DISTRIBUTIONS WHICH LEAD TO LINEAR REGRESSIONS

EVELYN FIX

UNIVERSITY OF CALIFORNIA, BERKELEY

1. In 1936, during the Oxford Conference of the Econometric Society, Ragnar Frisch proposed the following question.

Suppose it is known that two random variables X and Y have the following composition:

$$\left.\begin{aligned} X &= a\xi + \alpha \\ Y &= b\xi + \beta \end{aligned}\right\} \tag{1}$$

where ξ, α, and β are some mutually independent random variables and a and b are certain constant coefficients, the values of which are unknown. What are the conditions under which the regression of Y on X, and also that of X on Y, is linear, irrespective of the values of a and b?

A partial answer to the question was given by H. V. Allen [**1**].[1] Miss Allen proved this theorem: provided the first two moments of β and all the moments of ξ and α exist, the necessary and sufficient condition for linearity of regression of Y on X, whatever may be a and b, is that both ξ and α should be normally distributed. The proof was based on the construction of an infinite sequence of polynomials, normal and orthogonal with respect to the elementary probability law of X, and therefore the existence of all moments as postulated was necessary in this manner of proof. However, as the author herself points out, the condition that all moments exist is restrictive and makes her answer only a partial one. It may be noted that the distributions considered in many practical problems do have finite moments of all orders.

The purpose of this article is to consider the same problem under assumptions regarding the moments which are less restrictive than those of Miss Allen.

2. First, a precise interpretation must be given the problem of Ragnar Frisch. Let X and Y be any two random variables and let $F_{X,Y}(x,y)$ and $F_X(x)$ stand for the joint cumulative distribution of both variables and for the marginal cumulative distribution of X, respectively. Thus, for all real x and y,

$$F_{X,Y}(x,y) = P\{(X \leqq x)(Y \leqq y)\} \tag{2}$$

and

$$F_X(x) = P\{X \leqq x\} = \lim_{y\to\infty} F_{X,Y}(x,y)\,. \tag{3}$$

Statements concerning the regression of Y on X will be interpreted to presuppose the existence, for all real x except perhaps for a set of probability zero

[1] Boldface numbers in brackets refer to references at the end of the paper (see p. 91).