REMARKS ON A MULTIVARIATE TRANSFORMATION[1]

By Murray Rosenblatt

University of Chicago

The object of this note is to point out and discuss a simple transformation[2] of an absolutely continuous k-variate distribution $F(x_1, \cdots, x_k)$ into the uniform distribution on the k-dimensional hypercube. A discussion of related transformations has been given by P. Lévy [1].

Let $X = (X_1, \cdots, X_k)$ be a random vector with distribution function $F(x_1, \cdots, x_k)$. Let $z = (z_1, \cdots, z_k) = Tx = T(x_1, \cdots, x_k)$, where T is the transformation considered. Then T is given by

$$z_1 = P\{X_1 \leq x_1\} = F_1(x_1),$$
$$z_2 = P\{X_2 \leq x_2 \mid X_1 = x_1\} = F_2(x_2 \mid x_1),$$
$$\vdots$$
$$z_k = P\{X_k \leq x_k \mid x_{k-1} = x_{k-1}, \cdots, X_1 = x_1\} = F_k(x_k \mid x_{k-1}, \cdots, x_1).$$

One can readily show that the random vector $Z = TX$ is uniformly distributed on the k-dimensional hypercube, for

$$P\{Z_i \leq z_i ; i = 1, \cdots, k\}$$

$$= \int_{\{Z \mid Z_i \leq z_i\}} \cdots \int d_{x_k} F_k(x_k \mid x_{k-1}, \cdots, x_1) \cdots d_{x_1} F_1(x_1)$$

$$= \int_0^{z_k} \cdots \int_0^{z_1} dz_1 \cdots dz_k = \prod_{i=1}^k z_i,$$

when $0 \leq z_i \leq 1$, $i = 1, \cdots, k$. Hence $Z_1, \cdots, Z_k$ are uniformly and independently distributed on [0, 1].

Let $X_{(i)} = (X_{1(i)}, \cdots, X_{k(i)})$, $i = 1, \cdots, n$, be a random sample of n vectors from a population with distribution function $F(x_1, \cdots, x_k)$ and let $G(x_1, \cdots, x_k)$ be the corresponding sample distribution function. It has been found that the probability distribution of the Kolmogorov-Smirnov statistic,

$$\max_{x_1, \cdots, x_k} \mid F(x_1, \cdots, x_k) - G(x_1, \cdots, x_k) \mid,$$

is not the same for all continuous F when $k > 1$ [2]. The same can be said for the multidimensional von Mises statistic

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (F(x_1, \cdots, x_k) - G(x_1, \cdots, x_k))^2 \, d_{x_1 \cdots x_k} F(x_1, \cdots, x_k).$$

However, it would still be of interest to study the Kolmogorov-Smirnov and von Mises statistics for the case of sampling from a population uniformly distributed on the k-dimensional hypercube. One could test whether the $X_{(i)}$, $i = 1, \cdots, n$, are a sample from a population with distribution function $F(x_1, \cdots, x_k)$ by

---

[1] Work done under ONR contract.

[2] I have recently learned that J. H. Curtiss and I. R. Savage have also considered this transformation.

using either the Kolmogorov-Smirnov or von Mises statistic to test whether the $Z_{(i)} = TX_{(i)}$ are a sample from a population uniformly distributed on the k-dimensional hypercube. Both statistics are invariant under permutation of the $X_{(i)}$, $i = 1, \cdots, n$.

The transformation is of some interest in setting up regions of equal probability mass in a chi-square test of goodness of fit. Assume again that one wishes to test whether the $X_{(i)}$, $i = 1, \cdots, n$, come from a population with absolutely continuous distribution function $F(x_1, \cdots, x_k)$. This, as remarked before, is equivalent to testing whether the $Z_{(i)} = TX_{(i)}$, $i = 1, \cdots, n$, come from a population with distribution function $z_1 \cdots z_k$, where $0 \leqq z_j \leqq 1$, $j = 1, \cdots, k$. Divide the k-dimensional hypercube into the N^k cells

$$C_{j_1 \cdots j_k} = \left\{ z \left| \frac{j_i}{N} < z_i < \frac{j_i + 1}{N} \right., i = 1, \cdots, k \right\},$$

where $j_1, \cdots, j_k = 0, 1, \cdots, N - 1$. Each of these cells has probability mass $1/N^k$ under the null hypothesis. Let $\nu_{j_1 \cdots j_k}$ be the number of transformed observations $Z_{(i)} = TX_{(i)}$ in $C_{j_1 \cdots j_k}$. The chi-square statistic

$$\sum \frac{\left(\nu_{j_1 \cdots j_k} - \dfrac{n}{N^k} \right)^2}{n/N^k}$$

can be used to test whether the sample $X_{(i)}$ comes from a population with distribution function $F(x_1, \cdots, x_k)$. A test for goodness of fit based on the chi-square statistic above is consistent as n, $N \to \infty$.

