ABSTRACTS

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Santa Monica meeting of the Institute, June 15 and 16, 1951)

1. **First Passage in Random Walks.** T. E. HARRIS, The Rand Corporation.

Consider a random walk on the integers with transition probabilities p_r for $r \to r + 1$ and $q_r = 1 - p_r$ for $r \to r - 1$. Let N_{ij} be the first passage time from i to j. Explicit expressions are given for $P(N_{ij} < \infty)$ and $E(N_{ij} \mid N_{ij} < \infty)$, the latter expression sometimes being finite in transient chains. The second moment of N_{ij} can also be found. Sufficient conditions can be given for the limiting distribution of N_{0j}, as $j \to \infty$, to be exponential. The conditions include some walks with infinite mean recurrence times.

2. **Distinct Hypotheses and Convex Sets.** LUCIEN M. LE CAM, University of California, Berkeley.

Given a measure μ on a set $\mathfrak{X}$, the densities of probability with respect to μ on $\mathfrak{X}$ are considered as points in a Banach space B. A test is a point in the unit sphere of the conjugate Banach space. The distinctness of two hypotheses can be discussed using their convex hulls in B. Theorems by Mazur, Klee, etc., give necessary and sufficient conditions for distinctness. It is shown that the theorem by A. Berger and A. Wald corresponds to the case in which the closed convex hulls are disjoint. The result can thus be slightly extended.

3. **Uniform Convergence of Random Functions with Applications to Statistics.** HERMAN RUBIN, Stanford University.

Let $X_1, \cdots, X_n, \cdots$ be a sequence of independent and identically distributed variables with values in an arbitrary space X. Let T be a compact topological space, and let $f(t, x)$ be a complex-valued function on $T + X$, measurable in x for each $t \varepsilon T$. Let P be the common distribution of the X_i. Then if there is an integrable g such that $|f(t, x)| \leq g(x)$ for all $t \varepsilon T$ and $x \varepsilon X$, and if there is a sequence S_i of measurable sets such that $P(X - \cup_{i=1}^{\infty} S_i) = 0$ and for each i, $f(t, x)$ is equicontinuous in t for $x \varepsilon S_i$, then with probability one, $(1/n) \Sigma_{k=1}^{n} f(t, X_k) \to \int f(t, x)\, dP(x)$ uniformly for $t \varepsilon T$, and the limit of the function is continuous. Since $f(t, x) = e^{itx}$ satisfies the conditions of the theorem, the sample characteristic function converges to the population characteristic function uniformly with probability one in any bounded interval. Log $L(x \mid \theta) = f(x, \theta)$ satisfies the conditions of the theorem for many distributions, including the multivariate normal, Poisson, Cauchy, χ^2, and double exponential, and hence the almost certain convergence of maximum likelihood estimates to the true values if the parameter is restricted to a compact set is established for those cases. More difficult estimation procedures can also be shown to be consistent by this method.

4. **A Sequential Test for Linear Hypotheses.** PAUL G. HOEL, University of California, Los Angeles.

A sequential test for the general linear hypothesis is obtained by employing methods similar to those introduced by Wald in deriving his sequential t test. Optimum properties of the test are studied. An explicit expression for p_{1m}/p_{0m} is obtained, the evaluation of which requires incomplete Gamma function tables.

5. A Two Sample Test. Frank J. Massey, Jr., University of Oregon.

Consider two (or more) samples arranged together in order of size and let $z_1 < z_2 < \cdots < z_k$ be the α_1th observation, the α_2th observation, etc., in order of size in the combined samples. Let n_{ij} be the number of observations in the ith sample which are greater than z_{j-1} and less than or equal to z_j. Then the joint distribution of the n_{ij} is that of a two-variable contingency table with fixed marginal totals. This is an extension of the result of A. M. Mood and G. W. Brown for $k = 1$ (A. M. Mood, *Introduction to the Theory of Statistics*, 1950, p. 398). (This work was sponsored by the Office of Naval Research.)

