Journal of the Korean Statistical Society (2000), 29: 2, pp 247–260

On a Modified k-Spatial Medians Clustering [†]

Myoungshic Jhun[1] and Seohoon Jin[1]

ABSTRACT

This paper is concerned with a modification of the k-spatial medians clustering. To find a suitable number of clusters, the number k of clusters is incorporated into the k-spatial medians clustering criterion through a weight function. Proposed method for the choice of the weight function offers a reasonable number of clusters. Some theoretical properties of the method are investigated along with some examples.

Keywords: Cluster analysis; k-spatial medians clustering; Number of clusters.

1. INTRODUCTION

Suppose that independent multi-dimensional observations $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ are sampled from a distribution F on R^p $(p > 1)$. Let F_n be an empirical distribution function of the observations. It is desired to partition these observations into k clusters so that the observations within the same cluster are close in some sense and observations in different clusters are distant. We may use a procedure consists of

(i) finding $\underline{a}_n = (\underline{a}_{n1}, \underline{a}_{n2}, \cdots, \underline{a}_{nk})$ minimizing

$$\frac{1}{n}\sum_{i=1}^{n} \min_{1 \le j \le k} \eta(||\underline{x}_i - \underline{a}_j||)$$

where $||\cdot||$ is the usual Euclidean norm and $\eta(\cdot)$ is an increasing function, and

(ii) assigning each $\underline{x}_i$ to its nearest cluster center.

[†]This research was supported by Korea Science and Engineering Foundation(No. 981-0105-026-2).

[1]Department of Statistics, Korea University, Seoul 136-701, Korea

Some possible distance functions $\eta(\cdot)$ can be considered. If the distance function is $\eta(x) = x^2$, the procedure is the k-means clustering which minimizes within cluster sum of squares, and each $\underline{a}_{nj}$ is the mean of the observations in its cluster. Properties of the k-means clustering were studied by many authors including Hartigan (1975, 1978) and Pollard (1981, 1982). Even though the k-means clustering is one of the most widely used procedure, it is very much influenced by outliers, distant observations and data structure. If $\eta(x) = |x|$, the procedure minimizes within cluster sum of absolute deviations and each $\underline{a}_{nj}$ is the spatial median of the observations in its cluster. Statistical uses of the spatial median were discussed by Brown (1983). Since the spatial median is robust against outliers, we expect that the adoption of the spatial median into cluster analysis makes reasonable partition comparing with the k-means clustering. As an alternative partitioning method for decreasing effect of outliers and data structure, the k-spatial medians clustering was considered in some studies including Spath (1980), Butler (1986) and Jhun (1986).

For the procedures above, the number k of clusters should be given in advance. However, in most real life clustering situations, a data analyst is faced with the problem of choosing an appropriate number of clusters in the final solution. Accordingly, numerous procedures for determination of the number of clusters have been suggested and many studies have been carried out to compare the procedures. For example, Milligan and Cooper (1985) conducted a Monte Carlo evaluation of many indices for determining the number of clusters. On the while, Rost (1995) studied feasibility of the k-means clustering procedure together with a weight function, which can be used to obtain a proper number of clusters for a given data set.

In this paper, in the same vein of Rost (1995), a modified k-spatial medians clustering procedure is proposed. By using a weight function for the proposed procedure, reduction of within cluster sum of absolute deviations resulting from dividing the data set into k clusters will be compromised with losses of information from summarizing each observation as representative characteristics of its cluster. As a result, the number k of clusters depends on a given data set via the sample size and data structure. However, it is crucial to have a proper weight function for a satisfactory clustering results. We propose a method for the choice of the weight function and demonstrate its applicability. Asymptotic properties of the modified k-spatial medians clustering are also investigated along with some examples.

2. A MODIFIED k-SPATIAL MEDIANS CLUSTERING

Like other partitioning methods, determination of the number of clusters is a crucial problem for the k-spatial medians clustering. One of the natural way of finding a suitable number k of clusters is to incorporate it into a specified clustering criterion. Then, in the process of the algorithm, the modified criterion is minimized not only with respect to the unknown clusters but also with respect to the unknown cluster number k. This type of approach has been formally investigated by Peck et al. (1989) and Rost (1995). They modified the k-means clustering criterion by introducing a certain penalty factor which increases with k.

