The Annals of Statistics
1981, Vol. 9, No. 1, 141–145

STRONG LAW OF LARGE NUMBERS FOR MEASURES OF CENTRAL TENDENCY AND DISPERSION OF RANDOM VARIABLES IN COMPACT METRIC SPACES

BY HARALD SVERDRUP-THYGESON

University of Oslo

Given a sample of independent random variables $Z_1, Z_2, \cdots, Z_n$ with identical distribution p on a compact metric space (M, d), a measure of central tendency is a *sample centroid* (of order $r > 0$) defined as a point $\hat{X}_n$ in M satisfying

$$\frac{1}{n}\sum_{i=1}^{n} d^r(\hat{X}_n, Z_i) = \inf_{x\in M}\frac{1}{n}\sum_{i=1}^{n} d^r(x, Z_i).$$

A (population) *centroid* of Z is any point x^* in M such that

$$\int_M d^r(x^*, z)\,dp(z) = \inf_{x\in M}\int_M d^r(x, z)\,dp(z).$$

The quantity $\frac{1}{n}\sum_{i=1}^{n} d^r(\hat{X}_n, Z_i)$ itself is called the *sample variation*, whereas $\int_M d^r(x^*, z)\,dp(z)$ is the *variation* of Z. This paper establishes almost sure convergence for the sample centroid and variation to the corresponding population values for all orders $r > 0$.

Convergence is also proved for the case when the sample centroid is *restricted* to be one of the sample values.

1. Introduction. Consider a compact metric space (M, d) in which there is observed a random variable Z with probability distribution p. If a sample of several independent observations $Z_1, Z_2, \cdots, Z_n$ is available it may be desirable to summarize the sample, much as one does with ordinary variables, by measures of central tendency and dispersion. To this end one defines a *sample centroid* of order $r > 0$ as a point $\hat{X}_n$ in M satisfying

$$\frac{1}{n}\sum_{i=1}^{n} d^r(\hat{X}_n, Z_i) = \inf_{x\in M}\frac{1}{n}\sum_{i=1}^{n} d^r(x, Z_i).$$

The population value, that is a *centroid* of Z (or of p) of order r, is then any point $x^* \in M$ such that

$$\int_M d^r(x^*, z)\,dp(z) = \inf_{x\in M}\int_M d^r(x, z)\,dp(z).$$

The quantity $\frac{1}{n}\sum_{i=1}^{n} d^r(\hat{X}_n, Z_i)$ itself is called the *sample variation* of order r, whereas $\int_M d^r(x^*, z)\,dp(z)$ is the *variation* of Z of order r. Centroids are often not unique, a fact which causes some technical difficulties (see below).

These concepts have been discussed by MacQueen (1965) who notes that they reduce in Euclidean spaces to the mean, the median, the mode or the mid-range according as $r = 2$, $r = 1$, $r \to 0$ or $r \to \infty$. A variety of metric spaces are encountered in practice so that it may be worth establishing that these notions have the asymptotic stability (strong law of large numbers) so well known in the cases of Euclidean variables. This is accomplished here by

Received July 1979; revised November 1979.

AMS 1970 *subject classifications.* Primary 60F15; secondary 60B99.

Key words and phrases. Strong law of large numbers, compact metric space, central tendency, dispersion, centroid.

proving, in an appropriate sense, a.s. convergence of the sample centroids and sample variations to their corresponding population values (for all $r > 0$).

MacQueen gives a partial proof for the case of $r = 1$ which is based on some uniform convergence results of Parzen. The results presented here follow a different approach.

Previous work by other authors on laws of large numbers seems to be restricted to *linear* spaces (e.g., Hilbert, Banach, Fréchet spaces); see Padgett and Taylor (1973) for a review of such theory.

Closely related to the above notions is that of a *sample restricted centroid* (of order r), which is one of the sample points $\hat{Z}_n$ satisfying

$$\frac{1}{n}\sum_{i=1}^{n} d^r(\hat{Z}_n, Z_i) = \min_{1\le j\le n}\frac{1}{n}\sum_{i=1}^{n} d^r(Z_j, Z_i).$$

In this case, an appropriate *restricted* (population) *centroid* z^* is defined by

$$\int_M d^r(z^*, z)\,dp(z) = \inf_{z'\in W}\int_M d^r(z', z)\,dp(z),$$

where W is the support of the probability measure p. Further, the *restricted variation* is for the sample, $\frac{1}{n}\sum_{i=1}^{n} d^r(\hat{Z}_n, Z_i)$, and for the population $\int_M d^r(z^*, z)\,dp(z)$. These concepts are of special interest in situations where the objects in the sample do not have any easily quantified physical description, yet distances can still be measured. For example, suppose the sample is a set of 18th century paintings. Distances between all pairs can be obtained by judgement methods, but construction of an 18th century painting would not be so easy. In this case one of the sample points will serve quite well. Note that for purposes of comparing several groups of such objects, say 18th and 19th century paintings, the distances among all pairs being measured under the same conditions, the two restricted sample centroids potentially are of independent interest as exemplars. Knowing that they have (or do not have) the kind of stability of other measures of central tendency is of practical interest just as with the nonrestricted sample centroids defined above. It will be shown below that the restricted sample centroids will converge in an appropriate sense, as does the restricted sample variation.