There are $k!$ transformations T of the type described above corresponding to the $k!$ ways in which one can number the coordinates $x_1, \cdots, x_k$. One might think this unsatisfactory because of a suspicion that the experimenter might carry out one of the procedures described above with each of the $k!$ transformations and then choose that particular transformation which yields the result he wishes to obtain. But this situation can arise in any case where there is a multitude of tests in the same context.

The transformation T can be written down explicitly in several cases. We shall write down the transformation when $F(x_1, \cdots, x_k)$ is a normal distribution with mean $M = (m_1, \cdots, m_k)$ and covariance matrix $\Lambda = \{ \lambda_{ij} \}$, $i, j = 1, \cdots, k$. Let $\Lambda^{(r)} = \{ \lambda_{ij} \}$, $i, j = 1, \cdots, r \leqq k$, and $\Lambda_{ij}^{(r)}$ be the cofactor of $\{ \lambda_{ij} \}$ in $\Lambda^{(r)}$. The transformation T is then given by

$$F_1(x_1) = \Phi\left(\frac{x_1 - m_1}{\sqrt{\lambda_{11}}} \right),$$

$$F_2(x_2 \mid x_1) = \Phi\left(\frac{x_2 - m_2 + (\Lambda_{21}^{(2)}/\Lambda_{22}^{(2)})(x_1 - m_1)}{\sqrt{\Lambda^{(2)}/\Lambda_{22}^{(2)}}} \right),$$

$$\vdots$$

$$F_k(x_k \mid x_{k-1}, \cdots, x_1) = \Phi\left(\frac{x_k - m_k + \sum_{j=1}^{k-1} (\Lambda_{kj}/\Lambda_{kk})(x_j - m_j)}{\sqrt{\Lambda/\Lambda_{kk}}} \right).$$

Let $F(x_1, x_2)$ be a normal distribution with means m_1, m_2, variances σ_1^2, σ_2^2 and correlation coefficient ρ. The transformation can then be written as

$$F_1(x_1) = \Phi\left(\frac{x_1 - m_1}{\sigma_1}\right),$$

$$F_2(x_2 \mid x_1) = \Phi\left(\frac{x_2 - m_2 + \frac{\rho\sigma_1}{\sigma_2}(x_1 - m_1)}{\sigma_2\sqrt{1 - \rho^2}}\right).$$

REFERENCES

[1] P. Lévy, *Théorie de l'Addition des Variables Aléatoires*, Gauthier-Villars, Paris, 1937, pp. 71–73, 121–123.
[2] Paul B. Simpson, "Note on the estimation of a bivariate distribution function," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 476–478.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Eugene meeting of the Institute, June 19–21, 1952)

1. The Auditory Cortex—A Probability Model. Archie R. Tunturi, University of Oregon Medical School.

The role played by the brain in communication is well known, but in what manner the brain handles information is not understood. Some progress has been made in this direction by studying the anatomy and physiology of the auditory cortex in the anesthetized dog with controlled acoustic signals. Communication may be thought of as making a representation in a space of a representation in another space. In three of the four auditory areas (on one side of the brain), the entire frequency spectrum from 100 to 12800 cps is represented literally spacewise by groups of cells that respond only to a narrow range of frequencies. A special method increases the signal to noise ratio, by augmenting the electrical response of the cells, thereby permitting exact measurements of the characteristic frequency and intensity for each group of cells. This is similar to a narrow band filter, and does not reveal the effect of other frequencies on the information. The information capacity of the system can be inferred if it can be assumed that occurrence of the augmented response for the group of cells follows some probability function. These probabilities for all groups of cells can be assembled into a model representing the behavior of the system as a communication device. If there are 70 groups of cells between 100 and 12800 cps, the probability of any particular combination would be $1/2^{70}$, if the selections were equally probable. The effect of noise on this system will be considered. (Research sponsored in part by the Office of Naval Research.)

2. Testing Message Diffusion: The Utilization of Mathematical Models. Stuart C. Dodd, Richard J. Hill, and Susan Huffaker, University of Washington.

In connection with the study of interpersonal verbal communication, the Washington Public Opinion Laboratory designed an experimental procedure which yielded data on the temporal diffusion of thirty-three different messages in a population of 184 individuals. Data (including the recipient of each message, the initiator of communication, and the time of communication) were obtained on 5,522 separate instances of communication. The