Now, we propose a modified k-spatial medians clustering procedure as follows by incorporating a weight function $w(k)$ into the objective function. The procedure consists of

(i) finding the number of clusters k_n and sample k_n-spatial medians $\underline{a}_n = (\underline{a}_{n1}, \underline{a}_{n2}, \cdots, \underline{a}_{nk_n})$ satisfying (2.1),

$$\begin{aligned}\overline{W}(\underline{a}_n, k_n, F_n) &= w(k_n) \int \min_{1 \le j \le k_n} ||\underline{x} - \underline{a}_{nj}|| dF_n \\ &= \min_{k \le K} \left[w(k) \min_{(\underline{a}_1, \cdots, \underline{a}_k)} W(\underline{a}, k, F_n) \right] \end{aligned} \quad (2.1)$$

where $W(\underline{a}, k, F_n) = \int \min_{1 \le j \le k} ||\underline{x} - \underline{a}_j|| dF_n$, the weight function $w(k)$ is a monotone increasing function of k and K is the largest possible number of clusters, and

(ii) assigning each $\underline{x}_i$ to its nearest cluster center.

In the proposed procedure, by multiplying the weight function $w(k_n)$ to the objective function of the k-spatial medians clustering, one can compromise reduction of within cluster sum of absolute deviations resulting from dividing the data set into k groups with losses of information from summarizing each observation as representative characteristics of its cluster. Notice that the number k_n of clusters is a random number depending on the sample size n and data structure. Therefore the length of the sample optimal center vector $\underline{a}_n$ is not fixed, and the modified k-spatial medians clustering offers a data dependent choice of the number of clusters.

Now, let us define optimal number k^* of clusters and optimal k^*-spatial medians $\underline{a}^* = (\underline{a}_1^*, \underline{a}_2^*, \cdots, \underline{a}_{k^*}^*)$ as

$$\begin{aligned} \overline{W}(\underline{a}^*, k^*, F) &= w(k^*) \int \min_{1 \le j \le k^*} ||x - \underline{a}_j^*|| dF \\ &= \min_{k \le K} \left[w(k) \min_{(\underline{a}_1, \cdots, \underline{a}_k)} \int \min_{1 \le j \le k} ||x - \underline{a}_j|| dF \right] \end{aligned} \qquad (2.2)$$

Consistency of the proposed procedure can be shown in the following Theorem 2.1.

Theorem 2.1. *Let $\underline{a}_n = (\underline{a}_{n1}, \cdots, \underline{a}_{nk_n})$ be the vector of optimal centers from the modified k-spatial medians clustering for independent sampling from a distribution F on R^p. Suppose that $\int ||\underline{x}|| dF < \infty$ and that there exist a unique $k^* \in N$ and a unique vector $\underline{a}^* = (\underline{a}_1^*, \underline{a}_2^*, \cdots, \underline{a}_{k^*}^*)$. Then*

(i) $k_n \longrightarrow k^*$ almost surely as $n \to \infty$

(ii) $(\underline{a}_{n1}, \cdots, \underline{a}_{nk_n}) \longrightarrow (\underline{a}_1^*, \cdots, \underline{a}_{k^*}^*)$ almost surely as $n \to \infty$

Proof: See Appendix. □

Now, let us examine Example 2.1, which is similar to Exmaple 2.1 of Gaenssler (1988), to grasp applicability of the modified k-spatial medians clustering.

