A REGRESSION PROBLEM

By D. V. LINDLEY

Received 25 October 1950

1. *Introduction.* During the Oxford Conference of the Econometric Society in September 1936, Ragnar Frisch proposed a problem in regression theory. A partial solution was found in 1938 by Miss H. V. Allen (1). A more complete solution was given by C. R. Rao (6) in 1947, and in the same year the present author (5) obtained a solution as a particular case of a more general result. These last two papers contained a flaw, and a correct solution was provided by Miss E. Fix (2). This last solution still leaves a part of the problem unanswered, and in the present paper a result of P. Lévy's (4), is used to complete the solution. At the same time further generalizations of the problem are considered and, in the cases of most practical importance, complete solutions are obtained. It is advisable, both from the point of view of rigour and simplicity of analysis, to use a general definition of the conditional expectation of a random variable. Accordingly, the paper begins with a summary of the relevant definitions. These notions were introduced by Kolmogoroff (3). It has been thought worth while giving the definitions here, in forms which are slightly different from Kolmogoroff's and seem more suitable for applications, in order to explain the notation and nomenclature used. The relevant consequences of these definitions are also stated in the form in which they are used.

2. *Conditional Expectation.* Let X be any collection of elements and $\mathfrak{F}$ a completely additive class, or field, of sets in X. Let $p(A)$ be a measure defined for every A belonging to $\mathfrak{F}$. If $p(X) = 1$ we say p is a *probability* (measure) and speak of a *probability field*, $(\mathfrak{F}, p)$. Let Y be another collection and $\mathfrak{G}$ a field in Y. If, to each element a of X, corresponds a unique element b of Y, we write $b = Ta$ and speak of the *mapping* T of X on to Y. In general the mapping will be many-one. Let B be any set in Y. The set A of a in X such that $Ta = b$ belongs to B, is called the *origin* of B in X: we write $A = T^{-1}B$. The mapping is said to be *measurable* if $T^{-1}B$ belongs to $\mathfrak{F}$ for any B of $\mathfrak{G}$. Only measurable mappings will be considered, when it is easy to verify that the real positive number $q(B) = p(T^{-1}B)$ is then a probability measure over $\mathfrak{G}$, and also that the origins of sets belonging to $\mathfrak{G}$ form a field in X, which is a subfield of $\mathfrak{F}$: we denote it by $\mathfrak{F}_T$. In general, $\mathfrak{F}_T$ will be a proper subfield of $\mathfrak{F}$. In the special case where Y is the class of all real numbers, R^1, and $\mathfrak{G}$ is the field of all Borel sets on the real line, $\mathfrak{B}^1$, $x = Ta$ is called a *random variable*, and the subfield of $\mathfrak{F}$ corresponding to this mapping is denoted by $\mathfrak{F}_x$. x is then a real-valued function defined over X, such that the origin of the set of points for which x is less than k, for any real k, belongs to $\mathfrak{F}$. q is then called the *probability distribution* of x. The function $F(a) = q(x < a)$ is called the *distribution function* (d.f.) of x.

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

Let $T_1, T_2, \ldots, T_n$ be n mappings of a class X on to n other classes: these mappings are said to be *independent* if the equation

$$p\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} p(A_i) \tag{2·1}$$

holds for any sets A_i belonging to $\mathfrak{F}_{T_i}$. In the special case where the mappings are all random variables, the random variables are said to be independent.

A random variable, x, is said to have an *expectation*, written $E(x)$, whenever the Lebesgue integral

$$(\mathfrak{F})\int_X x\,dp \tag{2·2}$$

exists and is finite. In (2·2) the reference to the field can usually be omitted without confusion. If A is any set of $\mathfrak{F}$ with $p(A) \neq 0$ it is easy to see that $p_A(B) = p(AB)/p(A)$ is also a probability measure over $\mathfrak{F}$ and consequently $(\mathfrak{F}, p_A)$ is also a probability field. p_A is called the *conditional probability, given* A. The *conditional expectation of* x, *given* A, which we write $E_A(x)$, is then defined as

$$(\mathfrak{F})\int_X x\,dp_A \tag{2·3}$$

whenever it exists and is finite. An alternative form for this is

$$E_A(x) = \frac{1}{p(A)}\int_A x\,dp. \tag{2·4}$$

The conditional expectation is thus defined only with respect to an event of non-zero probability.

