ABSTRACTS

(Abstracts of papers presented at the Los Angeles Meeting of the Institute, December 27–28, 1957)

1. Non-parametric Multiple-Decision Procedures for Selecting That one of K Populations Which has the Highest Probability of Yielding the Largest Observation.

(Preliminary Report) ROBERT BECHHOFER, Cornell University AND MILTON SOBEL, Bell Telephone Laboratories. (By title)

Let X_i be chance variables with density function $f_i(x)$, and let

$$p_i = \text{Prob}\ \{X_i > \max_{j \neq i} X_j\} (i = 1, \cdots, k).$$

Then $\sum_{i=1}^{k} p_i = 1$. Let $p_{[1]} \leqq \cdots \leqq p_{[k]}$ denote the ranked p_i. Let θ^*, P^* $(1 < \theta^* < \infty, l/k < P^* < 1)$ be specified constants. The goal is to select the population associated with $p_{[k]}$; the procedure must guarantee, (*) Prob {Correct Selection | $p_{[k]} \geqq \theta^* p_{[k-1]}$} $\geqq P^*$. Procedure: "At the mth stage take the vector-observation $\mathbf{x}_m = (x_{lm}, \cdots, x_{km})$ where the x_{ij} $(j = 1, 2, \cdots)$ are independent observations from the ith population. Consider $\mathbf{y}_m = (y_{lm}, \cdots, y_{km})$ which is obtained by replacing the largest component of $\mathbf{x}_m$ by unity, and all other components by zero. Then $\mathbf{y}_m$ is an observation from a multinomial distribution with probability p_i associated with the ith component $(i = 1, 2, \cdots, k)$. (*) now can be guaranteed by continuing with procedures already proposed, e.g., these *Annals*, Vol. 27, p. 861. If $f_i(x) = g\{(x - \mu_i)/\theta\}$ $(i = 1, 2, \cdots, k)$, then the procedure can be used for selecting the population associated with the largest μ_i for any θ, known or unknown. Similar non-parametric procedures in which pairs of observations are taken from each population at each stage of experimentation, and which employ the range of each pair can be used for selecting that one of k populations which has the highest probability of yielding the largest sample range. If $f_i(x) = h\{(x - \mu_i)/\theta_i\}$ $(i = 1, 2, \cdots, k)$, then these latter procedures can be used for selecting the population associated with the largest θ_i for any set of μ_i, known or unknown. (Research supported in part by the U. S. Air Force through the Air Force Office of Scientific Research, ARDC, Contract No. AF 18(600)-331.) (Received September 25, 1957.)

2. The Asymptotic Efficiency of Friedman's χ_r^2-test.

PH. VAN ELTEREN, Mathematical Centre, Amsterdam. (By title)

Let $F(x)$ be a continuous cdf with density function $f(x) = F'(x)$ and let

$$x_{\mu\nu}\ (\mu = 1, 2, \cdots, m;\ \nu = 1, 2, \cdots, n)$$

be a chance variable with distribution $F_{\mu\nu}(x) = F(x + \theta_\nu + \eta_\mu)$. It is assumed for convenience, that $\sum_\nu \theta_\nu = 0$. Friedman (1937) has constructed the χ_r^2-test for the hypothesis $\theta_1 = \theta_2 = \cdots = \theta_n = 0$ (*J. Amer. Stat. Assn.*, Vol. 32, pp. 675–699). For alternatives $\theta_\nu = \theta_{\nu m} = \delta_\nu/\sqrt{m}$, where the δ_ν are given constants satisfying $\sum_\nu \delta_\nu = 0$, the asymptotic relative efficiency for $m \to \infty$ in the sense of Pitman of Friedman's test with respect to the corresponding 2-way-analysis of variance test is found to be $e_n = 12n(n + 1)^{-1}[\sigma \int f^2(x)\, dx]^2$, where σ^2 is the variance associated with $F(x)$. If $f(x)$ is normal, e_n reduces to $e_n = 3n/\pi(n + 1)$. (Received August 19, 1957.)

3. Experiments With Mixtures.

HENRY SCHEFFÉ, University of California.

Experiments with mixtures of q components are considered, whose purpose is the empirical prediction of the response to any mixture of the components, when the response depends on the proportions $x_1, x_2, \cdots, x_q$ of the components present but not on the total amount. The factor space is then the $(q - 1)$-dimensional simplex where $x_1 + \cdots + x_q = 1$, $x_i \geqq 0$. An experimental design called the *simplex lattice* and some modifications are treated; in the simplex lattice $x_i = 0, 1/m, 2/m, \cdots, 1$ for $i = 1, \cdots, q$ and some positive integer m, and the responses of all mixtures possible with these proportions are observed. The usual resolution of the response into general mean, main effects, and interactions does not seem possible, and so polynomial regression is employed. The problem of fitting an nth degree polynomial in $x_1, \cdots, x_q$ to the response is complicated by the fact that different polynomials give the same function on the simplex. Useful canonical forms are developed for $n \leqq 3$. The coefficients in these forms are interpreted as various kinds of synergisms. The analysis of experiments with these designs leads to classes of polynomials orthogonal on the lattices. The paper will appear in *J. Royal Stat. Soc., Series B.* (Received October 25, 1957.)

