Journal of Theoretical Probability, Vol. 17, No. 1, January 2004 (© 2004)

On a Statistical Framework for Estimation from Random Set Observations

De-Jun Feng[1,2] **and Ding Feng**[2,3]

Received June 5, 2002; revised June 26, 2003

Using the theory of random closed sets, we extend the statistical framework introduced by Schreiber[11] for inference based on set-valued observations from the case of finite sample spaces to compact metric spaces with continuous distributions.

KEY WORDS: Statistical inference; conditional distribution; random sets.

1. INTRODUCTION

This paper is about a theoretical framework for statistical inference based on imprecise observations. Let X be a random vector with unknown probability law μ_0 corresponding to the cumulative distribution function F on $\mathbb{R}^m$, and $X_1, X_2,..., X_n$ an i.i.d. F random sample. If the observations $X_1, X_2,..., X_n$ are observable, then the empirical dF_n based on this sample will be a good estimator of μ_0 $(= dF)$ for sufficiently large n. The inference problem about μ_0 becomes more delicate when we can only observe an i.i.d. random sample $S_1, S_2,..., S_n$ of a random set S such that $X_i \in S_i$, almost surely, $i = 1,..., n$. A statistical model for such imprecise observations was developed by Schreiber.[11] He only treated the case of finite sample spaces, i.e., when X takes values in a *finite subset* of $\mathbb{R}^m$. In this work, we shall elaborate and extend his work to *continuous random vectors*.

[1] Department of Mathematical Sciences, Tsinghua University, Beijing 100084, People's Republic of China. E-mail: dfeng@math.tsinghua.edu.cn

[2] Department of Mathematical Sciences, New Mexico State University, Las Cruces, New Mexico 88003-8001.

[3] To whom correspondence should be addressed. E-mail: dfeng@nmsu.edu

Statistical inference based on set-valued observations is best known in the statistical literature in *coarse data* situations such as missing data in multivariate analysis, censoring data in survival analysis, and grouped data in general. See, e.g., Heitjan and Rubin,[5] Gill *et al.*,[3] and van der Vaart and Wellner.[14] Though these works were concerned with statistical inference based on set-valued observations, their focus was on some special models where only observed data, namely the set-observations, are needed but not the distribution of the random set model. These models are called coarsening at random (CAR). The CAR model was introduced by Heitjan and Rubin[5] to describe a reasonable form of randomly grouped, censored, or missing data. The CAR assumption is popular, and applications abound. In the coarsening model that Heitjan and Rubin proposed, observations are not made in the sample space of the random vector of interest, but rather in its power set. The following is the definition of CAR (see, e.g., Heitjan and Rubin[5] or Gill *et al.*[3]). Suppose X is a random vector taking values in a finite set E. Let $\mathscr{E}$ denote the set of all subsets of E, and let S denote a random nonempty subset of E, i.e., S takes values in $\mathscr{E}\setminus\{\varnothing\}$. The random set S is a *coarsening* of X if, with probability 1, $X \in S$. Moreover, a coarsening S of X is called a *coarsening at random* (CAR) if the conditional distribution of S given $X = x$ satisfies the following coarsened at random (CAR) assumption:

$$P(S = A \mid X \in A) = P(S = A \mid X = x), \qquad \forall A \in \mathscr{E}\setminus\{\varnothing\}, \quad x \in A.$$

The basic existence result for the CAR model is the following (cf. Ref. 3): let S be a random nonempty set with distribution f on $\mathscr{E}\setminus\{\varnothing\}$. Then there exist CAR probabilities π: $\mathscr{E}\setminus\{\varnothing\} \to [0, 1]$ and a probability distribution p on E such that $f(A) = p(A)\,\pi(A)$ for any $A \in \mathscr{E}\setminus\{\varnothing\}$. Furthermore, the above $p(A)$ and $\pi(A)$ are uniquely determined if $f(A) > 0$. We remark that these CAR models are proved to exist only in the case of finite populations.