2. Terminology. Given a probability space $(\Omega, \mathscr{A}, P)$ and a compact metric space (M, d), let $\mathscr{B}$ be a σ-algebra containing the open subsets of M, and let $Z_1, Z_2, \cdots$ be independent, identically distributed random variables from $(\Omega, \mathscr{A}, P)$ into $(M, \mathscr{B})$, with their common probability distribution on $(M, \mathscr{B})$ denoted by p.

Let r be any positive constant. For each n define the set of *sample centroids*

$$\hat{C}_n := \left\{\hat{x} \in M : \frac{1}{n}\sum_{i=1}^{n} d^r(\hat{x}, Z_i) = \inf_{x\in M}\frac{1}{n}\sum_{i=1}^{n} d^r(x, Z_i)\right\}.$$

Also define the class of (population) *centroids*

$$C^* := \left\{x^* \in M : \int d^r(x^*, z)\,dp(z) = \inf_{x\in M}\int d^r(x, z)\,dp(z)\right\}.$$

Now let W_n be the sample set $\{Z_1, Z_2, \cdots, Z_n\}$ and let W be the support of p (i.e., the complement of the union of all open sets in M of p-measure 0). Then define the set of *sample restricted centroids*

$$\widehat{CR}_n := \left\{\hat{z} \in W_n : \frac{1}{n}\sum_{i=1}^{n} d^r(\hat{z}, Z_i) = \inf_{z\in W_n}\frac{1}{n}\sum_{i=1}^{n} d^r(z, Z_i)\right\}$$

and the set of *restricted* (population) *centroids*

$$CR^* := \left\{z^* \in W : \int d^r(z^*, z)\,dp(z) = \inf_{z'\in W}\int d^r(z', z)\,dp(z)\right\}.$$

Note that C^* and $\hat{C}_n$ are nonempty (since M is compact), as are CR^* and $\widehat{CR}_n$. Of course one or more of the sets may be a nonsingleton, as is illustrated by the median of a real valued random variable taking two distinct values each with probability ½. In the following, let x^* be an arbitrary element of C^*, and for each n define a function $\hat{X}_n : \Omega \rightarrow M$ by letting $\hat{X}_n$ be an arbitrary element of $\hat{C}_n$. Analogously, fix a z^* in CR^* and a function $\hat{Z}_n : \Omega \rightarrow M$ by $\hat{Z}_n \in \widehat{CR}_n$.

As there are many nontrivial examples in which the *full* sequence of sample centroids (or sample restricted centroids) does not converge, we will have to consider the sets

$$B := \{\text{cluster points of sequences } \langle \hat{x}_n \rangle_{n=1}^{\infty} \text{ in } \textstyle\prod_{n=1}^{\infty} \hat{C}_n\}$$

$$BR := \{\text{cluster points of sequences } \langle \hat{z}_n \rangle_{n=1}^{\infty} \text{ in } \textstyle\prod_{n=1}^{\infty} \widehat{CR}_n\}.$$

The question of measurability needs some special attention. Because $\hat{X}_n$ and $\hat{Z}_n$ need not be measurable, it is convenient to work with a slight modification of the concept of almost sure (a.s.) convergence, denoted as $\widetilde{\text{a.s.}}$: "Property x holds $\widetilde{\text{a.s.}}$" is to mean "The set of ω such that property x holds contains a set in $\mathscr{A}$ of P-measure 1". By using this definition we circumvent the usual need for functions on Ω to be measurable. Note though, that $\sum_{i=1}^{n} d^r(\hat{X}_n, Z_i) \equiv \inf_{x \in M} \sum_{i=1}^{n} d^r(x, Z_i)$ is measurable due to M being separable. Of course, $\sum_{i=1}^{n} d^r(\hat{Z}_n, Z_i) \equiv \min_{1 \le j \le n} \sum_{i=1}^{n} d^r(Z_j, Z_i)$ is measurable, even if M is not separable.

3. Results.

THEOREM.

