_2(2\pi)^{1/2}} e^{-(x-\theta_1)^2/(2\theta_2^2)},$$

the Cauchy distribution

$$f(\theta_1, \theta_2, x) = \frac{1}{\pi} \frac{\theta_2}{\theta_2^2 + (x - \theta_1)^2}$$

(each with $\Omega: -\infty < \theta_1 < +\infty, \theta_2 > 0$), and the distribution corresponding with Pearson's Type III curve

$$f(\theta_1, \theta_2, \theta_3, x) = \frac{1}{\theta_2 \Gamma(\theta_3 + 1)} \left(\frac{x - \theta_1}{\theta_2} \right)^{\theta_3} e^{-(x-\theta_1)/\theta_2}.$$

The following facts are observed to be true: The normal distribution admits a pair of sufficient statistics for (θ_1, θ_2) regarded as unknown, and also one for each of θ_1, θ_2 when the other is regarded as known. The Cauchy distribution does not admit any sufficient statistic for either parameter when the other is known, nor a pair for both when they are both unknown. The Pearson distribution admits sufficient statistics for θ_2 or θ_3 or both when θ_1 is regarded as known, but for no set of parameters involving θ_1 , no matter what assumptions are made with regard to the knowledge of θ_2, θ_3 .

Of these assertions, the positive ones are readily demonstrated by throwing the frequency function into the form (4) and taking the maximum likelihood statistics for ϕ_j .

In order to prove the negative assertions, we observe that if, in accordance with Theorem I, the function had the form (4), we should have

$$\frac{\partial^2}{\partial x \partial \theta_1} \log f = \sum_{k=1}^{\mu} \frac{\partial \Theta_k}{\partial \theta_1} \frac{dX_k}{dx}$$

(or, in dealing with the Cauchy distribution when θ_1 is known, a corresponding form in θ_2 ; here $\mu=1$ and the conclusion is immediate). If now we set $y = -\theta_1$, we see that the left-hand member is of the form $F(x+y)$, and satisfies the functional equation

$$F(x+y) = \sum_{k=1}^{\mu} A_k(x)B_k(y),$$

the solution of which (under the present conditions of differentiability) is known to be*

$$F(u) = \sum_{k=1}^N P_k(u)e^{\tau_k u},$$

where $P_k(u)$ is a polynomial, τ_k a constant. It is readily verified that this form does not apply in the cases to be considered here.

The same method, and the requirement of finiteness of $\int_{-\infty}^{+\infty} f dx$ (from which it follows that $\tau_k=0$ etc.), lead to the following:

THEOREM IV. *Let $f(\theta_1, \dots, \theta_r, x)$ satisfy the hypotheses of Theorem I, together with the further conditions that*

$$f(\theta_1, \dots, \theta_r, x) = f(\theta_2, \dots, \theta_r, x - \theta_1)$$

and that $\Omega \times R - T$ contain all the points of the θ_1 -axis (for every $\theta_2, \dots, \theta_r$). Then

$$f(\theta_1, \dots, \theta_r, x) = \exp \left[\sum_{k=1}^{\mu} Q_k(x - \theta_1)^{2k} + U + V \right],$$

where $Q_k = Q_k(\theta_2, \dots, \theta_r) < 0$, $U = U(\theta_1, \dots, \theta_r)$, $V = V(x, \theta_2, \dots, \theta_r)$.

* For references and an elegant proof, see P. Stäckel, *Sulla equazione funzionale $f(x+y) = \sum_{i=1}^n X_i(x)Y_i(y)$* , *Atti della Reale Accademia dei Lincei, Rendiconti*, vol. 22, pp. 392-393.