Another approach for statistical inference based on set-valued observations relies on the concept of selectors for random sets (see, e.g., Molchanov,[7] Scheiber,[11] and Norberg[9]), where a random vector X is said to be a *selector* of a random set S if $X \in S$ almost surely. Note that S is a coarsening of X iff X is a selector of the random set S.

Given a random set S, it is important to characterize the class of all its possible selectors, which corresponds to the class of all the possible distributions of the true outcome of the experiment.

Since the existence result for CAR models has so far only been established in the case of finite populations, in order to develop a framework for inference based on set-valued observations in the continuous case, we must elaborate the second approach, which is based on the theory of random sets developed by Matheron.[6]

Our statistical setup is a generalization of a statistical model in which the data provide only a sequence of *empirical capacity functionals* T_n, rather than a sequence of empirical measures. We shall view the random vector X as an *almost sure selector* of a random set S, which gives rise to the observations $S_1, S_2,..., S_n$. To analyze the unknown distribution μ_0 of X, we shall study the core of $T^{(n)}$. In the process we shall establish some convergence theorems for $T^{(n)}$, which will permit us in some cases to estimate μ_0 by a sequence $\mu_n \in T^{(n)}$. We shall also provide large deviation and central limit theorem type results about the rate of convergence of $T^{(n)}$. The reader may find all the definitions and formulated theorems in Section 3.

2. THE STATISTICAL PROBLEM

The statistical inference problem based on set-valued observations is as follows.

Let X be a random vector defined on some probability space $(\Omega, \mathscr{A}, P)$ with values in $\mathbb{R}^m$. The probability law μ_0 of X is the probability measure PX^{-1} on $\mathscr{B}(\mathbb{R}^m)$, the collection of all Borel subsets of $\mathbb{R}^m$.

Let $X_1, X_2,..., X_n$ be an i.i.d. random sample drawn from X. The Glivenko–Cantelli theorem asserts that μ_0 can be estimated consistently by the empirical measures dF_n, where

$$F_n(x) = \frac{1}{n} \#\{1 \leqslant i \leqslant n : X_i \leqslant x\}$$

is the empirical distribution based on the sample $X_1, X_2,..., X_n$, and # denotes the cardinality. In other words, with probability 1, for any $\epsilon > 0$ for all n sufficiently large, μ_0 is in the ϵ-neighborhood of dF_n defined by the supremum norm taken over an appropriate subset of the Borel subsets. See van der Vaart and Wellner[13] for details.

Suppose that we cannot observe the X_i's directly, but instead, we observe random sets $S_1, S_2,..., S_n$ with $X_i \in S_i$, $i = 1, 2,..., n$. In this situation, it is clear that in order to construct an estimator of μ_0 based on $S_1, S_2,..., S_n$, we must have an appropriate model. Schreiber[11] assumed that the observed sets $S_1, S_2,..., S_n$ are an i.i.d. sample from a random set S and the random vector X is an almost sure selector of S. We are going to extend Schreiber's work from finite sample spaces to *compact metric spaces* with a continuous distribution. In future work, we shall investigate the more general case of locally compact spaces like $\mathbb{R}^d$.

Thus, we consider now a random vector X taking values in a compact metric space (Y, d), such as $[0, 1]$ with the usual Euclidean metric, i.e., X is $\mathscr{A}$–$\mathscr{B}(Y)$-measurable, where $\mathscr{B}(Y)$ is the Borel-σ-field on Y generated by the metric d.

By a random set S, we mean a random closed set in Y in the sense of Matheron.[6] In fact for our statistical situation here, S will be a nonempty random closed set. Also since Y is compact, S is a compact random set. In view of Choquet's theorem (see Matheron[6]) the random evolution of S is characterized by its capacity functional T.

Recall that the core of T is defined by

$$\text{core}(T) = \{\mu \in \mathscr{M}(Y) : \mu \preccurlyeq T\},$$

where $\mathscr{M}(Y)$ is the collection of all Borel probability measures on Y, and we write $\mu \preccurlyeq T$ if $\mu(K) \leqslant T(K)$ for all $K \in \mathscr{K}$. Here $\mathscr{K}$ denotes the collection of all nonempty compact subsets of Y. In Section 3 the reader may find some examples of capacity functionals and their cores.