The definition can be extended to an event of zero probability provided that it is regarded as belonging to a subfield $\mathfrak{F}_T$ induced by a mapping. Let T be a mapping of X on to another class Y. In applications X will consist of the objects under study, Y will be n-dimensional Euclidean space. Let b be an element of Y and x a random variable over X. If $p(T^{-1}b) \neq 0$ we can define $E_{T^{-1}b}(x)$ by the above method. It is then more convenient to speak of the conditional expectation of x given b, rather than given $T^{-1}b$, and write $E_b(x)$. For the general case we define the *conditional expectation of* x *given* b, which will be a function of b, as the random variable z defined by the equations

$$E_A(z) = E_A(x) \tag{2·5}$$

for every A belonging to $\mathfrak{F}_T$ with $p(A) \neq 0$, where $\mathfrak{F}_z$ is a subfield of $\mathfrak{F}_T$. The last restriction ensures that z is a random variable over Y. We write $z = E_b(x)$, a real-valued function of b. The existence of z follows from the Radon-Nikodym theorem, since (2·5) can be written

$$\int_A z\,dp = p(A)\,E_A(x) = \int_A x\,dp$$

and $p(A) = 0$ implies that the additive set function $\int_A x\,dp$ is also zero; the latter exists by (2·4) since $E(x)$ exists. By the same theorem the function z is unique save for equivalence. Two random variables z_1, z_2, are *equivalent* if $z_1 = z_2$ except possibly on a set of measure zero. In the special case where T is also a random variable y, say, we write $E_b(x)$ as $E_y(x)$. If the mapping T is on to the class of elements (y_1, y_2) where

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

y_1 and y_2 are two random variables over X we write $E_b(x)$ as $E_{y_1, y_2}(x)$; and generally for any number of random variables. If x and y are independent random variables $E_y(x) = E(x)$: the converse is not true.

With these definitions and results the regression problem can be easily formulated.

3. Let $(\mathfrak{F}, p)$ be a probability field over a class X. Let ξ, η, u and v be four random variables defined over X, and suppose $E(\eta)$ exists. Then $E_\xi(\eta)$ exists and is termed the *regression of* η on ξ. The regression is said to be *linear* if

$$E_\xi(\eta) = \gamma\xi + \delta, \tag{3·1}$$

where γ and δ are real constants.

Define two new random variables x and y by the equations

$$x = \xi + u, \quad y = \eta + v, \tag{3·2}$$

then the problem to be considered in this paper is: given (3·1) and (3·2) under what condition is the regression of y on x also linear? That is, when does the equation

$$E_x(y) = cx + d \tag{3·3}$$

hold?

Particular cases of this are of especial interest. Suppose that $\eta = \gamma\xi + \delta$; then (3·1) is trivial and a *functional* (or structural) *relationship* is said to exist between ξ and η. Ragnar Frisch's original problem belongs to this case: he supposed that

$$x = a\xi + u, \quad y = b\xi + v, \tag{3·4}$$

and desired to know when (3·3) held for unknown a and b. This is equivalent to considering the problem for all values of a and b. He also supposed that ξ, u and v were independent.

The practical problems, of which the above are the mathematical models, are usually of the following form. ξ and η are quantities which it is desired to measure and in the population are supposed to be such that the average value of η, when ξ is fixed, is a linear function of ξ, or in the particular case of a functional relationship, η is a linear function of ξ, as would happen in the case of many physical laws. Measurements of ξ, η are not accurate and errors, which are random, u and v, are introduced, so that the final data do not consist of sets of values of ξ and η, but of x and y. The analysis would be simplified if y had a linear regression on x (there cannot, it is clear, be a functional relationship), and consequently it is natural to inquire when this is possible. In Frisch's problem the regression requirement is to hold for all a and b.

Without loss of generality $E(\eta)$, which must exist for the problem to have a meaning, can be supposed zero. Similarly δ, in (3·1), may be supposed zero, so that $E_\xi(\eta) = \gamma\xi$. Since

$$E \,.\, E_\xi(\eta) = E(\eta) = 0$$

it then follows that $E(\xi) = 0$. It will also be in agreement with practical requirements if attention is confined to the case where the errors are unbiased. In fact, it is convenient to assume that

$$E(u) = 0, \tag{3·5}$$

and

$$E_x(v) = 0. \tag{3·6}$$

((3·6) is implied by $E_{\xi, u}(v) = 0$, which is often a more convenient assumption in practice.)