Example 2.1 Consider a underlying distribution $F \in \{F^{(1)}, F^{(2)}, \cdots, F^{(K)}\}$, where $F^{(1)}$: uniform [0,1], $F^{(2)}$: uniform [0,1] ∪ [2,3], $F^{(3)}$: uniform [0,1] ∪ [2,3] ∪ [4,5] , $\cdots$. If $F = F^{(1)}$ then $W(\underline{a}^*, k, F) = \min_{\underline{a}} \int \min_{1 \le j \le k} ||\underline{x} - \underline{a}_j|| dF = \frac{1}{4k}$. Hence $W(\underline{a}^*, k, F)$ decreases at the rate of $1/k$ with the number k of clusters. In general, for $i \in N$, $W(\underline{a}^* = (a_1^*, a_2^*, \cdots, a_{i \times j}^*), i \times j, F^{(i)}) = \frac{1}{4j}$, $j \in N$ is satisfied. Thus there exists a certain weight function $w(k)$, for example $w(k) = k^{3/2}$, such that the optimal number k^* of clusters and cluster centroids $\underline{a}^*$ are uniquely determined for any $F \in \{F^{(1)}, F^{(2)}, \cdots, F^{(K)}\}$. One can easily check that the optimal number k^* of clusters and cluster centroids $(a_1^*, a_2^*, \cdots, a_{k^*}^*)$ which satisfy (2.2) are computed for each i, $(i = 1, 2, 3)$ as follows by considering the weight function $w(k) = k^{3/2}$.

$$\begin{aligned} F^{(1)} &: k^* = 1, a_1^* = 1/2 \\ F^{(2)} &: k^* = 2, (a_1^*, a_2^*) = (1/2, 5/2) \\ F^{(3)} &: k^* = 3, (a_1^*, a_2^*, a_3^*) = (1/2, 5/2, 9/2) \end{aligned}$$

The modified k-spatial medians clustering releases us from the problem of determining the number of clusters. □

The following Theorem 2.2 gives the limiting distribution of the sample k_n-spatial medians for the proposed procedure.

Theorem 2.2. *Let k_n be the optimal number of clusters and $\underline{a}_n = (\underline{a}_{n1}, \cdots, \underline{a}_{nk_n})$ be the vector of optimal centers from the modified k-spatial medians clustering for independent sampling from a distribution F on R^p. Suppose*

(i) the vector $\underline{a}^*$ that minimizes $\int \phi_{\underline{a},k}(\cdot)dF$ is unique, where $\phi_{\underline{a},k}(\cdot) = \min_{1\le j\le k} ||\cdot - \underline{a}_j||$,

(ii) the map $\underline{a} \to \int \phi_{\underline{a},k}(\cdot)dF$ has positive definite second order derivative Γ at $\underline{a} = \underline{a}^*$,

(iii) $\int ||\underline{x}||^2 dF < \infty$.

Let $\underline{\beta}_n = (\underline{a}_n{}', \underline{0}')'$ and $\underline{\beta}^* = (\underline{a}^{*\prime}, \underline{0}')'$, which are $pK \times 1$ vectors where K is the maximum number of clusters being considered. Then $\sqrt{n}(\underline{\beta}_n - \underline{\beta}^*)$ converges weakly to $N(\underline{0}, T)$ where

$$T = \begin{pmatrix} \Gamma^{-1}V\Gamma^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in R^{pK \times pK}$$

and $V = \int \eta(\cdot, \underline{a}^*)\eta(\cdot, \underline{a}^*)' dF$ where $\eta(\underline{x}, \underline{a}) = \frac{(\underline{x}-\underline{a})}{||\underline{x}-\underline{a}||}$.

Proof: See Appendix. □

Choice of the weight function

The modified k-spatial medians clustering uses a weight function $w(k)$ to deal with the decreasing sum of within-cluster absolute deviations caused by increasing the number of clusters. The number k of clusters in the modified k-spatial medians clustering largely depends on the type of weight function $w(k)$. If the weight function increases too fast(or slow) as k increases, the number of clusters tends to be selected as a small(or large) one. So it is important to find a proper weight function for successful performance of the modified k-spatial medians clustering. Marriott (1971) used the weight function of the form $w(k) = k^2$ in order to obtain the k_n for the k-means clustering. Krzanowski and Lai (1988) also gave a refinement of the method, and Rost (1995) applied the result for some cases. However, the method was motivated from uniform underlying distribution setting. We extend the method to general underlying distribution by proposing a weight function of the form $w_\varepsilon(k) = k^\varepsilon$ for the k-spatial medians clustering, where ε is a parameter for the amount of adjustment of the clustering criterion. Naturally the value of ε should be positive and it decides an extent of penalizing the subdivision of k clusters.