It is well known that $\text{core}(T) \neq \varnothing$ if Y is compact metric space (see, e.g., p. 102 of Molchanov[7]). Combining this with a result of Norberg[9] (see Proposition 6.1), we know that for every random set S, there exists a random vector X such that $P(X \in S) = 1$. This X is said to be an almost sure selector of S. Let us elaborate a little more on this. Probabilistic models are proposed to model observed data when uncertainty is present. Depending on the type of observed data, statistical procedures are derived in order to make inference about the random phenomenon under study. In the foreword to the pioneering work on random sets of G. Matheron,[6] G. Watson wrote "Modern statistics might well be defined as the application of computers and mathematics to data analysis. It must grow as new types of data are considered and as computing technology advances." Set-valued observations are an example of a new type of data, generalizing point-valued observations in standard statistical applications. They arise in several different contexts. Traditionally, statistics of random sets was investigated to study random patterns of objects such as the Boolean model (Matheron,[6] Molchanov[8]). Here, random closed sets in Euclidean spaces are used to model observed sets (as a generalization of point processes), and the associated statistics is concerned with the inference about various parameters of the random patterns under study such as the expected area, the expected perimeter, and the distribution of the random set model. Note that random set data can arise even in the standard framework of multivariate statistics. This is exemplified by the problem of probability density estimation using Hartigan's excess mass approach (Hartigan,[4] Polonik[10]), where random sets are used to estimate α-level sets of the unknown density function. In Biostatistics, set-valued observations arise as coarse data (Heitjan and Rubin[5] and Gill *et al.*[3]). Here, the random vector of interest X is not observable, but instead, one observes the values of some random set S containing X almost surely. From a modeling point of view, X is an almost sure selector of S, i.e., $P(X \in S) = 1$. In the above cited works in Biostatistics, the emphasis is on models of S which make inference about X feasible. This is the essence of the CAR model. In a related direction, Schreiber[11] set out to investigate a general framework for inference with set-valued observations. His result were for the case when the random vector of interest X takes a finite number of values. Our present work is an extension of Schreiber's framework to the continuous case. As in Schreiber's work, our emphasis is on models based upon the capacity functional T of the observed random set S. In Section 3, we shall give some examples illustrating capacity functionals and their cores.

Now, back to our general framework, the empirical capacity functional $T^{(n)}$ based on the i.i.d. random set sample $S_1, S_2, \ldots, S_n$ is defined on $\mathscr{K}$

$$T^{(n)}(K) = \frac{1}{n} \#\{1 \leqslant i \leqslant n : S_i \cap K \neq \varnothing\}.$$

Clearly by the strong law of large numbers, $T^{(n)}(K) \to T(K)$ almost surely as $n \to \infty$ for any $K \in \mathscr{K}$. Note that $dF_n \in \text{core}(T^{(n)})$ a.s. for any n.

The counterpart of the empirical measure dF_n is the core$(T^{(n)})$, which is a subset of $\mathscr{M}(Y)$. Since $\mu_0 \in \text{core}(T)$ our basic result concerning the estimation of μ_0 based on $S_1, S_2, \ldots, S_n$ relies on the approximation of core(T) by core$(T^{(n)})$. We shall show that the rate of convergence of core$(T^{(n)})$ to core(T) is exponential.

For statistical considerations, we assume, as in Schreiber,[11] that the unknown μ_0 belongs to an *a priori* known class Ξ of probability measures on $\mathscr{B}(Y)$. In the special case when $\Xi \cap \text{core}(T) = \{\mu_0\}$, our analysis of the approximation of core(T) by core(T_n) will lead to a consistent estimator of μ_0.

3. NOTATIONS AND MAIN RESULTS

Let (Y, d) be a compact metric space. For any $\epsilon > 0$ and $E \subset Y$, let $B_\epsilon(E)$ denote the ϵ-*neighborhood* of E in Y. That is,

$$B_\epsilon(E) := \{y \in Y : \exists x \in E \text{ with } d(x, y) < \epsilon\}.$$

For simplicity, we denote $B_\epsilon(y) = B_\epsilon(\{y\})$ for $y \in Y$.