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

4. Under these conditions an answer to the problem is given by the following

THEOREM 4·1. *Let ξ, η, u and v be four random variables defined over a class X with probability field $(\mathfrak{F}, p)$. Let x and y be defined by* (3·2). *Suppose $E(\eta) = 0$ and*

$$E_\xi(\eta) = \gamma\xi, \tag{4·1}$$

Then provided (3·5) *and* (3·6) *hold, and*

$$E_{\xi,u}(\eta) = E_\xi(\eta), \tag{4·2}$$

a necessary and sufficient condition that $E_x(y)$ exists and satisfies

$$E_x(y) = cx + d, \tag{4·3}$$

where c and d are constants, is that

$$\gamma E_x(\xi) = cx, \tag{4·4}$$

and then $d = 0$.

(4·3) may be written $$\int_A (cx+d)\,dp = \int_A y\,dp$$

for any set A belonging to $\mathfrak{F}_x$. By the definition of a random variable it is sufficient that this holds for the sets in X over which $x < a$, for all real a.

From (3·2) and (3·6) we have

$$\int_{x<a} (cx+d)\,dp = \int_{x<a} (\eta+v)\,dp, \quad = \int_{x<a} \eta\,dp,$$

which shows that $E_x(y)$ exists.

Since the set of points in X for which $x < a$ belongs to $\mathfrak{F}_{\xi,u}$, it follows from (4·2) and (4·1) that

$$\int_{x<a} (cx+d)\,dp = \int_{x<a} \gamma\xi\,dp.$$

If $a = +\infty$ in this last equation, the use of (3·5) shows that $d = 0$. Hence

$$\int_{x<a} cx\,dp = \int_{x<a} \gamma\xi\,dp$$

which is (4·4). Since the argument can be reversed the result is established.

Condition (4·2), like (3·6), is a condition on the errors which is likely to be satisfied in any practical application. (3·5) is only a convention to simplify the argument. There are important special cases in which one or other of these equations is true. We express the results in the following corollaries.

COROLLARY 4·1. *If u is independent of the pair (ξ, η), condition* (4·2) *may be omitted.*

Let A be the set of elements in X for which $\xi < a$ and $u < b$. Then by the definition of $E_{\xi,u}(\eta)$

$$\int_A E_{\xi,u}(\eta)\,dp = \int_A \eta\,dp. \tag{4·5}$$

By the assumption of the corollary,

$$\left.\begin{aligned} \int_A \eta\,dp &= \int_{u<b} dp \int_{\xi<a} \eta\,dp \\ &= \int_{u<b} dp \int_{\xi<a} E_\xi(\eta)\,dp \\ &= \int_A E_\xi(\eta)\,dp. \end{aligned}\right\} \tag{4·6}$$

By (4·5) and (4·6), (4·2) follows.

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

COROLLARY 4·2. *If $E(v) = 0$ and v is independent of the pair (ξ, u), condition* (3·6) *may be omitted.*

This is immediate since then $E_{\xi,u}(v) = E(v) = 0$ and hence, in particular $\int_{x<a} v\,dp = 0$ which is the only condition on v used in the proof of the theorem.

COROLLARY 4·3. *If $E(u)$ and $E(v)$ exists, v is independent of the pair (ξ, u) and u is independent of (ξ, η), condition* (4·4) *is necessary and sufficient for* (4·3) *to be true.*

By combining the two previous corollaries it follows that (4·2) and (3·6) are true. The remaining conditions, $E(u) = E(v) = 0$, only ensure $d = 0$. Their existence is sufficient to ensure the existence of the regression (4·3) with $d \neq 0$.

COROLLARY 4·4. *If $\eta = \gamma\xi$, $E(u)$ and $E(v)$ exist and v is independent of (ξ, u),* (4·4) *is necessary and sufficient for* (4·3) *to be true.*

When the functional relationship exists, (4·2) is trivial. Corollary (4·2) then completes the result. (This is the case considered by Rao (6).)

