Working Paper 92-36
September 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624 9849

A RATE OF CONVERGENCE IN CLUSTERING ANALYSIS

Juan Romo*

Abstract

We present a result about stochastic boundedness of stable empirical processes on Vapnik-Červonenkis classes of functions and we apply it to obtain a rate of convergence for the approximation between the sample and the populational variation in the k-centroids problem in clustering analysis.

Key words:
Clustering analysis; k-centroids; empirical processes; Vapnik-Červonenkis classes of functions.

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain.

1 INTRODUCTION

The k-means method in clustering analysis to partition n points into k groups can be described in the following way: given points $x_1, \ldots, x_n$ in $\mathcal{R}^d$ we find centers $b_1, \ldots, b_k$ minimizing

$$\sum_{i=1}^{n} \min_{1 \leq j \leq k} \| x_i - b_j \|^2$$

and then we assign x_i to the closest center ($\| \cdot \|$ is the euclidean norm). The name of the criterion is due to the fact that each center b_j is the mean of the points x_i in its cluster. This problem was already consider by W.D. Fisher (1958). MacQueen (1967) presented the more general case of k-centroids: given points $x_1, \ldots, x_n$ in $\mathcal{R}^d$, to find centers $a_1, \ldots, a_k$ minimizing

$$\sum_{i=1}^{n} \min_{1 \leq j \leq k} \| x_i - a_j \|^p, \quad 1 \leq p \leq 2.$$

The statistical version of this problem is as follows: let P be a probability measure on $(\mathcal{R}^d, \mathcal{B}_{\mathcal{R}^d})$ and let $\{X_n : n \in \mathcal{N}\}$ be a sequence of independent and identically distributed random variables with law P. Let P_n be the empirical measure associated to $X_1, \ldots, X_n$. For each set $a = a(k) = \{a_1, \ldots, a_k\} \subset \mathcal{R}^d$, we define

$$f_p(x, a) = \min_{1 \leq j \leq k} \| x - a_j \|^p, \quad 1 \leq p \leq 2.$$

Let

$$Pf_p(\cdot, a) \equiv \int_{\mathcal{R}^d} \min_{1 \leq j \leq k} \| x - a_j \|^p \, dP(x) = \sum_{j=1}^{k} \int_{A_j} \| x - a_j \|^p \, dP(x),$$

where A_j is the set of points closest to a_j than to any other center. Also, we have

$$P_n f_p(\cdot, a) = \frac{1}{n} \sum_{i=1}^{n} \min_{1 \leq j \leq k} \| X_i - a_j \|^p .$$

Let $\alpha = \alpha(k) = \{\alpha_1, \ldots, \alpha_k\}$ be the k-centroid minimizing $Pf_p(\cdot, a)$ over all sets $a \subset \mathcal{R}^d$ having at most k points, i.e.,

$$Pf_p(\cdot, \alpha) = \inf_{\{a : \sharp a \leq k\}} Pf_p(\cdot, a). \tag{1.1}$$

We will assume that α is unique. Let $\alpha^{(n)} = \alpha^{(n)}(k) = \{\alpha_1^{(n)}, \ldots, \alpha_k^{(n)}\}$ be the corresponding sample value, i.e.,

$$P_n f_p(\cdot, \alpha^{(n)}) = \inf_{\{a : \sharp a \leq k\}} P_n f_p(\cdot, a). \tag{1.2}$$

In clustering analysis is interesting to establish the convergence of the sample clusters to the populational ones and the convergence of the sample variation $P_n f_p(\cdot, \alpha^{(n)})$ to $Pf_p(\cdot, \alpha)$. The convergence of the sequence of sets must be understood in the sense of the Hausdorff's metric: if A and B are nonempty compact subsets of $\mathcal{R}^d$, the Hausdorff metric is

$$d_H(A, B) \equiv \max\{\sup_{x \in A} d(x, B), \sup_{x \in B} d(y, A)\},$$

where $d(x, C) = \inf_{y \in C} \| x - y \|$ for any $C \subset \mathcal{R}^d$. So $d_H(A, B) < \delta$ if, and only if, for any point in A there exists a point in B such that the distance between them is less than δ, and conversely. If follows that if A has k different points and δ is less than half of the minimum distance between points in A, B is a set with at most k points and such that $d_H(A, B) < \delta$, then B has to contain exactly k different points each of which lies within a distance δ of a unique point in A. It follows that the sequence of sets $\{\alpha^{(n)} : n \in \mathcal{N}\}$ converges to the set α if for every n there exists a labelling $(\alpha_1^{(n)}, \ldots, \alpha_k^{(n)})$ for the points in $\alpha^{(n)}$ and a labelling $(\alpha_1, \ldots, \alpha_k)$ for the points in α such that $\alpha_i^{(n)} \to \alpha_i$ for $1 \leq i \leq k$.