4. Least-Squares Estimation when Residuals are Correlated. M. M. Siddiqui, University of North Carolina.

Let $y_j, j = 1, \cdots, N$ be observations on a variate and let $y_j = \sum_{i=1}^{p} \beta_i x_{ij} + \Delta_j$, $j = 1, 2, \cdots, N$, where x_{ij} are non-stochastic, and $\Delta' = (\Delta_1, \cdots, \Delta_N)$ is a $N(0, \sigma^2 P)$ vector, where 0 is a zero vector and P is an $N \times N$ correlation matrix. Using the usual least-squares estimates, b_i, of β_i which are obtained by minimizing $\sum \Delta_j^2$, and s^2 of σ^2, the covariance matrix of b_i is obtained for general P and bounds are set on these covariances by first obtaining the maxima and minima of a quadratic and a bilinear form $u'Au$ and $u'Av$ where u and v are $N \times 1$ vectors and A is an $N \times N$ real symmetric matrix under the conditions $u'u = v'v = 1, u'v = 0$. (Received October 31, 1957.)

5. A Property of Additively Closed Families of Distributions. Edwin L. Crow, Boulder Laboratories, National Bureau of Standards.

Consider a one-parameter additively closed family of univariate cumulative distribution functions $F(x; \lambda)$ (H. Teicher, *Ann. Math. Stat.*, Vol. 25 (1954), pp. 775–778). Let three cumulants with orders in arithmetic progression exist and be non-zero. If all three orders are even, or if the first order is odd, it is also required that $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$. Consider linear combinations, with real, non-zero coefficients, of a finite number of independent variables with distributions in the family. It is proved that the only such linear combinations whose distributions are also in the family are those with coefficients unity. The additively closed families having this property may be called *strictly additively closed.* It can be shown that (one-parameter) additively closed stable families of distributions (normal and Cauchy in particular) with characteristic functions continuous in λ are not strictly additively closed, while Poisson, generalized Poisson, binomial, and gamma families are strictly additively closed. (Received October 31, 1957.)

6. Determining Sample Size for a Specified Width Confidence Interval. Franklin A. Graybill, Oklahoma State University.

If an experimenter decides to use a confidence interval to locate a parameter, he is concerned with at least two things: (1) Does the interval contain the parameter? (2) How wide is the interval? In general the answer to these questions cannot be given with absolute certainty, but must be given with a probability statement. The problem the experimenter then faces is: The determination of n, the sample size, such that (A) the probability will be equal to α that the confidence interval contains the parameter, and (B) the probability will be equal to β^2 that the width of the confidence interval will be less than d units (where α,

β^2, and d are specified). To solve this problem will generally require two things: (1) The form of the frequency function from which the sample of size n is to be selected; (2) Some previous information on the unknown parameters in the frequency function. This suggests that the sample be taken in two steps; the first sample will be used to determine the number of observations n to be taken in the second sample so that (A) and (B) will be satisfied. For a confidence interval on the mean of a normal population with unknown variance this problem has been solved by Stein for $\beta^2 = 1$. In this paper a theorem is proved which gives a method for determining n so that (A) and (B) will be satisfied. The theorem holds for parameters in the normal distribution and other distributions as well. (Received October 30, 1957.)

7. Nonparametric Estimation of Sample Percentage Point Standard Deviation. JOHN E. WALSH, Lockheed Aircraft Corporation.

The available data consists of a random sample $x(1) < \cdots < x(n)$ from a reasonable well-behaved continuous statistical population. The problem is to estimate the standard deviation of a specified $x(r)$ that is not in the tails of the sample. The estimates examined are of the form $a[x(r + i) - x(r - i)]$ and the explicit problem consists of determining suitable values for a and i. The solution $a = (1/2)(n + 1)^{-3/10}\{[r/(n + 1)][1 - r/(n + 1)]\}^{1/2}$ and $i \doteq (n + 1)^{4/5}$ appears to be satisfactory. Then the expected value of the estimate equals the standard deviation of $x(r)$ plus $O(n^{-9/10})$; also the standard deviation of this estimate is $O(n^{-9/10})$. That is, the fixed and random errors for this point estimate are of the same order of magnitude with respect to n. Solutions can be obtained which decrease the order of one of these types of error. However, these solutions increase the order of the other type of error, so that the over-all error magnitude exceeds $O(n^{-9/10})$. (Received November 7, 1957.)

