ABSTRACTS OF PAPERS

Presented December 22, 1947 at the Berkeley Meeting of the Institute

1. The Performance Characteristic of Certain Methods for Obtaining Confidence Intervals. B. M. BENNETT and J. NEYMAN, University of California, Berkeley.

Certain methods for obtaining confidence limits have been introduced by Bliss, R. A. Fisher and Paulson. Thus, e.g., let x_i, y_i $(i = 1, \cdots, n)$ represent a sample from a bivariate normal population with means $E(x_i) = \xi$, $E(y_i) = \alpha\xi$ and variances and covariance σ_x^2, σ_y^2, σ_{xy}. If $\bar{x}$, $\bar{y}$, S_x^2, S_y^2, S_{xy} are the sample means, variances and covariance respectively, then in order to determine confidence limits for α, the ratio:

$$u = \frac{\sqrt{n}(\bar{y} - \alpha\bar{x})}{\sqrt{S_y^2 - 2\alpha S_{xy} + \alpha^2 S_x^2}}$$

may be referred to the appropriate value t_ϵ of the Student-t distribution. The inequality: $| u | < t_\epsilon$ may, in general, be solved as a quadratic equation in α to yield two values $\underline{\alpha}$, $\bar{\alpha}$ which are presumed to be confidence limits for α. In this paper the probability π of being correct in using such a procedure, i.e., the performance or operating characteristic, is computed in the limiting case when σ_x^2, α_y^2, $\sigma_{xy} = \rho\sigma_x\sigma_y$ are assumed to be known. It is shown that π is a function $\pi(\alpha, \xi, \sigma_x, \sigma_y, \rho)$ of all the parameters, and in particular of α itself, the quantity for which confidence limits are supposed to be provided. Similar "quadratic" methods are also used in certain regression problems, e.g., in determining confidence limits for a value of x corresponding to an additional value of y when a previous sample regression of y on x is available; or in determining confidence limits for the intersection point of two population regression lines. The performance characteristic of each of these methods is shown to be a function of the quantity for which the method gives confidence limits.

2. Some Further Results on the Bernoulli Process. T. E. HARRIS, Douglas Aircraft Co.

Let $z_1, z_2, z_3, \cdots$, be a sequence of random variables defined as follows: $P(z_1 = r) = p_r$, $r = 0, 1, 2, \cdots, k$. If $z_n = 0$, $z_{n+1} = 0$. If $z_n = r$, $r \neq 0$, then z_{n+1} is distributed as the sum of r independent random variables, each having the same distribution as z_1. It is assumed that $x < 1$, where $x = E(z_1)$. Let N be the smallest value of n such that $z_{n+1} = 0$. A method is given for obtaining an expansion of the moment-generating function of N. In the case where $p_r = 0$ for $r \geq 3$, this expansion takes the form $1 + (1 - e^{-s})(1 - p_0)$ $F(s)$, where $F(s) = f_1(s) - p_2(1 - p_0)f_2(s) = 2xp_2^2(1 - p_0)^2f_3(s) - \cdots$, where $f_1(s) = (e^{-s} - x)^{-1}$, and $f_n(s) = f_{n-1}(s)(e^{-s} - x^n)^{-1}$. Certain restrictions on the constants p_r insure that this expansion converges for a complex neighborhood of $s = 0$.

3. Most Powerful Tests of Composite Hypotheses I. Normal Distributions. E. L. LEHMANN and C. M. STEIN, University of California, Berkeley, California.

Critical regions are determined for testing a composite hypothesis, which are most powerful against a particular alternative among all critical regions whose probabilities under the hypothesis tested are bounded above by the level of significance. These problems have been considered by Neyman, Pearson and others, subject to the condition that the critical region be similar. In testing the hypothesis specifying the value of the variance of a normal distribution with unknown mean against an alternative with larger variance, and in some other problems, the best similar region is also most powerful in the sense of this paper. However, in the analogous problem when the variance under the alternative hypothesis is less than that under the hypothesis tested, in the case of Student's hypothesis when the level of significance is less than $\frac{1}{2}$, and in some other cases, the best similar region is not most powerful in the sense of this paper. There exist most powerful tests which are quite good against certain alternatives in some cases where no proper similar region exists. These results indicate that in some practical cases the standard test is not best if the class of alternatives is sufficiently restricted.

4. **On the Selection of Forecasting Formulas.** PAUL G. HOEL, University of California, Los Angeles, California.

Given two competing formulas, $u = g(z_1, \cdots, z_m)$ and $v = h(z_1, \cdots, z_m)$, for forecasting a variable x, a significance test possessing optimum properties is designed for deciding whether one formula yields significantly better forecasts than the other. The test, which turns out to be a Student t test, is constructed as a test of the hypothesis $H_0 : m_i = u_i$ against the alternative $H_1 : m_i = v_i$, $(i = 1, \cdots, n)$, in which it is assumed that the variables $x_1, \cdots, x_n$, corresponding to the n samples, are independently normally distributed with means m_i and variances $\sigma_i^2 = \sigma^2$.