(i) $\frac{1}{n} \sum_{i=1}^{n} d^r(\hat{X}_n, Z_i) \rightarrow \int d^r(x^*, z)\, dp(z)$ a.s.

(ii) $B \subseteq C^*$ $\widetilde{\text{a.s.}}$

(iii) $\frac{1}{n} \sum_{i=1}^{n} d^r(\hat{Z}_n, Z_i) \rightarrow \int d^r(z^*, z)\, dp(z)$ a.s.

(iv) $BR \subseteq CR^*$ $\widetilde{\text{a.s.}}$

Conditions (ii) and (iv) describe the convergence properties of the sample centroids and the restricted sample centroids. Note that B and BR are nonempty for all ω since M is compact. Also note that (ii) implies the full sequence $\langle \hat{X}_n \rangle$ converges $\widetilde{\text{a.s.}}$ to x^* if x^* is unique. This is an example where $B = C^*$ $\widetilde{\text{a.s.}}$, which is also the case of the aforementioned two-valued random variable. To see that $B = C^*$ $\widetilde{\text{a.s.}}$ does not hold always, consider the cube $M := [-1, 1]^3$ with the metric $d((x_1, x_2, x_3), (y_1, y_2, y_3)) := |x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3|$ and let $Z := (U_1, U_2, U_3)$ be a random variable in M with $\Pr(U_k = -1) = \Pr(U_k = 1) = ½$ $(k = 1, 2, 3)$ where U_1, U_2, U_3 are independent. Let $r = 1$, and assume we have a set of independent observations $Z_1, Z_2, \cdots$. The set of centroids, C^*, is easily found to be M itself, since any point x in M makes $\int d(x, z)\, dp(z) \equiv \frac{1}{8} \sum_{z \in \{-1,1\}^3} d(x, z)$ equal to $\frac{24}{8}$. To find $\hat{C}_n$, let $S_{nk} := \sum_{i=1}^{n} U_{ik}$ $(k = 1, 2, 3)$ and let $S_n := (S_{n1}, S_{n2}, S_{n3})$. There are four mutually exclusive and jointly exhaustive events of interest:

$E_{n1} := \{S_n = (0, 0, 0)\}$,

$E_{n2} := \{S_n \in \{(t, 0, 0), (0, t, 0), (0, 0, t) : t \neq 0\}\}$,

$E_{n3} := \{S_n \in \{(s, t, 0), (s, 0, t), (0, s, t) : s, t \neq 0\}\}$ and

$E_{n4} := \{S_n = (r, s, t) \text{ where } r, s, t \neq 0\}$.

We easily convince ourself that $E_{n1} \Rightarrow \{\hat{C}_n = M\}$, $E_{n2} \Rightarrow \{\hat{C}_n = \text{some face of } M\}$, $E_{n3} \Rightarrow \{\hat{C}_n = \text{some edge of } M\}$ and $E_{n4} \Rightarrow \{\hat{C}_n = \text{some corner of } M\}$. It is then obvious that B equals $C^* = M$ if and only if E_{n1} occurs infinitely often, and if this is not the case then B at most equals the surface of M. Now $\Pr(E_{2n,1}) = \left[\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right]^3$, and using Stirling's formula we find $\Pr(E_{2n,1}) \approx (n\pi)^{-3/2}$. Since $\Pr(E_{n,1}) = 0$ when n is odd, we have $\sum_{n=1}^{\infty} \Pr(E_{n,1}) < \infty$ and the Borel-Cantelli lemma asserts that $\Pr(E_{n1}$ infinitely often$) = 0$, so $\Pr(B = C^*) = 0$. We can show similarly that $\Pr(B = \text{the whole surface of } M) = 1$. As regards the restricted case, though, we have $BR = CR^*$ (a.s.) which is the set of all corners of M.

PROOF OF THE THEOREM. Define

$$T_n(x) := \frac{1}{n}\sum_{i=1}^n d^r(x, Z_i) - \int d^r(x, z)\, dp(z),$$

$$T_n^*(x) := \frac{1}{n}\sum_{i=1}^n d^r(x, Z_i) - \int d^r(x^*, z)\, dp(z)$$

and

$$TR_n^*(x) := \frac{1}{n}\sum_{i=1}^n d^r(x, Z_i) - \int d^r(z^*, z)\, dp(z).$$

Clearly, by the ordinary strong law of large numbers for real-valued random variables,

$$T_n(x) \to 0 \quad \text{a.s.} \quad \text{for each } x \text{ in } M.$$

Furthermore, since M is compact, the results of Ranga Rao (1962) on uniform convergence can be applied, yielding

$$\sup_{x\in M} |T_n(x)| \to 0 \quad \text{a.s.},$$

which in turn implies

$$T_n(\hat{X}_n) \to 0 \quad \text{a.s.}$$

and

$$T_n(\hat{Z}_n) \to 0 \quad \text{a.s.}$$

By the definition of $\hat{X}_n$ and x^* we find

$$T_n(\hat{X}_n) \leq T_n^*(\hat{X}_n) \leq T_n(x^*),$$

so

$$|T_n^*(\hat{X}_n)| \leq \max\{|T_n(\hat{X}_n)|, |T_n(x^*)|\} \to 0 \quad \text{a.s.};$$

hence (i) is proved.