Pollard (1981) proved that if $\int_{\mathcal{R}^d} \| x \|^d \, dP(x) < \infty$, $1 \leq p \leq 2$, and for each $j \in \mathcal{N}$, $1 \leq j \leq k$, there exists a unique $\alpha(j)$ satisfying (1.1) then

$$\alpha^{(n)} \to \alpha \text{ almost surely as } n \to \infty \tag{1.3}$$

and

$$P_n f_p(\cdot, \alpha^{(n)}) \to P f_p(\cdot, \alpha) \text{ almost surely as } n \to \infty. \tag{1.4}$$

This improves the result of Sverdrup-Thygeson (1981) who obtained the same conclusion assuming compactness of the sample space.

For $p = 2$, Pollard (1982) established asymptotic normality of $n^{1/2}(\alpha^{(n)} - \alpha)$ assuming that $\int_{\mathcal{R}^d} \| x \|^2 \, dP(x) < \infty$, that P is absolutely continuous and under conditions ensuring that the second order differential of the map $a \mapsto Pf(\cdot, a)$ is positive definite. For $p = 1$ and $d = 1$, Butler (1986) has shown asymptotic normality of $n^{1/2}(\alpha^{(n)} - \alpha)$ assuming $\int_{\mathcal{R}^d} \| x \| \, dP(x) < \infty$ and positive definite second order derivative. See also Butler (1988) for central limit results for universally optimal and bounded locally optimal theory and for techniques for assesing the number of groups in the data.

Our main theorem will follow from a result about stochastic boundedness for empirical processes on Vapnik-Červonenkis classes of functions; so we recall some notation from the theory of empirical processes. For a probability space $(S, \mathcal{S}, P)$, $\mathcal{L}_p$ $(S, \mathcal{S}, P)$, $0 < p < \infty$, will be the class of real measurable functions f on S such that $\int_S | f |^p \, dP < \infty$. If $f \in \mathcal{L}_1(S, \mathcal{S}, P)$, we will write Pf instead of $\int_S f dP$. If $\mathcal{F}$ is a collection of measurable functions on S, we will assume that the envolope of $\mathcal{F}$, $F(s) \equiv \sup_{f\in\mathcal{F}} \mid f(s) \mid$, is finite for all $s \in S$. Although F is not necessarily measurable, there always exists a measurable $F^* : S \to \mathcal{R}$ (see Dudley (1984), Theorem 3.1.1) such that $F \leq F^*$ and for all measurable functions h such that $F \leq h$, we have that $F^* \leq h$ almost surely with respect to P. For $1 \leq p \leq 2$, we will write

$$\nu_n^{(p)} = n^{1-\frac{1}{p}}(P_n - P), \qquad n \in \mathcal{N}.$$

Since F is finite for all $s \in S$, we will have that for all $n \in \mathcal{N}$ and for all $p \in [1,2]$,

$$\{\nu_n^{(p)}(f) : f \in \mathcal{F}\} = \{n^{-1/p}\sum_{i=1}^{n}(f(X_i) - Pf) : f \in \mathcal{F}\}$$

is a stochastic process with sample paths in the space $l^\infty(\mathcal{F})$ of all real bounded functions defined on $\mathcal{F}$. If H is a real bounded function on $\mathcal{F}$, we will put $\| H \|_{\mathcal{F}} \equiv \sup_{f\in\mathcal{F}} \mid H(f) \mid$. We say that $\mathcal{F}$ is *supremum measurable* (for P) if

$$\sup_{f\in\mathcal{F}} \mid Q\{f(X_n)) : n \in \mathcal{N}\} \mid$$

is $P^{\mathcal{N}}$- completion measurable for each function Q which is a linear or quadratic function of finitely many of the $f(X_n)$ (see Alexander (1987)).