Now, we propose a data dependent choice of the parameter ε as the following manner. Consider ε_k which satisfies

$$W(\underline{a}^{*(1)}, 1, F) \quad = \quad k^{\varepsilon_k} W(\underline{a}^{*(k)}, k, F) \tag{2.3}$$

for $k = 1, \cdots, K$, where $\underline{a}^{*(k)}$ minimizes $\int \min_{1 \le j \le k} ||\underline{x} - \underline{a}_j|| dF$. Large ε_k implies small within cluster sum of absolute deviations and small ε_k implies large within cluster sum of absolute deviations for the clustering results when the number of clusters is k. By considering this trade-off, we may define a correct value ε^* as the average of ε_k for a distribution F as following.

$$\varepsilon^* = \sum_{k=2}^{K} \frac{\varepsilon_k}{K-1} \tag{2.4}$$

For a given data set, we also can define sample version ε_n

$$\varepsilon_n = \sum_{k=2}^{K} \frac{\varepsilon_{nk}}{K-1} \tag{2.5}$$

where ε_{nk} satisfies $W(\underline{a}_n^{(1)}, 1, F_n) = k^{\varepsilon_{nk}} W(\underline{a}_n^{(k)}, k, F_n)$ for $k = 1, \cdots, K$.

Consequently, we propose a data dependent choice of $\varepsilon = \varepsilon_n$ and weight function for the k-spatial medians clustering would be $w_{\varepsilon_n}(k) = k^{\varepsilon_n}$. Consistency of the choice ε_n can be obtained as in Theorem 2.3.

Theorem 2.3. *Suppose that $\int ||\underline{x}|| dF < \infty$ and that there is a unique vector $\underline{a}^*$ for which minimizes $\int \min_{1 \le j \le k} || \cdot - \underline{a}_j || dF$ for each $j = 1, 2, ..., k$. Then ε_n converges to ε^* almost surely.*

Proof: See Appendix. □

3. EXAMPLES

Following two examples show the applicability of the modified k-spatial medians clustering. The first example is analytic application of the modified k-spatial medians clustering for a underlying distribution and the second one is practical application of the procedure for a generated data set.

Example 3.1 Suppose that the underlying distribution F forms

$$F \ : \ \text{Uniform}(0,1) \cup \text{Uniform}(2,3)$$

One can easily find that

$$W(\underline{a}^*, k, F) = \begin{cases} 1, & k = 1 \\ \frac{1}{2k}, & k = 2, 4, 6, \cdots \\ \frac{k}{2(k^2-1)}, & k = 3, 5, 7, \cdots \end{cases}$$

Hence we have

$$\varepsilon^* = \begin{cases} \frac{1}{K-1}\left(\sum_{i=2}^{K} \frac{\log 2}{\log i} + \sum_{j=1}^{K/2-1} \frac{\log((2j+1)^2-1)}{\log(2j+1)} + 1\right), & K \text{ is even} \\ \frac{1}{K-1}\left(\sum_{i=2}^{K} \frac{\log 2}{\log i} + \sum_{j=1}^{(K-1)/2} \frac{\log((2j+1)^2-1)}{\log(2j+1)}\right), & K \text{ is odd} \end{cases}$$

The ε^* and $\overline{W}(\underline{a}^*, k, F)$ are given in Table 3.1. Those are computed for each K from 3 to 20. For every K, the minimum of $\overline{W}(\underline{a}^*, k, F)$ is achieved at $k = 2$. □

