The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 687–708

ABSTRACTS OF PAPERS

(Abstract of a paper presented at the Western Regional meeting, Missoula, Montana, June 15–17, 1967. Additional abstracts appeared in earlier issues.)

19. Methods for statistical analysis and testing in practical metric spaces.

J. MacQueen, Western Management Science Institute, University of California, Los Angeles. (Invited)

Let $M_1, M_2, \cdots M_N$ be compact metric spaces with distances $\rho_1, \rho_2, \cdots \rho_N$ respectively. Let M be the product space with distance $\rho = \sum_i \alpha_i \rho_i$, $\alpha_i \geqq 0$. For a probability distribution P on M, a centroid of order r is a point in the non-empty class C_r of points x such that $\int \rho^r(x, y)\, dP(y) = \inf_z \int \rho^r(z, y)\, dP(y)$; the latter quantity itself is termed the variation of order r of P. For a sample $x_1, x_2, \cdots, x_n$, $x_1 \varepsilon M$, a sample centroid minimizes $\sum_i \rho^r(z, x_i)/n$; the minimum value is the sample variation. Methods for statistical analysis based on these simple concepts are described. They are applicable to 'practical' metric spaces, which is to say, spaces where the necessary operations can be carried out on a computer. In addition to Euclidean space, three other spaces have been treated and a computer program developed which permits various kinds of statistical analysis in any of these spaces, or in products of various combinations of these spaces. For example, statistical tests with prescribed type I error can be performed to test whether or not several random variables with values in such spaces are independent or to test whether or not two such random variables have the same distribution. These tests are accomplished by sampling from the appropriate randomization distribution. The special domain of application of these methods is in connection with complex data problems, where the natural descriptions involve a variety of different kinds of information, e.g., medical case records or cross cultural surveys. Large samples are easily handled. The methods make use of an efficient algorithm for finding a partition of a sample of points in M which has low within-class variation. This algorithm can also be used for clustering, automatic file construction, and certain types of pattern recognition. (Received 4 December 1967.)

(Abstracts of papers to be presented at the Annual meeting, Washington, D. C., December 27–30, 1967. Additional abstracts appeared in the June, August, October, and December issues.)

89. $O(c)$-Bayes procedures in sequential design of experiments.

Gary Lorden, Northwestern University.

A sequential design procedure is $O(c)$-Bayes for a given prior distribution if its integrated risk exceeds the Bayes risk by $O(c)$ as c, the cost per observation, approaches zero. Kiefer and Sacks [*Ann. Math. Statist.* **34** 705–750] considered a very general formulation of the sequential design problem and gave simple procedures using non-randomized choices of designs whose integrated risk exceeds the Bayes risk (of order $c|\log c|$) by $o(c|\log c|)$. Their procedures, like all procedures requiring at most one switch of designs are, however, not $O(c)$-Bayes. The present investigation is restricted to a class of problems involving a finite number of states of nature. Simple $O(c)$-Bayes procedures are given, using only non-randomized choices of designs. They are similar to the procedures of Chernoff [*Ann. Math. Statist.* **30** 755–770], although the latter are $O(c)$-Bayes only for problems in which they specify non-randomized design choices. (Received 6 November 1967.)