Let T be a non-empty set and let $A \subset T$. If $\mathcal{C}$ is a class of subsets of T, we will write

$$\triangle^{\mathcal{C}}(A) \equiv \{C \cap A : C \in \mathcal{C}\}$$

and

$$m^{\mathcal{C}}(n) \equiv \max\{\triangle^{\mathcal{C}}(A) : A \subset T, \sharp A = n\}, \quad n \in \mathcal{N}.$$

Define the *(Vapnik-Červonenkis) index* of $\mathcal{C}$ as $V(\mathcal{C}) \equiv \min\{n \in \mathcal{N} : m^{\mathcal{C}}(n) < 2^n\}$ or $V(\mathcal{C}) \equiv +\infty$ if $m^{\mathcal{C}}(n) = 2^n$ for all $n \in \mathcal{N}$. We say that $\mathcal{C}$ is a *Vapnik-Červonenkis class of sets* if $V(\mathcal{C}) < \infty$. It f is a real function on S, its graph is $G(f) \equiv \{(s,t) \in S \times \mathcal{R} : 0 \leq t \leq f(s) \text{ or } f(s) \leq t \leq 0\}$. $\mathcal{F} \subset \mathcal{R}^S$ is a *Vapnik-Červonenkis class of functions* if $\{G(f) : f \in \mathcal{F}\}$ is a Vapnik-Červonenkis class of sets in $S \times \mathcal{R}$.

We will denote by $N(\varepsilon, \rho, \mathcal{F})$ the minimum number of balls with radius $\varepsilon > 0$ for the metric ρ and centers in $\mathcal{F}$ needed to cover $\mathcal{F}$. For any finite set $T \subset S$, let

$$d_T^{(2)}(f,g) = \left(\frac{\sum_{x\in T}(f-g)^2(x)}{\sum_{x\in T} F^2(x)}\right)^{\frac{1}{2}}$$

and define

$$N^{(2)}(\varepsilon, \mathcal{F}) = \sup_{T\in S, T\, finite} N(\varepsilon, d_T^{(2)}, \mathcal{F}).$$

If $\{Z(t) : t \in T\}$ is a subgaussian process, Marcus and Pisier (1978) proved that if

$$\int_0^{|T|_\sigma} (\log N(\varepsilon, \sigma, T))^{1/2} d\varepsilon < \infty$$

then Z has a version $\tilde{Z}$ with σ-uniformly continuous sample paths and, for $t_0 \in T$,

$$E \sup_{t \in T} | \tilde{Z}(t) | \leq E | \tilde{Z}(t_0) | + C \left[\int_0^{|T|_\sigma} (\log N(\varepsilon, \sigma, T))^{1/2} d\varepsilon + \phi(| T |_\sigma) \right],$$

where $\phi(\delta) = 4\delta(\log\log 4 \mid T |_\sigma \delta^{-1})^{1/2}, 0 < \delta \leq | T |_\sigma$ and $| T |_\sigma$ is the diameter of T for σ . If $Z(t) = \sum_{i=1}^n \varepsilon_i h_i(t)$ with h_i continuous in (T, σ) then $Z = \tilde{Z}$.

2 Results

Our main result (Theorem 3) gives a rate of convergence for (1.4) in the case $p = 1$; its proof relies on a technical lemma on the differentiability of the application $a \mapsto Pf(\cdot, a)$, where $f(x, a) = \min_{1 \leq j \leq k} \| x - a_j \|$, and on a proposition about stochastic boundedness of the empirical process. First, we present these results.

Lemma 1. Let P be absolutely continuous with respect to Lebesgue measure on $\mathcal{R}^d$ and such that $\int_{\mathcal{R}^d} \| x \|^p dP(x) < \infty$, $1 \leq p \leq 2$. The map $a \mapsto f(\cdot, a)$ from $\mathcal{R}^{kd}$ into $\mathcal{L}^p(\mathcal{R}^d, \mathcal{B}_{\mathcal{R}^d}, P)$ is $\| \cdot \|$-differentiable. As a consequence, the map $a \mapsto Pf(\cdot, a)$ is differentiable.

