ABSTRACTS OF PAPERS

(Abstracts of papers to be presented at the Central Regional meeting, Columbus, Ohio, March 23–25, 1967. Additional abstracts will appear in future issues.)

1. Interaction in multi-dimensional contingency tables (preliminary report). S. Kullback and H. H. Ku, George Washington University and the National Bureau of Standards.

The problem of interaction in multi-dimensional contingency tables is investigated from the viewpoint of information theory (Kullback, S., *Information Theory and Statistics*, (1959)). The hypothesis of no rth-order interaction is defined in the sense of an hypothesis of "generalized" independence of classifications with fixed rth-order marginal restraints. For a three-way table, with given cell probabilities π_{ijk}, it is shown that the minimum discrimination information for a contingency table with marginals $p_{ij.}$, $p_{.jk}$, and $p_{i.k}$ is given by the set of cell probabilities $p_{ijk}^* = a_{ij}b_{jk}c_{ik}\pi_{ijk}$ where a_{ij}, b_{jk}, and c_{ik} are functions of the given marginal probabilities, that is, $\ln(p_{ijk}^*/\pi_{ijk}) = \ln a_{ij} + \ln b_{jk} + \ln c_{ik}$, representing no second-order interaction. The minimum discrimination information statistic, asymptotically distributed as χ^2 with appropriate degrees of freedom, is $2\sum_{ijk} f_{ijk} \ln f_{ijk} - 2\sum_{ijk} f_{ijk}^* \ln f_{ijk}^* \geqq 0$ where f_{ijk} are the observed cell frequencies and f_{ijk}^* are the "no interaction" cell frequencies uniquely determined by a simple convergent iteration process of the marginals on π_{ijk}. For lower order marginal restraints, the usual independence hypotheses are generated. It is shown that the set p_{ijk}^* satisfies definitions of no second order interaction in a $2 \times 2 \times 2$ table given by Bartlett (*J. Roy. Statist. Soc. Suppl.* **2** (1935) 248–252) and no interaction in a $r \times s \times t$ table by Roy and Kastenbaum (*Ann. Math. Statist.* **27** (1956) 749–757), and is also related to that given by Good (*Ann. Math. Statist.* **34** (1963) 911–934). (Received 3 October 1966.)

2. Asymptotic distribution of the largest components in a sum of independent random variables. Allan H. Marcus, Case Institute of Technology.

Let $X_1, X_2, \cdots$ be independent positive random variables with a common distribution in the domain of attraction of a positive stable law with index a, $0 < a < 1$. Let X_{nj} be the jth largest of $X_1, \cdots, X_n$ and let $S_n = \sum_{i=1}^n X_i$. Let $Y_{nj} = X_{nj}/X_{n1}$ and $W_{nk} = (S_n - \sum_{i=1}^k X_{ni})/X_{n1}$. The largest component X_{n1} forms an appreciable proportion of the whole sum S_n; indeed, S_n is essentially dominated by a finite number k of its largest components. The asymptotic $(n \to \infty)$ characteristic function of the distribution of $W_{n1} + 1 = S_n/X_{n1}$ was derived by Darling (*Trans. Amer. Math Soc.* **73** (1952) 95–107). In the same way we derive the asymptotic characteristic function of the joint distribution of W_{nk}, $Y_{n2}, \cdots, Y_{nk}$. The asymptotic probability density function of $Y_{n2}, \cdots, Y_{nk}$ is derived explicitly and has a very simple form. Explicit formulas are given for some of the moments of W_{n1}, W_{n2}, and Y_{n2}. (Received 21 November 1966.)

(Abstracts of papers not connected with any meeting of the Institute.)

1. Asymptotically optimal statistics in some models with increasing failure rate averages (IFRA). Kjell Doksum, University of California, Berkeley.

Let F and G be defined by $F(t) = H(\gamma t)$ and $G(t) = H(\theta t)$ where H is unknown and $H(0) = 0$. For testing the equality of the means of F and G in the two-sample problem; it is shown that the Savage (*Ann. Math. Statist.* **27** (1956) 590–615) statistics maximized the minimum power over IFRA (or IFR) distributions asymptotically. Asymptotic uniqueness holds only in a class of rank tests. The results are extended to censored samples, the problem of estimating the ratio of the means, and the k-sample problem. (Received 4 October 1966.)