Let $\mathscr{K} := \mathscr{K}(Y)$ be the collection of all nonempty compact subsets of Y. Endow $\mathscr{K}$ with the Hausdorff metric ρ_d defined by

$$\rho_d(E, F) = \inf\{\epsilon \geqslant 0 : B_\epsilon(F) \subseteq E, B_\epsilon(E) \subseteq F\}, \qquad \forall E, F \in \mathscr{K}.$$

By Blaschke selection theorem, $(\mathscr{K}, \rho_d)$ is also a compact metric space (see, e.g., Falconer,[2] Theorem 3.16). Let $\mathscr{B}(Y)$ and $\mathscr{B}(\mathscr{K})$ denote the collections of Borel sets in (Y, d) and $(\mathscr{K}, \rho_d)$, respectively.

Let $\mathscr{M}(Y)$ be the collection of all Borel probability measures on Y, and $C(Y)$ the space of all continuous real functions on Y endowed with the uniform topology. Since Y is compact, there exists a sequence $\{f_i\}$ of continuous real functions dense in $C(Y)$. Define a metric Δ on $\mathscr{M}(Y)$ by

$$\Delta(\mu, \nu) = \sum_{i=1}^{\infty} \frac{|\int f_i \, d\mu - \int f_i \, d\nu|}{2^i \, \|f_i\|}, \tag{3.1}$$

where $\|f\| := \max_{y \in Y} |f(y)|$. It is well known that the metric Δ on $\mathscr{M}(Y)$ gives the weak-star topology, and $(\mathscr{M}(Y), \Delta)$ is a compact space (see, e.g., Walters,[15] Theorems 6.4 and 6.5).

Let $\mathscr{K}(\mathscr{M}(Y))$ be the collection of all nonempty compact subsets of $\mathscr{M}(Y)$. Endow $\mathscr{K}(\mathscr{M}(Y))$ with the Hausdorff metric ρ_Δ. (The definition of ρ_Δ is analogous to that of ρ_d.) Again, $(\mathscr{K}(\mathscr{M}(Y)), \rho_\Delta)$ is a compact metric space. Similarly we use $\mathscr{B}(\mathscr{K}(\mathscr{M}(Y)))$ denote the collection of all Borel sets in $\mathscr{K}(\mathscr{M}(Y))$.

Denote by $\mathscr{A}(Y)$ the class of all nonempty sets $E \subset Y$ such that

$$\{K \in \mathscr{K} : K \cap E \neq \varnothing\} \in \mathscr{B}(\mathscr{K}).$$

We shall see that the class $\mathscr{A}(Y)$ contains all the nonempty compact sets and open sets in (Y, d) (see Lemma 4.1). However, we don't know if $\mathscr{A}(Y) \supseteq \mathscr{B}(Y)$ in this general setting.

A *random set* S is a map defined on a probability space $(\Omega, \mathscr{F}, P)$ taking values in $\mathscr{K}$, and measurable with respect to $\mathscr{F}$–$\mathscr{B}(\mathscr{K})$. The *capacity functional* of S, denoted as T_s or simply T, is defined by

$$T(E) = P\{\omega \in \Omega : S(\omega) \cap E \neq \varnothing\}, \qquad \forall E \in \mathscr{A}(Y).$$

The *core* of T is defined as

$$\text{core}(T) = \{\mu \in \mathscr{M}(Y) : \mu \preccurlyeq T\},$$

where we write $\mu \preccurlyeq T$ if $\mu(K) \leqslant T(K)$ for all $K \in \mathscr{K}$. The concept of capacity functional is a natural generalization of probability measures. Taking S

to be singleton-valued, we know that every probability measure on Y is just a capacity functional, whose core just consists of the measure itself.

Let us first give two examples illustrating the capacity functionals and their cores.