Table 3.1 Result of the modified k-spatial medians clustering for Example 3.1

K	ε^*	$\overline{W}(\underline{a}^*, k, F)$ for each k																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	1.315	1.00	.622	.795																	
4	1.377	1.00	.649	.851	.843																
5	1.311	1.00	.620	.791	.769	.859															
6	1.326	1.00	.626	.804	.785	.880	.897														
7	1.302	1.00	.616	.784	.760	.847	.859	.919													
8	1.306	1.00	.618	.788	.765	.853	.866	.927	.946												
9	1.293	1.00	.612	.776	.751	.835	.846	.903	.920	.965											
10	1.294	1.00	.613	.777	.752	.836	.847	.905	.922	.966	.984										
11	1.285	1.00	.609	.769	.742	.824	.834	.889	.905	.948	.964	.999									
12	1.284	1.00	.609	.769	.742	.823	.833	.888	.904	.946	.963	.998	1.015								
13	1.278	1.00	.606	.763	.735	.815	.823	.877	.891	.933	.948	.982	0.998	1.026							
14	1.277	1.00	.605	.762	.734	.813	.821	.875	.889	.930	.946	.979	0.995	1.023	1.038						
15	1.271	1.00	.603	.758	.728	.806	.813	.866	.879	.919	.934	.967	0.982	1.010	1.024	1.048					
16	1.270	1.00	.603	.757	.727	.804	.811	.863	.877	.916	.931	.964	0.978	1.006	1.020	1.044	1.058				
17	1.266	1.00	.601	.753	.722	.799	.805	.856	.869	.908	.922	.954	0.968	0.995	1.008	1.032	1.045	1.065			
18	1.264	1.00	.600	.752	.721	.797	.803	.853	.866	.905	.919	.950	0.964	0.991	1.004	1.027	1.040	1.061	1.073		
19	1.260	1.00	.599	.749	.717	.792	.797	.847	.859	.897	.911	.942	0.955	0.981	0.994	1.017	1.030	1.050	1.062	1.080	
20	1.259	1.00	.598	.747	.716	.790	.795	.845	.857	.894	.908	.938	0.952	0.977	0.990	1.013	1.025	1.045	1.057	1.075	1.086

Example 3.2 Each of 40 data points are coming from three multivariate normal distributions $N(\underline{\mu}_i, \Sigma)$, $i = 1, 2, 3$ with $\underline{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\underline{\mu}_2 = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$, $\underline{\mu}_3 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$ and covariance structure $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Figure 3.1 represents the generated data set. Each point is plotted by its distribution identification number.

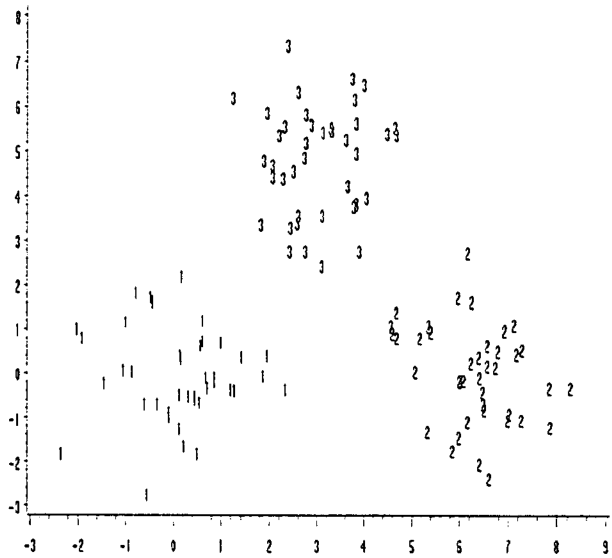

Figure 3.1 Scatterplot of the generated data

Table 3.2 shows the result of the modified k-spatial medians clustering for the generated data set. The ε of the weight function is computed as $\varepsilon_n = \sum_{i=2}^{K} \frac{\varepsilon_{ni}}{K-1} = 0.703$ where ε_{ni}'s satisfy $W(\underline{a}_n^{(1)}, 1, F_n) = 2^{\varepsilon_{n2}} W(\underline{a}_n^{(2)}, 2, F_n) = 3^{\varepsilon_{n3}} W(\underline{a}_n^{(3)}, 3, F_n) = \cdots = 9^{\varepsilon_{n9}} W(\underline{a}_n^{(9)}, 9, F_n)$. In the example, we consider $K = 9$ as the maximum number of clusters. As shown in Example 3.1, we can choose a larger number than the optimal number of clusters as the maximum number of clusters. The minimum value of $\overline{W}(\underline{a}_n, k, F_n)$ is achieved at $k = 3$, $\overline{W}(\underline{a}_n^{(3)}, 3, F_n) = 335.317$. Of course the resulting clusters of the modified k-spatial medians clustering accord with Figure 3.1. □