6. Some Slippage Problems for the Normal Distribution. (Preliminary Report.) Edward Paulson, University of Washington.

Let $\{X_{i\alpha}\}$ $(\alpha = 1, 2, \cdots, n)$ be a sample of n independent observations from Π_i with probability distribution $N(m_i\sigma_i)$ $(i = 1, 2, \cdots, K)$, and let $\bar{X}_i = \sum_{\alpha=1}^{n} X_{i\alpha}/n$, $s_i^2 = \sum_{\alpha=1}^{n} (X_{i\alpha} - \bar{X}_i)^2/(n-1)$. First consider the problem of choosing one of the $K + 1$ hypotheses $H_0, H_1, \cdots, H_K$, when H_0 is the hypothesis the K means are all equal, while $H_j (j = 1, \cdots, K)$ is the hypothesis the means are not all equal and $m_j = \max_i \{m_i\}$, when it is known a priori that the K standard deviations have a common known value σ^*. It is shown that of all decision procedures with a fixed probability α of making the wrong decision when H_0 is true, the procedure [select H_j if $\bar{X}_j = \max_i \{\bar{X}_i\}$ and $\bar{X}_j - \sum_{i=1}^{k} X_i/K_i > \lambda_\alpha \sigma^*$, otherwise select $H_0(\lambda_\alpha$ is a constant depending on $\alpha)$] maximizes the probability of making the correct decision when one of the means has slipped to the right, provided we restrict ourselves to decision procedures which are symmetric and invariant under a change of location parameter. An analogous slippage problem is considered with respect to the variances, and it is shown that the Cochran-Solomon-Eisenhart procedure [select Π_j if $s_j^2 = \max_i \{s_i^2\}$ and $s_j/\sum_{i=1}^{n} s_i^2 \geqq \lambda'_\alpha$, otherwise select H_0] is in a corresponding sense the 'best' possible.

7. Minimax Procedures for Two-valued Decision Problems When the Size of the Sample Is Fixed. S. G. Allen, Jr., Stanford University.

The problem considered is a minimax statistical decision procedure for choosing between two alternative actions, A_1 and A_2, after taking n independent observations on a random variable x. The probability density $p(x, \theta)$ of x is known except for the value of a real, one-dimensional parameter θ. The loss if decision A_1 is taken is zero for $\theta \leqq \theta_0$, and $w_1(\theta) \geqq 0$ otherwise; the loss if decision A_2 is taken is $w_2(\theta) \geqq 0$ for $\theta < \theta_0$ and zero otherwise. It is then shown that the likelihood ratio test is a minimax procedure if and only if there exists a triple (c, θ_1, θ_2), such that

$$\max_\theta \{w_1(\theta) Pr[\lambda(\theta_1, \theta_2) \leqq c \mid \theta]\} = \max_\theta \{w_2(\theta) Pr[\lambda(\theta_1, \theta_2) > c \mid \theta]\},$$

where $\theta_1 \leqq \theta_0 \leqq \theta_2$ and $\lambda(\theta_1, \theta_2) = \Pi_{i=1}^{n} [p(x_i, \theta_2)/p(x_i, \theta_1)]$. For the exponential class of distributions, sufficient conditions on the loss functions are found so that the above criterion is met. Finally, modifications of the results for the case of discrete exponential distributions are discussed.