Example 3.1. Take $Y = \{0, 1\}$. Let S be a random set with the distribution

$$P(S = \{0\}) = P(S = \{1\}) = P(S = \{0, 1\}) = \tfrac{1}{3}.$$

Let T be the capacity functional of S. Then we can check that

$$T(\{0\}) = 2/3; \qquad T(\{1\}) = 2/3; \qquad \text{and} \qquad T(\{0, 1\}) = 1.$$

The core of T is given by

$$\operatorname{core}(T) = \{\mu\colon 1/3 \leqslant \mu(\{0\}) \leqslant 2/3,\ \mu(\{1\}) = 1 - \mu(\{0\})\}.$$

Example 3.2. Take $Y = [0, M]$, where M is a fixed positive number. Let X and V be two independent random vectors taking values in Y. Define a random set S by

$$S = \begin{cases} [0, V], & \text{if} \quad X \leqslant V \\ [V, M], & \text{otherwise.} \end{cases}$$

This is a simple version of the model studied in van der Vaart and Wellner.[14] Now let T be the capacity functional of S, and μ_X, μ_V be the distributions of X, V, i.e., $\mu_X = PX^{-1}$ and $\mu_V = PV^{-1}$. For any nonempty compact set $E \subset Y$, we have

$$\begin{aligned} T(E) &= P(S \cap E \neq \varnothing) \\ &= P(S \cap E \neq \varnothing, X \leqslant V) + P(S \cap E \neq \varnothing, X > V) \\ &= P([0, V] \cap E \neq \varnothing, X \leqslant V) + P([V, M] \cap E \neq \varnothing, X > V) \\ &= P(\min E \leqslant V, X \leqslant V) + P(\max E \geqslant V, X > V) \\ &= \int_{\min E}^{M} \mu_X([0, t])\, d\mu_V(t) + \int_0^{\max E} \mu_X((t, M])\, d\mu_V(t). \end{aligned}$$

It is interesting to see that $T(E)$ depends only on $\min E$ and $\max E$. The core of T, is just

$$\operatorname{core}(T) = \left\{ \mu \in \mathscr{M}(Y) : \begin{array}{l} \mu([a, b]) \leqslant \int_a^M \mu_X([0, t])\, d\mu_V(t) \\ \qquad + \int_0^b \mu_X((t, M])\, d\mu_V(t), \forall [a, b] \subset Y \end{array} \right\}.$$

Especially we have $\mu_X \in \text{core}(T)$. To see this, we have for all $[a, b] \subseteq [0, M]$,

$$\int_a^M \mu_X([0, t])\, d\mu_V(t) + \int_0^b \mu_X((t, M])\, d\mu_V(t)$$

$$= \int_0^a \mu_X((t, M])\, d\mu_V(t) + \int_a^b 1\, d\mu_V(t) + \int_b^M \mu_X([0, t])\, d\mu_V(t)$$

$$\geqslant \mu_X([a, b]).$$

It is well known that in our general setting, core(T) is a nonempty compact subset of $\mathscr{M}(Y)$. In this paper we shall focus on the study of the perturbation property of core(T). We first define a pseudo-metric on the space of all capacity functionals on $\mathscr{A}(Y)$. Then we prove that core(T) depends continuously upon T. This result has some applications. Especially, we can use it to prove a convergence property of the empirical capacity functional, which generalizes a result of Schreiber.[11]

Now let us first define a pseudo-metric on the space of all capacity functionals on $\mathscr{A}(Y)$. A set $E \subset Y$ is called a *ϵ-spanning set* of Y if $B_\epsilon(E) = Y$. By the compactness of Y, for each $\epsilon > 0$ there exists a ϵ-spanning set consisting of finitely many points. For $n \geqslant 1$, we choose a $\frac{1}{n}$-spanning set H_n of Y such that H_n is a finite set. Define

$$\mathcal{O}_n = \{B_{\frac{1}{n}}(E) \colon E \subset H_n\}, \qquad n \geqslant 1. \tag{3.2}$$

For two capacity functionals T and T', define

$$\Lambda(T, T') = \sum_{n \geqslant 1} 2^{-n-\# H_n} \sum_{W \in \mathcal{O}_n} |T(W) - T'(W)|. \tag{3.3}$$

It is easy to show that Λ is a pseudo-metric. Furthermore we can show (see Proposition 4.5).

$$\Lambda(T, T') = 0 \Leftrightarrow T(K) = T'(K), \qquad \forall K \in \mathscr{K}.$$

This means that Λ is a metric restricted on $\mathscr{K}$.