Proof. We can write

$$f(x, a + h) = f(x, a) + h' \cdot \Delta(x, a) + \| h \| r(x, a, h) \tag{2.5}$$

where

$$\Delta(x, a) = \left(\frac{a_1 - x}{\| x - a_1 \|} I_{A_1}, \ldots, \frac{a_k - x}{\| x - a_k \|} I_{A_k} \right)'$$

and

$$r(x, a, h) \to 0 \quad \text{as} \quad h \to 0$$

for P-almost every x (in fact, except for $x = a_j$, $j = 1, \ldots, k$). Hence

$$| r(x, a, h) | \leq [\max_{1 \leq j \leq k} (\| x - a_j - h \| - \| x - a_j \| + | h' \cdot \Delta(x, a) |] \| h \|^{-1}$$

$$\leq \| h \|^{-1} \sum_{j=1}^{k} (\| x - a_j - h_j \| - \| x - a_j \|) + \| \Delta(x,a) \|$$

$$\leq \| h \|^{-1} \left(\min_{1 \leq j \leq k} (\| x - a_j - h_j \| + \| x - a_j \|) \right) \cdot$$

$$\cdot \sum_{j=1}^{k} (\| x - a_j - h_j \|^2 - \| x - a_j \|^2) + 1$$

$$\leq \| h \|^{-1} \left(\min_{1 \leq j \leq k} \| x - a_j - h_j \| + \| x - a_j \| \right) \sum_{j=1}^{k} \left(\| h_j \|^2 + 2h_j(x - a_j) \right) + 1$$

$$\leq C \left(1 + \sum_{j=1}^{k} \| x - a_j \| \right) \in \mathcal{L}^p(\mathcal{R}^d, \mathcal{B}, P).$$

for small enough h. Then

$$\| r(\cdot, a, h) \|_p \to 0 \quad \text{as} \quad h \to 0 \tag{2.6}$$

and the map $a \mapsto Pf(\cdot, a)$ is $\| \cdot \|_p$-differentiable. It follows that

$$\| r(\cdot, a, h) \|_1 \to 0 \quad as \quad h \to 0;$$

(2.5) and dominated convergence prove that the map $a \mapsto Pf(\cdot, a)$ is differentiable with derivative $P \Delta (\cdot, a)$. This finishes the proof.

Now we establish a sufficient condition for stochastic boundedness of the stable empirical process.

Proposition 1. Let $\mathcal{F} \subset \mathcal{L}_1(S, \mathcal{S}, P)$ be a supremum measurable Vapnik-Červonenkis class of functions with envelope F finite everywhere. Let $1 < p < 2$. If

$$sup_{t>0} t^p P\{F^* > t\} = K < \infty$$

then $\{\| n^{-\frac{1}{p}} \sum_{i=1}^{n} (f(X_i) - Pf) \|_{\mathcal{F}} : n \in \mathcal{N}\}$ is stochastically bounded.

Proof. From Andersen, Giné and Zinn (1988), it is enough to show that for some $\delta > 0$,

$$\lim_{M \to \infty} \limsup_{n \to \infty} Pr^* \{\| n^{-\frac{1}{p}} \sum_{i=1}^{n} \varepsilon_i f I_{\{F^* \leq \delta n^{\frac{1}{p}}\}}(X_i) \|_{\mathcal{F}} > M\} = 0. \tag{2.7}$$

Let $\delta = 1$ and define $\mathcal{F}_n = \{fI_{\{F^* \leq n^{1/p}\}}\}$ and $F_n = FI_{\{F^* \leq n^{1/p}\}}$. For each $x_1, \ldots, x_n$ fixed, consider the subgaussian process

$$\{Z(f) : f \in \mathcal{F}_n\} \equiv \{n^{-\frac{1}{p}} \sum_{i=1}^{n} \varepsilon_i f(x_i) : f \in \mathcal{F}_n\}$$

and its associated $\mathcal{L}_2$ distance

$$\sigma(f,g) = \left(\frac{\sum_{i=1}^{n} (f-g)^2(x_i)}{n^{2/p}} \right)^{\frac{1}{2}}.$$

The diameter of $\mathcal{F}_n$ for σ is

$$| \mathcal{F}_n |_\sigma \leq 2 \left(\frac{\sum_{i=1}^{n} F_n^{*2}(x_i)}{n^{2/p}} \right)^{\frac{1}{2}} \equiv D_n(F).$$

Let $T_n = \{X_1, \ldots, X_n\}$. Since

$$\sigma(f,g) = D_n(F) d_{T_n}^{(2)}(f,g),$$

we get

$$N(\varepsilon, \sigma, \mathcal{F}_n) = N(\frac{\varepsilon}{D_n(F)}, d_{T_n}^{(2)}, \mathcal{F}_n) \leq N^{(2)} \left(\frac{\varepsilon}{D_n(F)}, \mathcal{F}_n \right).$$