Table 3.2 The result of the modified k-spatial medians clustering for Example 3.2

k	$W(a_n^{(k)}, k, F_n)$	ε_{nk}	$\overline{W}(a_n^{(k)}, k, F_n)$
1	422.289		422.289
2	313.960	0.428	511.133
3	154.876	0.913	335.317
4	144.591	0.773	383.230
5	130.959	0.727	406.063
6	120.950	0.698	426.322
7	107.395	0.704	421.877
8	99.074	0.697	427.502
9	93.743	0.685	439.422

4. CONCLUSIONS

It is not uncommon to have outliers and particular structures in real life clustering situations. Since the spatial median is robust against outliers, we expect to obtain a reasonable partitioning result by using the k-spatial medians clustering. However, the number k of clusters should be given in advance for the k-spatial medians clustering as in other partitioning methods. We propose a modified k-spatial medians clustering procedure which determines the number of clusters data-dependently. This method incorporates the number k of clusters into the k-spatial medians clustering criterion through the weight function. It makes an application of the k-spatial medians clustering possible without pre-determined number of clusters. Anyhow, it is important to choose a proper weight function for reasonable resulting clusters. We also propose a method for the choice of a proper weight function in general setting. Applicability of the modified k-spatial medians clustering is shown by giving some examples with properly determined weight function. Moreover, we show consistency of the proposed number k_n of clusters and that of sample k_n-spatial medians. Asymptotic normality of sample k_n-spatial medians is also obtained.

APPENDIX

Proof of Theorem 2.1 Put $\overline{\phi}_{\underline{a},k}(\underline{x}) = w(k)\phi_{\underline{a},k}(\underline{x})$, $\underline{x} \in R^p$, where $\phi_{\underline{a},k}(\cdot) = \min_{1\le j\le k} ||\cdot - \underline{a}_j||$. Let $\mathcal{T} = \{\overline{\phi}_{\underline{a},k}; (\underline{a}_1, \cdots, \underline{a}_k) \in C_k, k \le (Nk^* = K)\}$, where $(\underline{a}_1, \cdots, \underline{a}_k) \in C_k$ if and only if $\underline{a}_i \in B(M)$ for at least one i out of $1 \le i \le k$, where the closed ball $B(M)$ centered at origin and of radius M which is large enough to satisfy $(\underline{a}_{n1}, \cdots, \underline{a}_{nk_n}) \in C_{k_n}$ for sufficiently large n. If all $\underline{a}_i$, $1 \le i \le k_n$, lies outside of $B(M)$, then

$$\begin{aligned}\overline{W}(\underline{a}_n, k_n, F_n) \quad &\ge \int_{B(\frac{M}{2})} |\frac{M}{2}| dF_n \\ &\to \int_{B(\frac{M}{2})} |\frac{M}{2}| dF \qquad a.s.\end{aligned}$$

Since $k_n \le Nk^*$ and $w(k)$ is (strictly) increasing function we have

$$\begin{aligned}\overline{W}(\underline{a}_n, k_n, F_n) \quad &\le w(k_n) W(\underline{0}, k_n, F_n) \ \text{ because } \underline{a}_n \text{ is optimal for } F_n \\ &\le w(Nk^*) W(\underline{0}, k_n, F_n) \\ &\to w(Nk^*) \int ||\underline{x}|| dF \qquad a.s. \\ &< \infty\end{aligned}$$