In the case of the restricted variation (iii), we first show

(1) $$\min_{1\leq i\leq n} | TR_n^*(Z_i) - TR_n^*(z^*) \mapsto 0 \quad \text{a.s.}$$

To this end, let

$$s(\delta) := \sup_{z\in M}\sup_{d(x,y)<\delta} |d^r(x, z) - d^r(y, z)|$$

and observe that $s(\delta) \to 0$ as $\delta \to 0$ due to the compactness of M. Easily, we find

$$\sup_{d(x,y)<\delta} |TR_n^*(x) - TR_n^*(y)| \leq s(\delta).$$

Furthermore, for any $\delta > 0$, $O_\delta := \{x \in M : d(x, z^*) < \delta\}$ is a set which must have *positive p*-measure α, else z^* wouldn't be in the support. Hence with probability at least $1 - (1 - \alpha)^n$ there is some Z_i, $i \leq n$, lying in O_δ. This implies

$$\limsup_n \min_{1\leq i\leq n} |TR_n^*(Z_i) - TR_n^*(z^*)| \leq s(\delta) \quad \text{a.s.}.$$

Letting $\delta \to 0$ then proves the assertion (1).

As we continue the proof of (iii), observe that $T_n(z^*)$ is equivalent to $TR_n^*(z^*)$. It is easy to see that

$$\min_{1\leq i\leq n} TR_n^*(Z_i) \leq T_n(z^*) + \min_{1\leq i\leq n} |TR_n^*(Z_i) - TR_n^*(z^*)|.$$

By the definition of $\hat{Z}_n$ and z^* we find

$$T_n(\hat{Z}_n) \leq TR_n^*(\hat{Z}_n) = \min_{1\leq i\leq n} TR_n^*(Z_i),$$

hence

$$|TR_n^*(\hat{Z}_n)| \leq \max\{|T_n(\hat{Z}_n)|, |T_n(z^*)| + \min_{1\leq i\leq n} |TR_n^*(Z_i) - TR_n^*(z^*)|\} \to 0 \quad \text{a.s.}$$

which proves (iii).

We now turn to proving the convergence properties of the sample centroids and the restricted sample centroids.

Writing

$$\left| \int d^r(\hat{X}_n, z)\, dp(z) - \int d^r(x^*, z)\, dp(z) \right| = |-T_n(\hat{X}_n) + T_n^*(\hat{X}_n)| \leq |T_n(\hat{X}_n)| + |T_n^*(\hat{X}_n)|$$

we see that

$$(2) \qquad \int d^r(\hat{X}_n, z)\, dp(z) \to \int d^r(x^*, z)\, dp(z) \quad \widetilde{\text{a.s.}}$$

Now let $X \in B$, say $X = \lim_k \hat{X}_{n_k}$ where $\hat{X}_n \in C_n$ ($\forall n$). Then $d^r(\hat{X}_{n_k}, z) \to_k d^r(X, z)$ for all z in M, so by the bounded convergence theorem we have

$$\int d^r(X, z)\, dp(z) = \lim_k \int d^r(\hat{X}_{n_k}, z)\, dp(z)$$

and by (2) the right-hand side equals $\int d^r(x^*, z)\, dp(z)$ $\widetilde{\text{a.s.}}$, so $X \in C^*$ $\widetilde{\text{a.s.}}$ and hence $B \subseteq C^*$ $\widetilde{\text{a.s.}}$

Proving that $BR \subseteq CR^*$ $\widetilde{\text{a.s.}}$ is quite analogous to the proof just presented for the nonrestricted centroids. □

Acknowledgment. The author is indebted to J. MacQueen for discussions leading to formulation of the problem.

REFERENCES

[1] MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* (LeCam, L. L. and Neyman, J. Eds.), 281–296.

[2] Padgett, W. J. and Taylor, R. L. (1973). *Laws of Large Numbers for Normed Linear Spaces and Certain Frechet Spaces. Lecture Notes in Math.* No. 360, Springer, Berlin.

[3] Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

Department of Statistics
Mathematics Institute
University of Oslo
Blindern
N-Oslo 3
Norway