8. The Asymptotic Properties of Bayes Estimates. R. C. Davis, U. S. Naval Ordnance Test Station, China Lake.

Let $X_1, X_2, \cdots$, etc., be independently and identically distributed chance variables possessing the common cumulative distribution function $F(x, \theta)$, in which $F(x, \theta)$ admits an elementary probability law $f(x, \theta)$ depending upon an unknown parameter θ. Let also $\lambda(\theta)$ be an assumed a priori distribution for θ. θ may assume values in Ω, a closed subset of the real axis. Denote by $E(\theta \mid x_1, x_2, \cdots, x_n; \lambda)$ the conditional expected value of the a posteriori distribution of θ given a sample $x_1, x_2, \cdots, x_n$ of values of $X_1, X_2, \cdots, X_n$. This is termed the Bayes estimate of θ. It has been surmised for several years that under some set of regularity conditions every Bayes estimate is BAN (best asymptotically normal). It is proved that this surmise is correct. It turns out that for each property, i.e., consistency, asymptotic normality, and asymptotic efficiency, the property of the Bayes estimate can be established under regularity conditions quite similar to those assumed in proving the respective property of the maximum likelihood estimate as treated by Doob and Cramér. Moreover under conditions quite similar to those assumed by Wald to establish the almost certain convergence of the maximum likelihood estimate, the Bayes estimate may be shown to converge almost certainly to the true value of the unknown parameter θ.

9. Problems of Estimation and Hypothesis Testing in Connection with Birth-and-Death Stochastic Processes. Eric R. Immel, University of California, Los Angeles.

Methods of estimation and of hypothesis testing are developed for the two parameters of the distribution of the integer-valued chance variable $x(t)$ associated with the time-homogeneous birth-and-death process of the Markov type. For a certain class of processes maximum likelihood estimates are obtained which are unbiased estimates of the ratio of the parameters, and the estimates are efficient in the Cramér sense. More general estimates are obtained which are asymptotically optimum. The estimates yield confidence limits for the ratio of the parameters and for the parameters. The methods used yield approximate estimates for all processes of the type considered. A large-sample method of discriminating between two processes is discussed, as well as nonsequential and sequential methods for testing simple hypotheses against certain sets of alternatives.

10. The Identifiability of n-dimensional Linear Structures. T. A. Jeeves, University of California, Berkeley.

Notation: Italic letters denote n-dimensional random vectors. Boldface letters denote $n \times n$ matrices of sure numbers. Consider the random vectors $Y = S + U$ with S and U independent. Vector S is assumed to satisfy the condition $S = X\mathbf{B}$. *Definition:* $\mathbf{B}$ is said to be identifiable if $S = X^*\mathbf{B}^*$ implies that $\mathbf{B}$ and $\mathbf{B}^*$ are row equivalent If U has a multinormal distribution then $\mathbf{B}$ is unidentifiable if and only if $S = Z\mathbf{C} + V$ with rank $\mathbf{C} <$ rank $\mathbf{B}$, where Z and V are independent and V has a multinormal distribution.

11. A Probabilistic Study of Runs in Egg Production. (Preliminary Report.) Dorothy Cruden Lowry, University of California, Berkeley.

This is a report on an analysis of the production records of some 450 hens of the same age and subject to the same environment The random variable studied, X_i, is assigned the value 1 if the hen laid an egg on the ith day of a given period and 0 otherwise, and the questions investigated are as follows. Are the variates X_i identically and independently distributed and, if not, what sort of stochastic model can be devised for which the distribution agrees reasonably well with the observed production record? A test based on the total number of runs in a 30-day period results in failure to reject the hypothesis of independence, as does a test based on the serial correlation coefficient computed for families of hens. However the test based on the total number of runs in a 60-day period results in rejection of the hypothesis of independence in more cases than the chosen probability level would indicate. As several simple stochastic models failed to fit the data, more complex models are to be tried.