5. The above results show that condition (4·4) is the essential condition which must be satisfied in order that the regression exist and be linear. This condition may be written

$$\gamma E_{\xi+u}(\xi) = c(\xi + u), \tag{5·1}$$

and is thus a relationship between ξ and u only. It is clear that for any distribution of ξ, there is a joint distribution of ξ and u satisfying (5·1) and thus given any distribution of true value, there is an error distribution which preserves the linearity of the regression. This error will usually be correlated with ξ, and it is of interest to examine whether it is possible that ξ and u be independent, the case considered by previous writers. To do this it is convenient to use the *characteristic function* (c.f.) of a random variable, x, defined as $E(e^{itx})$ where t is real: we write $\phi_x(t)$. $\phi_x(0) = 1$ and $\phi_x(t)$ is continuous so that $\log_e \phi_x(t)$ is defined in some neighbourhood of the origin. This function is called the *cumulative function* (cm.f.) of x: we write $K_x(t)$. The *joint characteristic function* of two random variables (x, y) is similarly defined as $E(e^{itx+iuy})$.

THEOREM 5·1. *Under the conditions of corollary* (4·3) *a necessary and sufficient condition for equation* (4·3) *is that the cm.f. of u be a constant times the cm.f. of ξ.* (Lindley (5) under less general ccnditions.)

We have to show the condition of the theorem is equivalent to (4·4). (4·4) is

$$c\int_{x<a} x\,dp = \gamma\int_{x<a} \xi\,dp, \tag{5·2}$$

since this holds for all a, and $|e^{itx}| = 1$, we have

$$c\int_X e^{itx}(\xi + u)\,dp = \gamma\int_X e^{itx}\xi\,dp. \tag{5·3}$$

ξ and u are independent so that (5·3) may be written

$$c\int_X e^{it\xi}\,dp\int_X e^{itu}u\,dp = (\gamma - c)\int_X e^{it\xi}\xi\,dp\int_X e^{itu}\,dp.$$

By the uniformity of convergence of the integral obtained by differentiating with respect to t it follows that

$$c\phi_\xi(t)\,\phi'_u(t) = (\gamma - c)\,\phi'_\xi(t)\,\phi_u(t).$$

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

This equation integrates to give, since $\phi_\xi(0) = \phi_u(0) = 1$ and both functions are continuous,

$$\phi_\xi(t)^{c-\gamma}\phi_u(t)^c = 1.$$

Since neither function is zero, this is

$$(\gamma - c)\,K_\xi(t) = cK_u(t). \tag{5·4}$$

The argument may be reversed, and the result is established.

Since the modulus of a c.f. is not greater than one, it follows that $\gamma - c$ and c must have the same sign. Thus if $\gamma > 0$ and ξ is not a constant, $\gamma > c > 0$. The regression line is tilted toward the ξ-axis.

6. It follows from the above analysis that if the regression condition is satisfied for one value of γ it is satisfied for all γ. Thus even if γ is unknown the theorems are still applicable. Ragnar Frisch's problem is different in that it is a question concerning regression for all values of a and b (see (3·4)). This is equivalent to a change of scale in the ξ and η axes. The latter is seen not to matter since the conditions of Theorem 4·1 are invariant under such a change. The change of scale in ξ, however, affects condition (4·4), which must now be replaced by

$$\gamma E_{a\xi+u}(a\xi) = c(a\xi + u) \tag{6·1}$$

for all permitted a. We shall suppose a is confined to A, some closed interval on the real line. Frisch's problem will therefore be answered if we can discover when (6·1) is satisfied for all a in A. Frisch (and the other writers mentioned in §1) considered the case where u is independent of ξ, and we shall deal with this case first. The fundamental relation is (5·4). Since $K_{a\xi}(t) = K_\xi(at)$, this becomes, under the new conditions,

$$K_u(t) = f(a)\,K_\xi(at) \tag{6·2}$$

for a in A, where $f(a)$ is some function of a. Without loss of generality it can be supposed that $a = 1$ is in A, whence combining (6·2) with the same equation with $a = 1$, $K_u(t)$ may be eliminated to obtain the result