Now,

$$Pr^*\{\| n^{-\frac{1}{p}} \sum_{i=1}^{n} \varepsilon_i f(X_i) \|_{\mathcal{F}_n} > M\}$$

$$\leq E_X \left[1 \wedge \frac{1}{M} E_\varepsilon \| \frac{\sum_{i=1}^{n} \varepsilon_i f(X_i)}{n^{1/p}} \|_{\mathcal{F}_n} \right]$$

(by Marcus and Pisier (1978))

$$\leq \frac{C_p}{M} E_X \left[\int_0^{D_n(F)} (\log N(\varepsilon, \sigma, \mathcal{F}_n))^{\frac{1}{2}} d\varepsilon + 4(\log\log 4)^{\frac{1}{2}} D_n(F) \right]$$

$$\leq \frac{C_p}{M} E_X \left[\int_0^{D_n(F)} (\log N^{(2)}(\frac{\varepsilon}{D_n(F)}, \mathcal{F}_n))^{\frac{1}{2}} d\varepsilon + 4(\log\log 4)^{\frac{1}{2}} D_n(F) \right]$$

$$= \frac{C_p}{M} E_X D_n(F) \left[\int_0^1 (\log N^{(2)}(\varepsilon, \mathcal{F}_n))^{\frac{1}{2}} d\varepsilon + 4(\log\log 4)^{\frac{1}{2}} \right]$$

$$\leq \frac{C_p}{M} E_X D_n(F) \left[\int_0^1 (\log N^{(2)}(\varepsilon, \mathcal{F}))^{\frac{1}{2}} d\varepsilon + 4(\log\log 4)^{\frac{1}{2}} \right]$$

Since F is a Vapnik-Červonenkis class, the integral between brackets is finite; so to obtain (2.7), it is enough to show that

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{E_X D_n(F)}{M} = 0.$$

This will follow if we show that $E_X D_n(F)$ is bounded in n; but

$$E_X D_n^2(F) = 2nE_X \left[\frac{F^{*2}}{n^{2/p}} I_{\{F^* \leq n^{1/p}\}}\right]$$

$$= 4\int_0^\infty ntPr\{\frac{F^*}{n^{1/p}} I_{\{F^* \leq n^{1/p}\}} > t\}dt$$

$$\leq 4\int_0^1 ntPr\{F^* > tn^{1/p}\}dt$$

$$\leq 2\int_0^1 K_p t^{1-p} dt = \frac{2K_p}{2-p} < \infty,$$

because $1 < p < 2$, and the result follows.

Our final result gives the rate of convergence for the approximation of the sample variation to the populational one.

Theorem 3. Let P be absolutely continuous with respect to Lebesgue measure on $\mathcal{R}^d$ and such that $\int_{\mathcal{R}^d} \| x \|^p \, dP(x) < \infty$, $1 \leq p \leq 2$. Then

$$n^{1-\frac{1}{p}}(P_n f(\cdot, \alpha^{(n)}) - Pf(\cdot, \alpha)) \longrightarrow 0 \qquad \textit{in probability.}$$

Proof. If $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ minimizes (1.1) (for p=1), consider the class of functions

$$\{r(\cdot, \alpha, a - \alpha) : a \in \mathcal{U}(\alpha)\}$$

where $\mathcal{U}(\alpha)$ is a neighborhood of α such that

$$| r(x, \alpha, a - \alpha) | \leq C(1 + \sum_{i=1}^{k} \| x - a_j \|)$$

for all $x \in \mathcal{R}$ and all $a \in \mathcal{U}(\alpha)$. Let B_i be the set of points which are closer to α_i than to any other of the elements of α. Then

$$r(x, \alpha, a - \alpha) = \| a - \alpha \|^{-1} (f(x, a) - f(x, \alpha) - (a - \alpha)' \Delta (x, \alpha)) =$$

$$= \sum_{i,j=1}^{k} \| a - \alpha \|^{-1} (\| x - a_j \| - \| x - \alpha_i \| + \| a_i - \alpha_i \| \frac{x - \alpha_i}{\| x - \alpha_i \|}) I_{A_j} I_{B_i}.$$

By Problem II. 28 in Pollard (1984), the class

$$\mathcal{F} = \{ r(\cdot, \alpha, a - \alpha) : a \in \mathcal{U}(\alpha) \}$$

verifies the conditions of Proposition 1 (since the envelope

$$F(x) = C(1 + \sum_{i=1}^{k} \| x - \alpha_i \|) \in \mathcal{L}^p$$

because $\int_{\mathcal{R}^d} \| x \|^p \, dP(x) < \infty$) ; so Proposition 1 proves that $\{\nu_n^{(p)} : n \in \mathcal{N}\}$ is stochastically bounded in $l^\infty(\mathcal{F})$, i.e.,

$$\lim_{M \to \infty} \limsup_{n \to \infty} Pr^* \{ \| \nu_n^{(p)} \|_{\mathcal{F}} > M \} = 0. \tag{2.8}$$