12. Estimation in a Set of Distribution Functions of the Same Type. ROBERT R. PUTZ, University of California, Berkeley.

Two distribution functions $\Phi^{(1)}$ and $\Phi^{(2)}$ are said to be of the *same type* in case there exist constants a and b such that $\Phi^{(1)}(x) = \Phi^{(2)}(ax + b)$, with $a > 0$. A family $\mathfrak{F}$ of identical-type distribution functions $\Phi_{\mu,\sigma}$, the mean μ and standard deviation σ being components of a random vector $\theta = (\mu, \sigma)$, is considered. For a random experiment in which there is observed a distribution function selected from $\mathfrak{F}$ in accordance with the distribution of θ, knowledge of the latter distribution may be used to predict the observed distribution function $\Phi_{\mu,\sigma}$. For any estimated distribution function $\Phi_{\hat{\mu},\hat{\sigma}}$, the error $\epsilon(x) = \Phi_{\hat{\mu},\hat{\sigma}}(x) - \Phi_{\mu,\sigma}(x)$ is considered in relation to its dependence on the relative parameter errors $\Delta_1 = (\hat{\mu} - \mu)/\mu$ and $\Delta_2 = (\hat{\sigma} - \sigma)/\sigma$. In the Gaussian case explicit expressions are obtained for $\max_x | \epsilon(x) |$ and for the maximizing argument value $\hat{x}$. Here in the particular case $\Delta_2 = 0$ these become $(\hat{x} - \mu)/\sigma = \Delta_1/2\delta$, where $\delta = \sigma/\mu$, and for small Δ_1, $\max_x | \epsilon(x) | = | \Delta_1 | /\delta\sqrt{2\pi} + O(\Delta_1^5)$, while in the case $\Delta_1 = 0$, they reduce for small Δ_2 to $| \hat{x} - \mu | /\sigma = 1 + \frac{1}{2}\Delta_2 + O(\Delta_2^2)$ and $\max_x | \epsilon(x) | = \gamma | \Delta_2 | + O(\Delta_2^2)$, where $\gamma = 0.242 \cdots$. The difference $e(x) = \Phi_{\hat{\mu},\hat{\sigma}}(x) - F_{\mu,\sigma}^{(n)}(x)$, where $F_{\mu,\sigma}^{(n)}$ is the distribution function of a sample of size n drawn from the distribution $\Phi_{\mu,\sigma}$, is considered.

13. On the Practical Application of Confidence Intervals for the Mean of a Normal Population. ARTHUR SHAPIRO, Stanford University.

Let $X_1, \cdots, X_n$ be independent normal variables having the same unknown mean ξ and the same, also unknown, variance σ^2. The shortest unbiased confidence interval for ξ, corresponding to the confidence coefficient α, is give by (1), $\bar{X} - st_\alpha/\sqrt{n-1} \leqq \xi \leqq \bar{X} + st_\alpha/\sqrt{n-1}$, where $\bar{X}, s, t_\alpha$ have the usual meaning. However, if the length of the confidence interval exceeds a certain limit, the practical value of the result of estimation is nil. In all practical cases, formula (1) is used for estimating ξ only if $st_\alpha/\sqrt{n-1} \leqq \tau^*$. As a result, the probability that statement (1) regarding ξ will be correct is not equal to the confidence coefficient α. For fixed τ^*, σ, α, denote by $P(\tau^*, \sigma, \alpha)$ the probability that (i), $st_\alpha/\sqrt{n-1} < \tau^*$, and (ii), (1) will be correct. Then $P(\tau^*, \sigma, \alpha)$ can be shown to be a monotonic function of τ^*/σ. In some cases, the experimenter may be sure that his $\sigma \leqq \sigma_0$. In these cases it is important to be able to answer the question: what is the smallest value $n(\theta_0, \alpha, \alpha')$ such that $n \geqq n(\theta_0, \alpha, \alpha')$ implies $P(\tau^*, \sigma, \alpha) \geqq \alpha'$ where $\theta_0 = \tau^*/\sigma_0$, $\alpha' < \alpha$. The purpose of this work is to tabulate $n(\theta_0, \alpha, \alpha')$ for several combinations of α and α'. For example, if $\alpha = .95$ and $\alpha' = .90$ then for $\tau^*/\sigma_0 = 1.07$, $n(\theta_0, \alpha, \alpha') = 9$ and for $\tau^*/\sigma_0 = .200$, $n(\theta_0, \alpha, \alpha') = 121$.

14. Some Results Concerning Random Numbers of Random Variables. (Preliminary Report.) ROBERT F. TATE, University of California, Berkeley.