Remark. As we have previously mentioned, any probability measure on Y can be viewed as a capacity functional. Restricted to $\mathscr{M}(Y)$, (3.3) becomes

$$\Lambda(\mu, \mu') = \sum_{n \geqslant 1} 2^{-n-\# H_n} \sum_{W \in \mathcal{O}_n} |\mu(W) - \mu'(W)|.$$

It is not hard to check that if $\Delta(\mu_k, \mu) \to 0$, then $\rho_\Delta(\mu_k, \mu) \to 0$; but the converse is not true.

Now, we can formulate one of our main results:

Theorem 3.3. Suppose Y is an arbitrary compact space. Let T_1 and T_2 be two capacity functionals on $\mathscr{A}(Y)$. Then for any $n \geqslant 1$,

$$\rho_\Delta(\mathrm{core}(T_1), \mathrm{core}(T_2)) \leqslant \#H_n \cdot 4^{\#H_n} \cdot \max_{W \in \mathcal{O}_n} |T_1(W) - T_2(W)| + 2 \sum_{i=1}^{\infty} \frac{C_{f_i}\left(\frac{6}{n}\right)}{2^i \|f_i\|}, \tag{3.4}$$

where $\mathcal{O}_n$ is defined as in (3.2), the sequence $\{f_i\}$ is given as in (3.1), and $C_f(\epsilon) = \sup\{|f(x) - f(y)|: d(x, y) \leqslant \epsilon\}$.

The proof of the above theorem will be given in Section 6. As a direct corollary, we have

Corollary 3.4. Suppose Y is an arbitrary compact space. Let T_k $(k = 1, 2, \ldots)$ and T be capacity functionals on $\mathscr{A}(Y)$ satisfying $\lim_{k \to \infty} \Delta(T_k, T) = 0$. Then

$$\lim_{k \to \infty} \rho_\Delta(\mathrm{core}(T_k), \mathrm{core}(T)) = 0.$$

Remark. Under the condition of Corollary 3.4, we deduce from the definition of the Hausdorff metric that, for any $\mu \in \mathrm{core}(T)$, there exists a sequence $\mu_k \in \mathrm{core}(T_k)$ such that $\Delta(\mu_k, \mu) \to 0$ and thus μ_k converges to μ in the weak-star topology.

The above results have an important application in the analysis of the convergence property of empirical capacity functionals.

Now suppose T is the capacity functional on $\mathscr{A}(Y)$ of a random set $S: \Omega \to \mathscr{K}$. Let $\{S_i\}$ be a sequence of i.i.d. random sets with the same distribution as S. For each $\omega \in \Omega$, define a sequence of set functions $\chi_i(\omega, \cdot)$ on $\mathscr{A}(Y)$ by

$$\chi_i(\omega, E) = \begin{cases} 1, & \text{if } S_i(\omega) \cap E \neq \varnothing, \\ 0, & \text{otherwise.} \end{cases} \qquad \forall E \in \mathscr{A}(Y), \tag{3.5}$$

and define $T_\omega^{(n)}$ $(n \in \mathbb{N})$ on $\mathscr{A}(Y)$ by

$$T_\omega^{(n)}(E) = \frac{1}{n} \sum_{i=1}^{n} \chi_i(\omega, E) = \frac{1}{n} \#\{1 \leqslant i \leqslant n, S_i(\omega) \cap E \neq \varnothing\}. \tag{3.6}$$