$$f(a)\,K(at) = f(1)\,K(t), \tag{6·3}$$

where the suffix ξ has been omitted for convenience. If we ignore the trivial case where ξ and u are constants, $f(a)$ may be eliminated between (6·3) and the same equation with $t = 1$, and we have

$$K(1)\,K(at) = K(a)\,K(t). \tag{6·4}$$

Let t be in A and r be a positive integer. Then

$$\frac{K(t^r)}{K(1)} = \left(\frac{K(t)}{K(1)}\right)^r. \tag{6·5}$$

It follows that the same equation holds for all rational r and thus, by continuity, for all $r > 0$. If A is an interval of the *positive* axis we see that

$$K(z) = K(1)\,z^\alpha \tag{6·6}$$

for $z > 0$, with $\alpha \geqslant 0$. With $a = z$, $t = -1/z$ in (6·4) we have

$$K(z) = K(-1)\,|z|^\alpha \tag{6·7}$$

for $z < 0$. The form of $K(z)$ is then given by (6·6) and (6·7), with, of course, $K(0) = 0$. If A is an interval of the *negative* axis the same result ensues. If A includes 0 as an interior point we obtain the result that $K(1) = K(-1)$.

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

Again the argument may be reversed and hence the necessary and sufficient condition for (6·1) with ξ independent of u is that $K_\xi(t)$ satisfies (6·6) and (6·7), and $K_u(t)$ is a multiple of $K_\xi(t)$. It remains to show that these are possible forms for cm.f.'s. If, for example, they were not, there would be no solution to Frisch's problem.

To do this we use Lévy's result (4) that (6·6) and (6·7) are only equations of a cm.f. if $K(-1)$ is the conjugate of $K(1)$, $0<\alpha\leqslant 2$, and $K(1) = -c_0+ic_1$, where c_0 and c_1 are real and satisfy

$$c_0 \geqslant 0, \quad |c_1 \cos \tfrac{1}{2}\pi\alpha| \leqslant c_0 \sin \tfrac{1}{2}\pi\alpha. \tag{6·8}$$

In fact such cm.f.'s are the cm.f.'s of the stable distributions. A distribution is said to be stable if, whenever x_1 and x_2 are two independent random variables having this distribution, and c_1 and c_2 are any two positive constants, the random variable x defined by the equation $c_1x_1+c_2x_2 = cx$ has the same distribution for a suitable chosen positive constant c. In our case the mean exists, so that $\alpha \geqslant 1$, and our result is

THEOREM 6·1. *Let ξ, u and v be three random variables, which are not constants, and*

$$x = a\xi+u, \quad y = b\xi+v.$$

Then, provided $E(\xi) = E(u) = E(v) = 0$, u is independent of ξ and $E_{\xi,u}(v) = 0$, a necessary and sufficient condition that the regression of y on x be linear for all a and b is

(a) If $a<0$ or $a>0$, that ξ and u both have stable distributions.

(b) If values of a of both signs are permitted then in addition the c.f.'s must be real. (This result has been obtained by Miss Fix (2) without the demonstration that (6·6) and (6·7) are necessarily cm.f.'s.)

COROLLARY 6·1. *If the variance of ξ exists, the distribution must be normal.*

For then $\alpha = 2$ and hence, by (6·8), $c_1 = 0$, so that $K(z) = -c_0z^2$. The normal distribution is the only stable distribution with finite variance.

7. We now consider Frisch's problem without the restriction that ξ and u be independent. (6·1) may be written as

$$c\int_X e^{it(a\xi+u)}(a\xi+u)\,dp = \gamma\int_X e^{it(a\xi+u)}a\xi\,dp,$$

which in terms of $\phi(s,t)$, the joint characteristic function of ξ and u, is

$$(\gamma-c)\,a\frac{\partial\phi(s,t)}{\partial s} = c\frac{\partial\phi(s,t)}{\partial t},$$

with $s = at$. That is

$$\frac{\partial\phi(s,t)}{\partial s} = f\left(\frac{s}{t}\right)\frac{\partial\phi(s,t)}{\partial t}, \tag{7·1}$$

where $f(s/t) = f(a) = c/(\gamma-c)\,a$ is some function of a. (γ is not equal to c since this would imply $\phi(0,t) = 1$, that is u is a constant.)

