NOTE ON REGRESSION FUNCTIONS IN THE CASE OF THREE SECOND ORDER RANDOM VARIABLES

By Clyde A. Bridger

The study of the correlation of two second-order random variables has received the attention of several authors, among them Yule [1], Charlier [2], Wicksell [3, 4], and Tschuprow [5]. Yule writes of them under the guise of "attributes." The study of three or more second order random variables has lagged behind. In this note we shall examine the regression function of one second order random variable on two others by considering the problem from the point of view of Tschuprow's [6] paper on the correlation of three random variables.

A variable X that takes on m values $x_1, \cdots, x_m$ with corresponding probabilities $p_1, \cdots, p_m$ subject to the condition $\sum_i p_i = 1$ is defined as a random variable of order m. (In particular, if X takes on only two values, x and x' with probabilities p and q, where $p + q = 1$, X is a random variable of second order.) The system of values x and probabilities p constitute the law of distribution of X. In the case of two random variables, X and Y, there exists a joint distribution law, covering all possible combinations of X and Y, together with their associated probabilities $p_{11}, \cdots, p_{mn}$ the joint distribution law contains all of the information regarding the stochastical dependence of X and Y.

The extension to more than two variables is immediate. Let p_{ijk} represent the probability of the simultaneous occurrence of the set of values x_i, y_j, z_k of three random variables X, Y, and Z; p_{ij}, that of the simultaneous occurrence of x_i, y_j together without reference to Z; p_i, that of the occurrence of x_i without reference to Y or Z; etc. Then, we have relationships of the types $\sum_i \sum_j \sum_k p_{ijk} = \sum_i \sum_j p_{ij} = \sum_i p_i = 1$; $\sum_i p_{ijk} = p_{jk}$; $\sum_i \sum_j p_{ijk} = \sum_j p_{jk} = \sum_i p_{ik} = p_k$. Similarly, let $p_{jk}^{(i)}$ be the probability of the simultaneous occurrence of y_j and z_k on the condition that X takes on the value x_i; $p_j^{(i)}$, that of the occurrence of y_j without reference to Z, on the same condition; etc. Then

$$\sum_k p_k^{(i)} = \sum_k p_k^{(ij)} = \sum_j \sum_k p_{jk}^{(i)} = 1; \qquad \sum_j p_{jk}^{(i)} = p_k^{(i)}; \qquad p_i p_j^{(i)} = p_{ij};$$

$$p_{ij} p_k^{(j)} = p_i p_{jk}^{(i)} = p_i p_j^{(i)} p_k^{(ij)} = p_{ijk}; \qquad \sum_i p_i p_j^{(i)} = p_j; \text{ etc.}$$

Denoting by $E(x)$ or simply Ex the expression "the mean value or mathematical expectation of x," we have $m_{fgh} = EX^f Y^g Z^h = \sum_i \sum_j \sum_k p_{ijk} x_i^f y_j^g z_k^h$. In particular, the mean values of the distributions are given by $m_X = EX$