5. **On the Power Function of the "Best" t-test Solution of the Behrens-Fisher Problem.** J. E. WALSH, Douglas Aircraft Company

The most powerful t-test solution of the Behrens-Fisher problem (one-sided and symmetrical) was obtained by Scheffé in *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 35-44. This note derives (approximately) the power efficiency of this t-test for the case in which the ratio of the variances of the normal populations is also known. Let the t-test be based on m sample values from the first normal population and n sample values from the second normal population, where $m \leq n$. For fixed values of m and n, a symmetrical t-test with significance level 2α has the same power efficiency as a one-sided t-test with significance level α. For one-sided t-tests with significance level α, the power efficiency is approximately $50[B + \sqrt{B^2 - 8(m + n)A}]/(m + n)$, where $B = 2 + (m + n)A + K_\alpha^2/2$, $A = 1 - K_\alpha^2/2(m - 1)$, and $K\alpha$ is the standardized normal deviate exceeded with probability α. This approximation is reasonably accurate for $m \geq 4$ if $\alpha = .05$, $m \geq 5$ if $\alpha = .025$, $m \geq 6$ if $\alpha = .01$, $m \geq 7$ if $\alpha = .005$. Intuitively the power efficiency of a test measures the percentage of available information per observation which is utilized by that test.

6. **On Sequences of Experiments.** CHARLES STEIN, University of California, Berkeley, California.

One performs a sequence of N experiments to decide between two simple hypotheses regarding probability distributions of certain observable quantities. At each stage there is a choice among L experiments and the one chosen yields a random variable. One wishes to achieve certain upper bounds α and β to the probabilities of first and second kind errors respectively, and, subject to these restrictions, to minimize the expected cost under a third hypothesis. The cost of each particular sequence of experiments is known. A solution is obtained, essentially by applying Lagrange's method and working back from the end of the experiment. This can be generalized to multiple decision problems. The results are applied to two-sample tests with the second sample of variable size, and to Wald's sequential analysis. As another problem, suppose (X_1, Y_1), $(X_2, Y_2) \cdots$ are independently distributed with bivariate normal distributions having mean ξ and covariance matrix Σ, both unknown. One tests $H_0 : \xi = 0$ against $H_1 : \xi'\Sigma^{-1}\xi = \delta$. A test (not necessarily optimum) valid within the usual approximation is obtained from the ratio of the p.d.f.

of Hotelling's T^2 under H_1 to that under H_0. Analogous results hold for the multiple correlation coefficient, ratio of two variances and test for linear hypothesis.

7. The Effect of Selection Above Definite Lower Limits of Linear Functions of Normally Distributed Correlated Variables on the Means and Variances of Other Linear Functions. G. A. BAKER, University of California, Davis, California.

Sometimes certain variables in a system can be observed before other economically or socially important variables. These variables or linear combinations of them can be used as a basis of selection at given levels. The question is: How does selection on these earlier or more easily available variables affect the mean and variance of the economically or socially more important variables or, perhaps, linear functions of the more important variables. The general procedure is clear. We transform to a new system of variables which contains the linear functions on which selection is performed and the linear functions of which the means and variances are required as separate variables. The remaining new variables are eliminated by integration. The final calculation involves the numerical evaluation of integrals whose integrands are the product of polynomials and normal multivariate functions and whose limits-depend on the given levels of selections. The general ideas are simple but the actual labor of computation in a given case is tedious. An example is considered in detail.

8. An Inversion Formula for the Distribution of a Ratio of Random Variables. J. GURLAND, University of California, Berkeley, Calif.

The repeated Cauchy principal value of integrals applied to characteristic functions is used in obtaining inversion formulae for distribution functions. Let the random variables X_1 and X_2 have a joint distribution function with corresponding characteristic function $\phi(t_1, t_2)$. Suppose $P\{X_2 \leq 0\} = 0$. Let $\oint g(t)\,dt = \lim_{\substack{\epsilon \to 0 \\ T \to \infty}} \left(\int_{-T}^{-\epsilon} + \int_{\epsilon}^{T} \right) g(t)\,dt$ for any function $g(t)$. If $G(x)$ is the distribution function of X_1/X_2 then $G(x) + G(x - 0) = 1 - \frac{1}{\pi i} \oint \frac{\phi(t, -t_x)}{t}\,dt$. This formula is free of restrictions which accompany the formula given by Cramer in the case where X_1 and X_2 are independent; and differentiation extends a result of Geary to a much larger class of distribution functions. Further generalizations of the theory are obtained, and as an example the distribution function of the ratio of quadratic forms of random variables $X_1, X_2 \cdots X_n$ is considered in the case where $X_1, X_2 \cdots X_n$ have a multivariate normal distribution.