The integration of (7·1) yields

$$\phi(s,t) = F\left[sG\left(\frac{s}{t}\right)\right],$$

where F is an arbitrary function and G is a function derived from the f of (7·1). On putting $s = t$ we obtain

$$\phi(s,t) = \phi\left[th\left(\frac{s}{t}\right), th\left(\frac{s}{t}\right)\right], \tag{7·2}$$

where

$$h(s/t) = G(s/t)/G(1).$$

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

We confine our attention to the more important case where the variables, ξ and u, possess variances, and where a can have any value. If $s = at$ in (7·2), the coefficients of t^2 in the Taylor series expansion of $\phi(at, t)$, which exists thus far because the variances exist, may be compared. From this it follows that

$$h(a) = \sqrt{(A + Ba + Ca^2)}. \tag{7·3}$$

Hence
$$\phi(at, t) = \phi(th(a), th(a)), \tag{7·4}$$

with $h(a)$ given by (7·3). For fixed a, $\phi(at, t)$ is the c.f. of $a\xi + u$: so that since the c.f. uniquely determines the distribution, (7·4) is equivalent to

$$p\left[\frac{a\xi + u}{h(a)} < c\right] = p(\xi + u < c) \tag{7·5}$$

for any constant a. Thus the distribution of $a\xi + u$ is, apart from a scale factor, $h(a)$, independent of a. Let the class of distributions for which this is true be denoted by D: we have to specify D. Since $h^2(a)$ is a quadratic in a, it follows that by a linear transformation of ξ and u, we can change to new variables, ξ' and u', for which $h(a) = \sqrt{(1 + a^2)}$ and D is left invariant.

Hence
$$\phi\left(\frac{at}{\sqrt{(1+a^2)}}, \frac{t}{\sqrt{(1+a^2)}}\right), \tag{7·6}$$

the c.f. of ξ' and u', does not depend on a. Now

$$\phi(s, t) = \iint \exp[is\xi' + itu']\, dF(\xi', u'), \tag{7·7}$$

where $F(\xi', u')$ is the d.f. of ξ' and u'. Transform to variables

$$s = \rho\cos\psi, \quad \xi' = r\cos\theta,$$
$$t = \rho\sin\psi, \quad \eta' = r\sin\theta,$$

(7·7) is then
$$\Phi(\rho, \psi) = \iint \exp[i\rho r\cos(\theta - \psi)]\, dG(r, \theta), \tag{7·8}$$

where $G(r, \theta)$ is the d.f. of r and θ, and by (7·6) this does not depend on ψ. To find the form of $G(r, \theta)$ such that this holds, we use the inversion formula

$$G(R, \Theta) = \frac{1}{4\pi^2}\int_0^{\to\infty} \rho\, d\rho \int_0^{2\pi} d\psi \int_0^R \int_0^\Theta \exp[-i\rho R\cos(\Theta - \psi)]\, R\, dR\, d\Theta\, \Phi(\rho, \psi), \tag{7·9}$$

which holds for any continuity sector of $G(R, \Theta)$. A sector $0 \leqslant r \leqslant R$, $0 \leqslant \theta \leqslant \Theta$ is called a *continuity sector* if $\int dG(r, \theta)$ over the boundary of the sector, excluding the origin, is zero. To prove this result we substitute (7·8) in the right-hand side of (7·9) before the limit has been taken, and obtain

$$\begin{aligned}
&\frac{1}{4\pi^2}\int_0^\rho \rho\, d\rho \int_0^{2\pi} d\psi \int_0^R \int_0^\Theta \int_0^\infty \int_0^{2\pi} \exp\{i\rho[r\cos(\theta - \psi) - R\cos(\Theta - \psi)]\}\, R\, dR\, d\Theta\, dG(r, \theta) \\
&\quad = \frac{1}{2\pi}\int_0^R \rho\, d\rho \int_0^R \int_0^\Theta \int_0^\infty \int_0^{2\pi} J_0\{\rho\sqrt{[R^2 + r^2 - 2Rr\cos(\theta - \Theta)]}\}\, R\, dR\, d\Theta\, dG(r, \theta) \\
&\quad = \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} dG(r, \theta) \int_0^\rho \rho\, d\rho \int_0^R \int_0^\Theta J_0\{\rho\sqrt{[R^2 + r^2 - 2Rr\cos(\theta - \Theta)]}\}\, R\, dR\, d\Theta,
\end{aligned}$$

where J_0 is the Bessel function of zero order.