8. On the Structure of Distribution-Free Statistics. C. B. BELL, Xavier University of Louisiana and Stanford University.

Let $X_1, \cdots, X_n$ be a sample of a one-dimensional random variable X which has continuous cpf F. It has been observed that the distribution-free statistics commonly appearing in the literature can be written in the form $\Phi[F(X_1), \cdots, F(X_n)]$, where Φ is a measurable symmetric function defined on the unit cube. Such statistics are said to have structure (d). In establishing that having structure (d) is equivalent to being symmetric and strongly distribution-free for properly closed, symmetrically complete classes of cpf's, this paper extends a result of Birnbaum and Rubin while employing different methodology. These results interest a statistician because (1) they indicate that one should construct a statistic of structure (d) whenever one wishes to design a distribution-free statistic; and (2) they guarantee that each symmetric, strongly distribution-free statistic is of structure (d), and, hence, that the value of its cpf at any point is the volume of a polyhedral region in the unit cube. Under such circumstances the work of numerous statisticians indicate that it should be possible to evaluate the cpf explicitly; reduce it to recursion formulae; tabulate it with high-speed computers; or evaluate its limiting distribution. (Received November 7, 1957.)

9. On the Supremum of the Poisson Process. RONALD PYKE, Stanford University.

Let $\{X(t); t \geqq 0\}$ be a Poisson process (with shift) for which $\log E(e^{iwX(t)}) = -itw\alpha + \lambda t(e^{iw} - 1)$, $w \in R_1$, $\alpha, \lambda > 0$. Define $\sigma(x, T) = Pr\left\{\sup_{0 < t \leqq T} X(t) \leqq x\right\}$. Let $X_1, X_2, \cdots X_n$ be the ordered random variables of n independent and uniform $-$ (0, 1) random variables. The distribution function, $Pr\left\{\max_{1 \leqq i \leqq n} (ai - X_i) \leqq x\right\}$, is obtained for all $a, x \in R_1$. For $a = 1/n$, this reduces to the distribution function of D_n^+ (cf. Birnbaum and Tingey, A.M.S., Vol. 22). Utilizing this result, $\sigma(x, T)$ is obtained explicitly. Applications of these expressions to queueing theory and distribution-free statistics are given.

10. On the Distributions of Various Sums of Squares in an Analysis of Variance Table for Different Classifications With Correlated and Non-homogeneous Errors. B. R. BHAT, Karnatak University. (Preliminary Report) (By Title)

The distributions of various sums of squares in an analysis of variance table for two way classification have been obtained by Box (Ann. Math. Stat., Vol. 25, pp. 484–498) under the assumption that the vectors $X_{.j}$ for $j = 1, 2, \cdots q$ are independent vector observations from a p-variate normal population with mean μ and covariance matrix Σ. The vector $X_{.j}$ for each jth level of the factor B denotes p observations corresponding to the p levels of the other factor A. This paper gives the distributions of the various sums of squares for any n-way classification under similar normality assumptions. It is noted that these distributions, in general, follow a simple pattern and so is their mutual dependence. For $n = 3$, if we have a third factor C at r levels in addition to the above factors A and B and if we assume that $X_{..k}$ for $k = 1, 2, \cdots r$ are independent vector observations from a pq-variate normal population, then, according to the general pattern the first set consists of distributions of the sums of squares for the correction term, main effects A and B and their interaction $A\ B$. The second set (the only remaining set) consists of the distributions of the sums of squares for the remaining main factor C and its interactions with the effects in the first set. Any two distributions, not belonging to the same set are independent, whereas, the distributions in the same set are mutually dependent. (Received May 10, 1957)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

During 1957–58 T. W. Anderson will be a Fellow at the Center for Advanced Study in the Behavioral Sciences in Stanford, California.

Dr. Robert S. Aries is now Chairman of the Board of Aries Associates, whose offices for general consultation were recently transferred to 77 South Street, Stamford, Connecticut.

John Bailey has recently joined the staff of the Waltham Laboratories of Sylvania Electric Products, Inc., as an Engineer in their Applied Engineering Department.

Colin R. Blyth, on leave from the University of Illinois, will be at Stanford University for the academic year of 1957–58.

John V. Breakwell has taken a position as Staff Scientist with the Lockheed Missile Systems Division in Palo Alto.

D. M. Brown is now studying for the degree of Ph.D. in statistics at Princeton University on a RAND Corporation Fellowship.