ABSTRACTS OF PAPERS

Presented on September 16, 1945 at the Rutgers meeting of the Institute

1. On The Variance of a Random Set in n Dimensions.

HERBERT ROBBINS, Lieutenant USNR Postgraduate School, Annapolis, Md.

Using a general formula for the moments of the measure of a random set X (*Ann. Math. Stat.* Vol. XV (1944), pp. 70–74) we find the mean and variance in the case where X is a random sum of n-dimensional intervals with sides parallel to the coordinate axes, thus generalizing the results previously found (loc. cit.) for the case $n = 1$.

2. The Non-Central Wishart Distribution and its Application to Problems in Multivariate Statistics.

T. W. ANDERSON, Princeton University.

The non-central Wishart distribution is the joint distribution of sums of squares and cross-products of deviations of observations from multivariate normal distributions with identical variance-covariance matrices and with different sets of means. The rank of the non-central Wishart distribution is defined as the rank of the matrix of sets of means. In a previous paper (by M. A. Girschick and the present author) the non-central Wishart distribution is given explicitly for the rank one and two cases and indicated for the case of any rank. In the present paper the characteristic function of the non-central Wishart distribution is given for general rank. The distribution, which is given in the form of a multiple integral, is the product of a central Wishart distribution and a symmetric function of the roots of a determinantal equation involving the matrix of squares and cross products of observations and the matrix of population means. It is shown that the convolution of two non-central Wishart distributions is again a non-central Wishart distribution if the variance-covariance matrices are the same. The moments of the generalized variance and the moments of the likelihood ratio criterion for testing certain linear hypotheses (for example, the hypothesis that the means of a set of populations are identical, given that the matrices of population variances and covariances are the same) are obtained for the linear and planar non-central cases in terms of infinite series. Likelihood ratio criteria are developed for testing the dimensionality of the means of a set of multivariate populations (with identical variances and covariances) on the basis of one sample from each. The criterion for testing whether the dimensionality is h in the space of p dimensions is a symmetric function of $p - h$ smallest roots of the determinantal equation involving the sample estimate of the matrix of variances and covariances and the sums of squares and cross-products of deviations of sample means. The maximum likelihood estimate of the hyperplanes and positions of means on them are obtained. The asymptotic distributions of the criteria are χ^2-distributions.

3. The Effect on a Distribution Function of Small Changes in the Population Function.

BURTON H. CAMP, Wesleyan University.

It is generally assumed in the application of distribution theory that, if the actual population function is not very different from the one used in the theory, then the true sampling distribution of a statistic will not be very different from the one obtained in the theory. But elsewhere in mathematics we do not assert that a conclusion will be only slightly modified by a small deviation in the hypothesis. This paper presents some theorems which are useful in determining the maximum effect on a sampling distribution of certain kinds of small changes in the population function.

4. Composite Distributions. Casper Goffman and Benjamin Epstein, Westinghouse Electric Corporation.

Let $f(x; \theta_1, \theta_2, \cdots, \theta_n)$ be a function such that for every point $\theta_1 = \theta_{10}, \cdots, \theta_n = \theta_{n0}$ in parameter space, x is a random variable with p.d.f. $f(x; \theta_{10}, \cdots, \theta_{n0})$. Suppose further that the parameters $\theta_1, \theta_2, \cdots, \theta_n$ are themselves random variables whose p.d.f.'s are given respectively by $\phi(\theta_1), \cdots \phi(\theta_n)$. Using a concept of "probability contained in an interval" and an axiom based on this concept, we show that x is a random variable with p.d.f. $g(x)$ given by the formula

$$(1) \qquad g(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x; \theta_1 \cdots, \theta_n) \phi(\theta_1) \cdots \phi(\theta_n) \, d\theta_1 \cdots d\theta_n .$$

In this paper we consider statistical properties of the function $g(x)$ in cases of particular interest in applications. The cases treated here are (a) where the mean, $\bar{x}$, is the only variable parameter, (b) where the standard deviation, σ, is the only variable parameter, and (c) where the mean $\bar{x}$, and the standard deviation, σ, are both variable parameters; $\bar{x}$ and σ being independent.

It is shown that problems (a) and (b) are equivalent respectively to the sum and product of two independent random variables, one of which has zero mean. Formulae for the moments in problem (c) are then derived in terms of the formulae obtained for (a) and (b).

5. Population, Expected Values and Sample. E. J. Gumbel, New School for Social Research.

Let x be an unlimited continuous variate, and let $F(x)$ be the probability of a value equal to, or less than, x. Then the expected m^{th} values $\hat{x}_m$, for n observations, are approximations to the most probable m^{th} values and defined by $F(\hat{x}_m) = F_1 + (F_n - F_1)(m - 1)/(n - 1)$, where F_1 and F_n are the probabilities of the most probable first and the most probable last value. The probabilities F_1, $1 - F_n$ and $(F_n - F_1)/(n - 1)$ are of the order of magnitude $1/n$.

The distribution of the expected values $\hat{x}_m$ differs from the distribution of the sample and from the theoretical distribution. However, for a symmetrical distribution the mean and the odd moments about mean calculated from the expected values coincide with the mean and the moments of the population. For the normal distribution, the expected standard deviation $\sigma(n)$ divided by the standard deviation σ of the population and traced on normal probability paper approximates a linear function of $\sqrt{\log n}$. The approach of $\sigma(n)$ toward σ is slow. For 500 observations, $\sigma(n)$ is about 99% of σ. The moments of the distribution of the expected values exist even in the case that the moments of the theoretical distribution diverge.

6. On Optimum Estimates for Stratified Samples. Morris H. Hansen and William N. Hurwitz, Bureau of the Census.

A stratified sample is drawn from a population with R strata. Neyman found the optimum sample allocation for the "best unbiased linear estimate." However, biased but consistent estimates of the form $\frac{x_i'}{y_i'}$ where both x_i' and y_i' are random variables have been found to give more reliable results in a large class of problems. Even more efficient estimates can be obtained by finding the values of n_i (the sample size) and w_i which minimize the mean square error of estimates of the form $\Sigma w_i \frac{x_i'}{y_i'}$ or $\frac{\Sigma w_i x_i'}{\Sigma w_i y_i'}$.

7. Pearsonian Correlation Coefficients Associated with Least Squares Theory.
PAUL S. DWYER, University of Michigan. (Read by Title).

In least squares theory we have the predicting variable x, the observed value of the predicted variable, y, the residual e, and the predicted value of the predicted variable y. The purpose of this paper is to study the Pearsonian coefficients resulting from correlating all these variables in pairs (a) in the case of a single predicted variable and (b) in the case of two or more predicted variables. The results yield such coefficients as multiple correlation, multiple alienation, partial correlation, part correlation, and new coefficients not previously in use. The results are given in expanded, determinant, and matrix form. A simplified calculational technique is provided.