From (2.5), with $a = \alpha$ and $a + h = \alpha^{(n)}$, we get

$$f(x, \alpha^{(n)}) = f(x, \alpha) + (\alpha^{(n)} - \alpha)' \Delta(x, \alpha) + \| \alpha^{(n)} - \alpha \| \, r(x, \alpha, \alpha^{(n)} - \alpha). \tag{2.9}$$

Integrating with respect to $\nu_n^{(p)}$, we obtain

$$\nu_n^{(p)} f(\cdot, \alpha^{(n)}) =$$

$$= \nu_n^{(p)} f(\cdot, \alpha) + (\alpha^{(n)} - \alpha)' \nu_n^{(p)} \Delta(\cdot, \alpha) + \| \alpha^{(n)} - \alpha \| \, \nu_n^{(p)} r(\cdot, \alpha, \alpha^{(n)} - \alpha). \tag{2.10}$$

The theorem of Pollard (1981) proves that $\alpha^{(n)} \longrightarrow \alpha$ almost surely. Hence (2.8) implies that

$$\| \alpha^{(n)} - \alpha \| \, \nu_n^{(p)} r(\cdot, \alpha, \alpha^{(n)} - \alpha) \longrightarrow 0 \qquad \textit{in probability.}$$

By the law of large numbers in finite dimensions, the second term in the right hand side of (2.5) also tends to zero in probability. Then, from (2.5) it follows that

$$\nu_n^{(p)} f(\cdot, \alpha^{(n)}) - \nu_n^{(p)} f(\cdot, \alpha) \longrightarrow 0 \qquad \textit{in probability,}$$

i.e.

$$\frac{1}{n^{1/p}}\sum_{i=1}^{n}(f(X_i,\alpha^{(n)}) - Pf(\cdot,\alpha^{(n)})) - \frac{1}{n^{1/p}}\sum_{i=1}^{n}(f(X_i,\alpha) - Pf(\cdot,\alpha)) \longrightarrow 0$$

in probability. By the weak law of large numbers in $\mathcal{R}^d$ and since the map $a \mapsto Pf(\cdot, a)$ is continuous, we obtain

$$n^{1-\frac{1}{p}}(P_n f(\cdot,\alpha^{(n)}) - Pf(\cdot,\alpha)) \longrightarrow 0 \qquad \textit{in} \quad \textit{probability}.$$

This concludes the proof.

References

ALEXANDER, K. S. (1987). The central limit theorem for empirical processes on Vapnik- Červonenkis classes of functions. *Ann. Probab.*, 15, 178-203.

ANDERSEN, N. T., GINÉ, E. and ZINN, J. (1988). The central limit theorem for empirical processes under local conditions: the case of Radon infinitely divisible limits without Gaussian component. *Trans. Amer. Math. Soc.,* 308, 603- 635.

BUTLER, R.W. (1986). Optimal stratification and clustering on the line using the L_1-norm. *J. Mult. Analysis* 19, 142-155.

BUTLER, R. W. (1988). Optimal clustering on the real line. *J. Mult. Analysis* 24, 88-108.

DUDLEY, R.M. (1984). A course on empirical processes. *Lect. Notes in Math.* Springer-Verlag.

FISHER, W. D. (1958). On grouping for maximum homogeneity. *J. Amer. Statis. Assoc.* 53, 789-798.

MACQUEEN, J.B. (1967). Some methods for classification and analysis for multivariate observations. *Proc. Fifth Berk. Symp. Math. Stat. Prob.* 281-296.

MARCUS, M. R. and PISIER, G. (1978). *Random Fourier Series with Applications to Harmonic Analysis.* Princeton U. Press, Princeton, New Jersey.

POLLARD, D. (1981). Strong consistency of k-means clustering. *Ann. Statist.* 9, 135-140.

POLLARD, D. (1982). A central limit theorem for k-means clustering. *Ann. Probab.*, 10, 919- 926.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, Berlin.

SVERDRUP- THYGESON, H. (1981). Strong law of large numbers for measures of central tendency and dispersion of random variables in compact spaces. *Ann. Statist.* 9, 141-145.