Serinko, Regis J. (1-PAS); Babu, Gutti Jogesh (1-PAS)

Weak limit theorems for univariate $k$-mean clustering under a nonregular condition.


For an integer $k$ and a given distribution $F$ with finite second moment, the $k$-cluster center vector is defined to be the vector $a = (a_1, \ldots, a_k)'$ which minimizes $W(a) = \int \min_{i \leq k} (x - a_i)^2 \, dF(x)$. Using the cluster center $a$, one may partition the space into $k$ groups or clusters $C_j$ according to the smallest distance to the components of the vector $a$.

Then, the $i$th component of the split point vector is defined by $p_i = \sum_{j=1}^k \int_{C_j} dF(x), \ i = 1, \ldots, k-1$. Under regularity conditions, e.g., the Hessian matrix of $W(a)$ is nonsingular, it is known that the sample cluster center and the sample split point vector are asymptotically normal with a rate $1/\sqrt{n}$. In this paper, the limiting distributions of these statistics are investigated in the irregular cases. It is shown that the limiting distribution is not normal and that the rate is only $n^{-1/4}$.