If we choose M such that $w(Nk^*) \int ||\underline{x}|| dF < \int_{B(\frac{M}{2})} |\frac{M}{2}| dF$, then there must eventually be at least one of the optimal centers $\underline{a}_{ni}$, $1 \le i \le k_n$, within $B(M)$,thus $\overline{\phi}_{\underline{a}_n,k_n} \in \mathcal{T}$. Since the graphs of elements of $\mathcal{T} = \{\overline{\phi}_{\underline{a},k}; (\underline{a}_1, \cdots, \underline{a}_k) \in C_k,\ k \le Nk^*\}$ form a Vapnik-Chervonenkis class (see Pollard, 1984) and $\sup_{\mathcal{T}} \overline{\phi}_{\underline{a},k} \le \Phi$ for a Φ with $\int \Phi(\cdot) dF < \infty$ provided $\int ||\underline{x}|| dF < \infty$ and $k \le Nk^*$, we have the follows by using properties of Vapnik-Chervonankis classes.

$$\begin{aligned} & \sup_{\mathcal{T}} |\int f dF_n - \int f dF| \xrightarrow{a.s} 0 \\ \Rightarrow \quad & \liminf \inf_{\mathcal{T}} (\int f dF_n - \int f dF) \ge 0 \\ \Rightarrow \quad & \liminf (\int \overline{\phi}_{\underline{a}_n,k_n} dF_n - \int \overline{\phi}_{\underline{a}_n,k_n} dF) \ge 0 \quad \text{as } \overline{\phi}_{\underline{a}_n,k_n} \in \mathcal{T} \\ \Rightarrow \quad & \liminf (\overline{W}(\underline{a}_n, k_n, F_n) - \overline{W}(\underline{a}_n, k_n, F)) \ge 0\end{aligned}$$

Therefore for a given $\epsilon > 0$,

$$\overline{W}(\underline{a}_n, k_n, F) < \overline{W}(\underline{a}_n, k_n, F_n) + \epsilon_1 \qquad \text{(A.1)}$$

and

$$\begin{aligned}\overline{W}(\underline{a}_n, k_n, F_n) &\leq \overline{W}(\underline{a}^*, k^*, F_n) \text{ because } \underline{a}_n, k_n \text{ is optimal for } F_n \\ &\to \overline{W}(\underline{a}^*, k^*, F) \quad a.s. \\ &\leq \overline{W}(\underline{a}_n, k_n, F) \text{ because } \underline{a}^*, k^* \text{ is optimal for } F\end{aligned}$$

Hence

$$\begin{aligned}\Rightarrow \overline{W}(\underline{a}_n, k_n, F_n) &\leq \overline{W}(\underline{a}^*, k^*, F) + \epsilon_2 \\ &\leq \overline{W}(\underline{a}_n, k_n, F) + \epsilon_2 \\ &\leq \overline{W}(\underline{a}_n, k_n, F_n) + \epsilon_1 + \epsilon_2 \qquad \text{by (A.1)}\end{aligned}$$

From this we get

$$\overline{W}(\underline{a}_n, k_n, F) - \overline{W}(\underline{a}^*, k^*, F) < \epsilon_1 + \epsilon_2$$

As $\epsilon_1, \epsilon_2 > 0$ was arbitrarily chosen, we thus obtain

$$\overline{W}(\underline{a}_n, k_n, F) \longrightarrow \overline{W}(\underline{a}^*, k^*, F) \; a.s. \tag{A.2}$$

According to the uniqueness assumption of k^* and $\underline{a}^*$ finally, it follows from (A.2) theorem holds true. □