Two theorems concerning random numbers of random variables are stated and proved. The first theorem asserts that if X_i $(i = 1, 2, \cdots, \mu)$ and Y_j $(j = 1, 2, \cdots, \nu)$ are sets of observations each of which consists of independent, equi-distributed random variables, where μ and ν are positive, integer-valued random variables, then, even assuming the X_i independent of the Y_j, we have $\sum_{i=1}^{\mu} X_i$ and $\sum_{j=1}^{\nu} Y_j$ independent if and only if μ and ν are independent random variables. The generalization to k sets of observations is immediate. The second theorem constitutes a generalization of the Lindeberg-Lévy form of the Central Limit Theorem for k-dimensional random vectors. The random vectors $(X_n^{(1)}, \cdots, X_n^{(k)})$ are generalized to $(X_{\mu_{1n}}^{(1)}, \cdots, X_{\mu_{kn}}^{(k)})$, where μ_{in} are random variables which take on the values $0, 1, 2, \cdots, n$, and whose probability distributions degenerate at infinity. As $n \rightarrow \infty$ the Central Limit Theorem will now hold, with the same limiting normal law as before. The proof follows from considerations involving the Continuity Theorem for characteristic functions.

15. Inspection Plans Which Improve Lot Quality. ZIVIA S. WURTELE, University of California, Berkeley.

Assume that all defective items found during inspection of the lot are replaced by non-defective items and that when inspection is terminated the entire lot is accepted and used. Let the cost be linear in the number of items inspected, of defective items not removed, and of defective items replaced. Consider the Poisson limiting case; let t be the expected number of defective items in the lot. For any a priori distribution of t the Bayes plan is characterized by a set of stopping points $\{(d, x_d)\}$, $d = 0, \cdots, d_0$, where d is the number of defective items found in a proportion x_d of the lot and where $x_{d+1} > x_d$. Inspection continues until a stopping point is reached or until more than d_0 defective items are found, in which case the entire lot is inspected. Methods are obtained for finding Bayes procedures and for calculating the risk associated with any plan of this type. For any such plan there exists an a priori distribution of t, other than a trivial one, with respect to which this plan is Bayes.

16. Nonparametric Discrimination, II. Small-Sample Performance in Normal Populations. EVELYN FIX AND J. L. HODGES, JR., University of California, Berkeley.

Consider the usual discrimination problem of assigning an individual I to one of two populations, from each of which a sample is available, on the basis of p measurements made on all individuals concerned. By defining a distance function in the p-dimensional space, and observing the population of origin of those sampled individuals "near" I, one may obtain a class of (sequential and nonsequential) nonparametric classification rules whose performance is asymptotically optimum as the sample sizes are increased, regardless of the populations being discriminated. In the present study, numerical results on the probabilities of misclassification are obtained, primarily for small samples from bivariate normal populations.

17. On Certain Estimators for the Parameters of the Distribution of Largest Values. JULIUS LIEBLEIN, National Bureau of Standards.

Estimators for the two parameters of the cdf of largest values, Prob. $\{X \leqq x\} = \exp(-e^{-\alpha(x-u)})$, have been given by E. J. Gumbel and B. F. Kimball, who have studied their asymptotic behavior in infinitely large samples. This paper considers the behavior of estimators of one parameter when the other is known, and attempts to evaluate exactly the bias and efficiency for samples of finite size. This turns out to be possible in only a few situations, and it is found that numerical methods of approximation would be necessary for most cases.