9. Independence of Parameters and Sufficient Statistics. E. W. BARANKIN, University of California, Berkeley, California.

The notions of *complete set of independent parameters* and *minimal set of sufficient statistics* are suitably defined for a class of families of probability densities $\{p(x_1, \cdots, x_n; \vartheta_1, \cdots, \vartheta_\nu)\}$, and the order of each of these sets is determined as the rank of a certain matrix. Second order continuous differentiability is eventually required of the function p; and certain other conditions are laid down, designed to ensure that the behavior of p in the large is similar to its behavior in the small when only continuous differentiability is assumed. The problem of determining the order of a minimal set of sufficient statistics is made, by certain device, to become identical in character with that of finding the order of a complete set of independent parameters. (This is in the nature of these concepts.)

An explicit method is given for finding a complete set of independent parameters and a minimal set of sufficient statistics.

(Presented December 30, 1947 at New York at the Annual Meeting of the Institute)

1. Distribution of the Circular Serial Correlation Coefficient for Residuals from a Fitted Fourier Series (*Preliminary Report*). R. L. ANDERSON, University of North Carolina, Raleigh, North Carolina and T. W. ANDERSON, Columbia University.

Given a set of N observations $\{X_i\}$, which are defined as follows:

$$X_i - \mu_i = \rho\cdot(X_i - L - \mu_i - L) + \epsilon_i ,$$

where the residuals $\{\epsilon_i\}$ are assumed to be normally and independently distributed with zero means and equal variances and L is the lag. A statistic for testing the null hypothesis: $\rho = 0$ is ${}_LR$, the circular serial correlation coefficient of residuals e_i from a regression line fitted by least squares: $X_i = M_i + e_i$. The following regression line is considered:

$$M_i = a_0 + \sum_k{}' a_k \operatorname{Cos} \frac{2\pi ki}{N} + \sum_k{}' b_k \operatorname{Sin} \frac{2\pi ki}{N},$$

where k ranges over some subset of the integers $1, 2, \cdots, \frac{1}{2}(N-1)$ or $\frac{1}{2}(N)$, depending on whether N is odd or even (if N is even, $b_{\frac{1}{2}N}$ is not used). Hence ${}_LR$ is defined as:

$${}_LR = \frac{e_1 e_{L+1} + e_2 e_{L+2} + \cdots + e_N e_{L+N}}{\Sigma e_i^2},$$

with $e_{i+N} = e_i$.

The distribution of this ${}_LR$ has the same general form as that presented by R. L. Anderson for $\rho = 0$ ["Distribution of the serial correlation coefficient," Annals of Math. Statistics 13:1-13(1942)]; and for $\rho \neq 0$ by W. G. Madow ["Note on the distribution of the serial correlation coefficient," *Annals of Math. Statistics* 16:308-310(1945)].

For M_i consisting of terms of only *one* period, $\frac{N}{k} = 2, 3, 4, 6, 12$ and 24, exact values of the 1% and 5% significance levels of ${}_1R$ have been computed for $N = 12$ and 24. Approximate significance levels have been computed for $N = 12(12)96$. More of the exact significance levels are being computed, and all computations will be extended to include some multiple periods and some lags greater than 1.

2. Some New Methods for Distributions of Quadratic Forms. HAROLD HOTELLING, Institute of Statistics, University of North Carolina, Chapel Hill.

Any homogeneous quadratic form in normally distributed variates of zero means has the same distribution as $q = \frac{1}{2}(a_1x_1^2 + \cdots + a_nx_n^2)$, where the a_i are roots of a determinantal equation based on the coefficients of the given form and the parameters of the normal distribution, and where the x_i are normally and independently distributed with zero means and unit variances. We take $\Sigma a_i = n$, and begin by expanding the distribution of a positive definite form in a series of powers of q whose coefficients are polynomials in the reciprocals of the a_i. This series shows the analyticity of the function, which is then expressed as the product of a χ^2 distribution function of a series of Laguerre polynomials with coefficients which are simple polynomials in the moments of the a_i. Indefinite forms and certain ratios of forms are dealt with by convolutions of these series and by other means.

3. Frequency Functions Defined by the Pearson Difference Equation. LEO KATZ, Michigan State College, East Lansing, Michigan.