2. The problem of the two nearest neighbors. BRADLEY EFRON, Stanford University.

Suppose that $X_1, X_2, \cdots, X_n$ are independent and identically distributed random variables in d-dimensional Euclidean space (or more generally on a d-dimensional Riemannian manifold Q), whose common distribution is given by a bounded density function $f(x)$ with respect to Lebesgue measure. We are interested in the limiting distribution of the smallest distance between any two of the points, say $M = \min_{i \neq j} \|X_i - X_j\|$, as n gets large. Under mild regularity conditions on Q, we have the following:

THEOREM. $\lim_{n\to\infty} P(M > x/n^{2/d}) = \exp[-cx^d]$ *where* $c = (\pi^{d/2}/2\Gamma(\frac{1}{2}d + 1))\int_Q f^2(x)\, dx$. (That is, M is asymptotically Weibull in distribution.) Abramson, (*Ann. Math. Statist.* **37** 1421) has proved a stronger version of this result for $d = 1$. Note that the uniform distribution $f(x)$ = constant on Q asymptotically maximizes the minimum distance in a stochastic sense. (Received 2 November 1966.)

3. The problem of changing means (preliminary report). CHANDAN K. MUSTAFI, Columbia University.

We consider n observations whose means are changing randomly according to the model suggested by Chernoff and Zacks (*Ann. Math. Statist.* **35** (1964) 999–1018). Under the assumption that the probabilities of change remain constant over time (i) we derive a sequence of estimates t_n of the current mean which are unbiased and whose variances are uniformly bounded for each finite n. Asymptotically t_n converges almost surely to an estimate t whose variance can exceed at most by an arbitrarily small amount the variance of any other linear unbiased estimate based on a finite number of recent observations. (ii) An empirical Bayes procedure has been developed to estimate the location of all such pairs of points between which changes have taken place. (Received 14 November 1966.)

4. A likelihood ratio test paradox. B. V. SHAH, Research Triangle Institute.

A comparison of likelihood ratio tests for testing the mean of a normal population from the sample of size N is made. The test T_1 is the likelihood ratio test for testing a simple hypothesis $H_0 : \mu = \mu_0, \sigma = \sigma_0$ against the composite null hypothesis $H_1 : \mu, \sigma$ unspecified. The test T_2 is the well known t-test, which tests the null hypothesis $H_0^*: \mu = \mu_0, \sigma$ unspecified against the same H_1. It is shown that for every given observation $x_1, x_2, \cdots, x_n$, there exists a confidence level α_0 and a σ_0 such that the test T_2 will reject H_0^* at the confidence level α_0, but the test T_1 will accept the hypothesis H_0. This is contrary to the basic principles of logic. A new definition of the support provided by an experimental outcome to the statistical hypothesis H_0 against the alternative H_1 is given. The application of the definition of the support to the theory of testing of hypotheses and the interval estimation is given. (Received 20 November 1966.)

5. A curvilinear ranking test. EDWARD WALTER and REINHARD ROSSNER, Universitat Freiburg.

To each x_i of n fixed values $x_1 < \cdots < x_n$ an observation y_i is given. The following curvilinear rank correlation coefficient r_c is proposed: Let $R_1, \cdots, R_n$ be the usual ranking of the y_i's, and $R_1^*, \cdots, R_n^*$ defined as 1, 4, 5, $\cdots$, n, $\cdots$, 7, 6, 3, 2 (a similar ranking is given by Siegel and Tukey, *J. Amer. Statist. Assoc.* **55** (1960) 429–445), then $r_c = 1 - 6(n(n^2 - 1))^{-1} \sum_{i=1}^{n} (R_i^* - R_i)^2$. Under H_0 (equal probability for every permutation) r_c is distributed as Spearman's rank correlation coefficient r_S. The correlation between r_S and r_c is $r = 3(n^2 - 1)^{-1}$ for $n \equiv 0 \bmod 2$ and $= 3(n(n \pm 1))^{-1}$ for $n \equiv \pm 1 \bmod 4$. This allows us to use both statistics simultaneously. (Received 18 October 1966.)

6. Tables of inverse Gaussian probabilities. M. T. WASAN and L. K. ROY, Queen's University, Kingston.

Given the family of inverse Gaussian probability density functions

$$f(x) = \lambda^{\frac{1}{2}}(2\pi x^3)^{-\frac{1}{2}}e\lambda(x - \mu)^2(2\mu^2 x)^{-1}$$

for positive x with parameters $\mu > 0$ and $\lambda > 0$, a table of the values of $P[X \leqq A]$ is presented for $\lambda = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 2, \cdots, 10, 16, 24, 32$ with μ held fixed at 5. If an inverse Gaussian variate Y has a distribution with parameters μ_1 and λ_1 where $\mu_1 \neq 5$, then the transformation $X = 5Y/\mu_1$ produces an inverse Gaussian distribution with $\mu = 5$ and $\lambda = 5\lambda_1/\mu_1$, and if this λ is not one for which a table has been presented, a method of interpolation using the existing tables is demonstrated which will find $P[X \leqq A \mid \lambda]$ for any fixed A. It is also shown that if λ is greater than 32 or less than $\frac{1}{4}$, existing tables of the normal and gamma distributions respectively may be used instead. (Received 28 November 1966.)