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

By Neumann's integral theorem (Watson, (7)) this equals

$$\frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} dG(r,\theta)\, K(\rho;\, r,\theta),$$

where $\lim_{\rho\to\infty} K(\rho;\, r,\theta)$ has the values:

2π inside the sector $0 \leqslant r \leqslant R, 0 \leqslant \theta \leqslant \Theta$;
π on the boundary of the sector, excluding the origin;
0 outside the sector; and
Θ at the origin.

The result then follows, provided that we make the convention that $G(R,\Theta)$ includes a proportion, $\Theta/2\pi$, of the probability that $r = 0$, $\theta = 0$. A d.f. is uniquely determined by its values on the continuity sectors, and hence (7·9) determines $G(R,\Theta)$ everywhere. It is, of course, the analogue in polar coordinates of the Lévy inversion formula.

When $\Phi(\rho,\psi)$ does not depend on ψ, the integration with respect to ψ in (7·9) may be performed. The integrand does not then depend on Θ and hence

$$G(R,\Theta) = G(R)\,\Theta/2\pi.$$

Hence Θ is uniformly distributed independently of R. The distribution of ξ' and u' is then, for obvious reasons, termed *circular*. The distribution of ξ and u is then *elliptical*.

To obtain the class D it is sufficient to find those distributions of ξ which are such, that it is possible to find another variable, u, so that the joint distribution of (ξ, u) is circular. Let $\psi = 0$ in (7·8). Then $\Phi(\rho,0) = \Phi(\rho)$ is the characteristic function of ξ'. Integrate with respect to ψ and then Θ in (7·9), we obtain

$$G(R) = \int_0^{\to\infty} \rho\, d\rho \int_0^R J_0(\rho R)\, R\, dR\,.\,\Phi(\rho). \tag{7·10}$$

Then $G(R)$ is the standardized d.f. of $R = \sqrt{(\xi^2+u^2)}$. A d.f. is standardized if the value at any discontinuity point is equal to the mean of the limits to the right and left at the point. Hence the class D is characteristic by the fact that if $\Phi(\rho)$ is the characteristic function of a distribution belonging to D, then the transform of $\Phi(\rho)$ by (7·10) must be a standardized d.f. We have therefore

THEOREM 7·1. *There exist random variables ξ and u with $E(\xi^2)$ and $E(u^2)$ finite and non-zero, $E(\xi) = E(u) = 0$, satisfying* (6·1) *for all a if, and only if, ξ and u have an elliptical distribution. In order that there be such a distribution for a given characteristic function $\Phi(\rho)$ of ξ it is necessary and sufficient that the function $G(R)$, defined by* (7·10) *be a standardized d.f.*

Equation (7·5) is a natural generalization of the idea of a stable distribution. Two random variables, x, y not necessarily independent, may be called *jointly stable* if, for any constants a and b, the random variable z, defined by $ax+by = cz$, has the same distribution as x for a suitably chosen constant c. The above theory shows that if the variances of x and y exist, then jointly stable variables must have an elliptical distribution.

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press

REFERENCES

(1) ALLEN, H. V. *Statistical Research Memoirs*, 2 (1938), 60.
(2) FIX, E. *Proceedings of the Berkeley Symposium on mathematical statistics and probability* (California, 1949), 79.
(3) KOLMOGOROFF, A. *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Berlin, 1933). [English translation by N. Morrison (Chelsea, New York, 1950).]
(4) LÉVY, P. *Théorie de l'addition des variables aléatoires* (Paris, 1937). §§30 and 56.
(5) LINDLEY, D. V. *Suppl. J. Roy. Stat. Soc.* 9 (1947), 218.
(6) RAO, C. R. *Econometrica*, 15 (1947), 245. (Correction—*Econometrica*, 17 (1949), 212.)
(7) WATSON, G. N. *Bessel functions* (Cambridge, 2nd ed. 1948), p. 475, equation 3, and the first note on p. 471.

STATISTICAL LABORATORY
CAMBRIDGE

https://doi.org/10.1017/S0305004100026694 Published online by Cambridge University Press