Frequency "links" formed from the Pearson difference equation provide an efficient means of fitting functions to observed distributions. These links, involving three constants which are determined by the first four moments of the observed series, correspond to a three-parameter family of discrete frequency functions. This family of functions is just as broad as that defined by the differential equation, containing functions of equally diverse types; in addition, it has the very important advantage that the graduation process is the same for any type. Further, the simpler functions of the family all correspond to points lying in one plane of the parameter space. This plane, giving a two-parameter family of functions (depending upon the first three moments), is studied intensively, rather complete results being obtainable for areas, moments, sampling characteristics of moments, etc. It is also shown that the problem of discrimination among simple discrete frequency functions for graduating observed data is resolvable (in the plane) to the sampling distribution of one statistic. A special case of the two-parameter family depending on only the first two moments was previously discussed.

4. Distribution of the Sum of Roots of a Determinantal Equation under a Certain Condition. D. N. NANDA, Institute of Statistics, University of North Carolina, Chapel Hill.

Let $x = \| x_{ij} \|$ and $x^* = \| x^*_{ij} \|$ be two p-variate sample matrices with n_1 and n_2 degrees of freedom. Then $S = xx'/n_1$ and $S^* = x^*x^{*\prime}/n_2$ are, under the null hypothesis, independent estimates of the same population covariance matrix. The distribution of a root, specified by its rank order, of the determinantal equation $| A - \theta(A + B) | = 0$, where $A = n_1S$ and $B = n_2S^*$, has already been given by S. N. Roy, and by the author, who has also obtained the limiting distribution of any root when one of the samples becomes infinitely large. The moment generating function of the sum of the roots when $n_1 = p \pm 1$ can be derived from the limiting distribution of the largest root. The probability distributions of the sum of roots under this condition have been formulated for the determinantal equations having two, three, and four roots. The moments of these distributions have also been obtained. The method is applicable for the determinantal equation of any order. These probability distributions can easily be tabulated, as they involve only simple algebraic and incomplete beta functions.

5. Applications of Carnap's Probability Theory to Statistical Inference. GERHARD TINTNER, Iowa State College, Ames, Iowa.

The new theory of probability of Rudolf Carnap ("On inductive logic," *Philosophy of Science*, vol. 12, 1945, pp. 72 ff. "The two concepts of probability," *Philosophy and Phenomenological Research*, vol. 5, 1944, pp. 513 ff.) introduces a distinction between probability$_1$, the degree of confirmation, and probability$_2$, related to relative frequency. It is believed, that the ideas developed are useful in clarifying the problems of statistical inference.

As an example, consider the case of "inverse inference," i.e. inference from a sample to the population. The evidence is that in a sample of size s there are s_1 individuals with a certain property M and $s_2 = s - s_1$ without the property. The hypothesis is that in the population consisting of n individuals there are n_1 individuals with property M and $n_2 = n - n_1$ individuals without this property. The degree of confirmation is then:

$$c^* = \frac{\binom{n_1 + w_1 - 1}{s_1 + w_1 - 1}\binom{n_2 + w_2 - 1}{s_2 + w_2 - 1}}{\binom{n + k - 2}{n - s}}$$

In this formula we have: w_1 the logical width of the property M, w_2 the logical width of the property non-M, $k = w_1 + w_2$. It should be noted that for $w_1 = w_2 = 1$ the formula becomes the classical result, i.e. a term of the hypergeometric distribution.

This idea may be applied to statistical estimation. We could for instance choose n_1 in such a fashion that c^* becomes a maximum. This would be estimation by the principle of maximum degree of confirmation, analogous to maximum likelihood. In a similar fashion we may also use c^* to establish limits for n_1 similar to confidence or fiducial intervals.

6. **Circular Probable Error of an Elliptical Gaussian Distribution.** HALLETT H. GERMOND, S. W. Marshall & Co., Consulting Engineers, Washington, D. C.

Preliminary tables are presented, giving the radii of distribution-centered circular cylinders enclosing various percentages of the volume under an elliptical bivariate Gaussian surface. These tables are further interpreted in terms of a correlated bivariate Gaussian distribution. The application of these tables to impact analysis is illustrated.

(Presented December 29, 1947 at the Chicago Meeting of the Institute)

1. **The Asymptotic Analogue of the Theorem of Cramér and Rao.** HERMAN RUBIN, Institute for Advanced Study, Princeton, N. J.

The author generalizes the results of Cramér and Rao on the minimum variance of estimates to the case of the asymptotic distribution of an estimate. He shows that if certain regularity conditions are satisfied, the formula given by Cramér and Rao remains valid. The main results are obtained in the case of consistent estimates, but with a stronger set of hypotheses, the results remain true for estimates which are not consistent. The method used to obtain these results is to construct statistics to which the theorem of Cramér and Rao can be applied, and whose variance converges to the variance of the limiting distribution. This procedure is also applied to the case in which there is no limiting distribution, and in which two sequences of distributions are considered which act as if they approach each other.