Proof of Theorem 2.2 By Theorem 2.1 k_n converges to k^* almost surely. Thus $k_n = k^*$ for sufficiently large n. Let $\underline{c}_n = (\underline{c}_{n1}, \underline{c}_{n2}, \cdots, \underline{c}_{nk^*})$ satisfy

$$W(\underline{c}_n, k^*, F_n) = \inf_{\underline{a}} W(\underline{a}, k^*, F_n)$$

and

$$\begin{aligned}\underline{v}_n &= (\underline{v}_{n1}, \underline{v}_{n2}, \cdots, \underline{v}_{nk^*}) \\ &= I(k_n = k^*)(\underline{a}_{n1}, \underline{a}_{n2}, \cdots, \underline{a}_{nk^*}) + I(k_n \neq k^*)(\underline{c}_{n1}, \underline{c}_{n2}, \cdots, \underline{c}_{nk^*})\end{aligned}$$

where $I(\cdot)$ is an indicator function. If we set a $pK \times 1$ vector $\nu_n = (\underline{v}_n, \underline{0})$, then $\underline{\nu}_n = \underline{\beta}_n$ for sufficiently large n, and $\sqrt{n}(\underline{\nu}_n - \underline{\beta}^*)$ and $\sqrt{n}(\underline{\beta}_n - \underline{\beta}^*)$ are stochastically equivalent. Since

$$W(\underline{v}_n, k^*, F_n) = \inf_{\underline{a}} W(\underline{a}, k^*, F_n)$$

by C.L.T. of sample k-spatial medians

$$\sqrt{n}(\underline{v}_n - \underline{a}^*) \xrightarrow{d} N(\underline{0}, \Gamma^{-1}V\Gamma^{-1})$$

$$\sqrt{n}(\underline{\nu}_n - \underline{\beta}^*) \xrightarrow{d} N(\underline{0}, T)$$

Then, we have $\sqrt{n}(\underline{\beta}_n - \underline{\beta}^*) \xrightarrow{d} N(\underline{0}, T)$ □

Proof of Theorem 2.3 By the result of Pollard (1981), $W(\underline{a}_n, k, F_n)$ converges to $W(\underline{a}^*, k, F)$ almost surely. Since ε_{nk} is a linear combination of $W(\underline{a}_n, k, F_n)$'s, it is clear that ε_{nk} converges to ε_k of (2.3) almost surely. Therefore, ε_n from (2.5) converges to ε^* from (2.4) almost surely as n tends to ∞. □

ACKNOWLEDGEMENT

This work was supported by grant No.981-0105-026-2 from the Basic Research program of the Korea Science & Engineering Foundation.

REFERENCES

Brown, B. M. (1983). "Statistical uses of the spatial median", *J. R. Statist. Soc.* **B 45**, 25-30.

Butler, R. W. (1986). "Optimal stratification and clustering on the line using the L_1-norm", *Journal of Multivariate Analysis*, **18**, 142-155.

Gaenssler, P. (1988). "On a modification of the k-means clustering procedure", in Proceedings of the fifth international symposium on data analysis and informatics. *Data analysis and informatics, V*, Elsevier Science Publisher B.V., 365-371.

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley & Sons.

Hartigan, J. A. (1978). "Asymtotic distributions for clustering criteria", *Annals of Statistics*, **6**, 177-131.

Jhun, M. (1986). "Bootstrap Method for k-Spatial Medians", *Jouranl of the Korean Statistical Society*, **15**, 1-8.

Krzanowski, W. J. & Lai, Y. T. (1988). "A criterion for determining the number of groups in a data set using sum-of-squares clustering", *Biometrics*, **44**, 23-34.

Marriott, F. H. C. (1971). "Practical problems in a method of cluster analysis", *Biometrics*, **27**, 501-514.

Milligan, G. W. & Cooper, M. C. (1985). "An examination of procedures for determining the number of clusters in a data set", *Psychometrika*, **50**, 159-179.

Peck, R., Fisher, L., & van Ness, J. (1989). "Approximate confidence intervals for the number of clusters", *Journal of the American Statistical Association*, **84**, 184-191.

Pollard, D. (1981). "Strong consistency of k-means clustering", *Annals of Statistics*, **9**, 135-140.

Pollard, D. (1982). "A central limit theorem for k-means clustering", *Annals of Probability*, **10**, 919-926.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer Verlag.

Rost, D. (1995). "A simulation study of the weighted k-means cluster procedure", *Journal of Statistical Computing and Simulation*, **53**, 51-63.

Spath, H. (1980) . *Clustering Analysis Algorithms for data reduction and classification of objects,* John Wiley & Sons.