18. A Theorem on the Impossibility of Affine Resolvable Designs. S. S. SHRIKHANDE, Nagpur College of Science, India.

The following theorem is proved. *Suppose a Balanced Incomplete Block Design (B.I.B.D.) with parameters* $v = b = n^2t + n + 1$; $r = k = nt + 1$; $\lambda = t$ *exists, and a B.I.B.D. with parameters* $v = b = n(n^2t + n + 1) + 1$; $r = k = n^2t + n + 1$; $\lambda = nt + 1$ *does not exist, then an Affine Resolvable B.I.B.D.* (R. C. BOSE, "A note on the resolvability of balanced incomplete block designs," *Sankhyā* Vol. 6(1942), pp. 105–110) *with parameters* $v = nk = n^2(n - 1)t + n^2$; $b = nr = n(n^2t + n + 1)$; $\lambda = nt + 1$ *does not exist.*

The proof depends upon the fact that the second design can be constructed if the other two exist. Making use of a result given by Schützenberger ("A non-existence theorem for an infinite family of symmetrical block designs," *Annals of Eugenics*, Vol. 14 (1949), pp. 286–287) and others it is proved that an Affine Resolvable B.I.B.D. with parameters $v = 63$, $b = 93$, $r = 31$, $k = 21$, $\lambda = 10$ does not exist.

19. The Nonexistence of Difference Sets for Group Designs. S. S. SHRIKHANDE, Nagpur College of Science, India.

The following theorem is proved. *Let* $v = mn$, *where* n *is a prime congruent to* 3 (*mod.* 4). *Let nonnegative integers* λ_1 *and* λ_2 *satisfy the relation* $k(k - 1) = (m - 1)\lambda_1 + (n - 1)m\lambda_2$. *Define* $\theta = k + \lambda_1(m - 1) - \lambda_2 m$ *and let* ϕ *be a prime factor of* θ *occurring in it to an odd degree. Then if* $(-n/\phi) = -1$, *where* (q/p) *stands for the Legendre symbol, there does not exist a difference set of* k *integers which gives rise to a group design* (K. R. NAIR AND C. R. RAO, "A note on partially balanced incomplete block designs," *Science and Culture*, Vol. 7 (1942), pp. 615–616) *with parameters* $v = b = mn$, $r = k$, λ_1, λ_2, *where there are* n *groups of* m *treatments each, in* b *blocks of size* k, *such that each pair of varieties from the same group occurs in* λ_1 *blocks while each pair of varieties coming from different groups occurs in* λ_2 *blocks.* This generalizes a result of Chowla ("On difference sets," *Proc. Nat. Acad. Sci.*, Vol. 35 (1949), pp. 92–94), the proof following along the lines of his paper.

20. Concerning Large-Sample Tests and Confidence Intervals for Mortality Rates. JOHN E. WALSH, Bureau of the Census.

This is an extension of the paper "Large-sample tests and confidence intervals for mortality rates" which appeared in the June, 1950 issue of the *Journal of the American Statistical Association*. The results of this other article are placed on an axiomatic basis and the validity of these axioms is discussed. The basic underlying concepts are explained and some numerical examples of applications are worked out. Also additional significance tests and confidence intervals are presented.

NEWS AND NOTICES

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

Personal Items

The Shewhart Medal for outstanding service and leadership in the field of quality control was awarded to Dr. Martin A. Brumbaugh, Director of Statistics at Bristol Laboratories, Inc., Syracuse, New York, at the convention of the American Society for Quality Control in Cleveland May 23 and 24. Dr. Brumbaugh is a founder of the American Society for Quality Control and at this time is first vice-president of the group.

Dr. W. Edwards Deming visited Japan in July, 1950, and delivered a number of lectures and conducted two 8-day courses in quality control in Tokyo and Fukuoka. Considerable interest in statistical methods and quality control among Japanese engineers and industrialists has arisen, largely as result of this visit.

Mr. John C. Hintermaier, formerly Superintendent of Development, Standards Testing Laboratories, Cluett Peabody & Co., Research Division, Troy, New York, is now with the Textile Materials Engineering Laboratory, Philadelphia Quartermaster Depot, Philadelphia 45, Pennsylvania.

Mr. Roy R. Kuebler, Jr. is returning to his position of Associate Professor of Mathematics at Dickinson College, Carlisle, Pennsylvania, after a year's leave.

Mr. Marvin Masel has resigned his position as Engineering Statistician with the