$T_\omega^{(n)}$ is called the nth *empirical capacity functional* based on $\{S_i\}$. In fact, for each ω, $T_\omega^{(n)}$ is really the capacity functional of a random set. To see this, define a probability space $(\Omega_n, 2^{\Omega_n}, P_n)$ by $\Omega_n = \{1, 2,\ldots, n\}$ and

$$P_n(E) = \frac{\#E}{n}, \qquad \forall E \subset \Omega_n.$$

The random set $V_\omega: \Omega_n \to \mathscr{K}$ is defined by $V_\omega(i) = S_i(\omega)$, $\forall i \in \Omega_n$. One may check that

$$T_\omega^{(n)}(E) = P_n\{i: V_\omega(i) \cap E \neq \varnothing\}, \qquad \forall E \in \mathscr{A}(Y).$$

That is, $T_\omega^{(n)}$ is the capacity functional of V_ω. The following result gives a natural relation between empirical measures and empirical capacity functions.

Theorem 3.5. Suppose T is the capacity functional on $\mathscr{A}(Y)$ of a random set $S: \Omega \to \mathscr{K}$. Let X be a random vector satisfying $P(X \in S) = 1$. Let $\{S_n\}$ be a sequence of i.i.d. random sets with the same distribution as S, and X_i a sequence of i.i.d. random vectors with the same distribution as X satisfying $P(X_i \in S_i) = 1$. Let $\mu_\omega^{(n)}$ denote the empirical measure based on $X_1(\omega),\ldots, X_n(\omega)$, and $T_\omega^{(n)}$ the empirical capacity functional based on $S_1(\omega),\ldots, S_n(\omega)$. Then

$$\mu_\omega^{(n)} \in \mathrm{core}(T_\omega^{(n)}), \quad \text{a.s. } \omega.$$

Proof. Let $(\Omega_n, 2^{\Omega_n}, P_n)$ and V_ω defined as in the last paragraph. It is clear that

$$\mu_\omega^{(n)} = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i(\omega)},$$

where δ_y denotes the Dirac measure at y. Set

$$v_\omega(i) = X_i(\omega), \qquad \forall i \in \Omega_n.$$

Then v_ω is a random variable on Ω_n which has the distribution $\mu_\omega^{(n)}$. Note that $v_\omega \in V_\omega$ for a.s. ω. We have

$$\mu_\omega^{(n)} \in \mathrm{core}(T_\omega^{(n)}), \quad \text{a.s. } \omega.$$

This finishes the proof of the theorem. □

In the following theorem we give a rate estimate of the convergence of empirical capacity functionals.

Theorem 3.6. Suppose Y is an arbitrary compact metric space, T is the capacity functional on $\mathscr{A}(Y)$ of a random set $S\colon \Omega \to \mathscr{K}$, and $\{T_\omega^{(n)}\}_{n\geqslant 1}$ is a sequence of empirical capacity functionals generated by a sequence of i.i.d. random sets $\{S_n\}_{n\geqslant 1}$ (with the same distribution as S). Then $\lim_n \rho_\Delta(\mathrm{core}(T_\omega^{(n)}), \mathrm{core}(T)) = 0$ almost surely. Moreover, for any $\epsilon > 0$, there exist n_ϵ and $L_\epsilon > 0$ such that

$$P\{\omega \in \Omega : \rho_\Delta(\mathrm{core}(T_\omega^{(n)}), \mathrm{core}(T)) > \epsilon\} \leqslant e^{-nL_\epsilon}$$

for all $n \geqslant n_\epsilon$.

The proof of Theorem 3.6 will be given in Section 7. We remark that under the conditions of this theorem, for any $\mu_0 \in \mathrm{core}(T)$ and for almost all $\omega \in \Omega$, there exists a sequence $\mu_n(\omega) \in \mathrm{core}(T_\omega^{(n)})$, $n \geqslant 1$, such that $\mu_n(\omega)$ converges to μ in the weak-star topology. Moreover, the above $\mu_n(\omega)$ can clearly be constructed using the steps in the proofs of Lemmas 6.3 and 6.4.

Let Ξ be a nonempty subset of $\mathscr{M}(Y)$ considered as a statistical model in Schreiber.[11] For any capacity functional T on $\mathscr{A}(Y)$, define

$$\hat{\Delta}(T \mid \Xi) = \inf\{\Delta(\mu, v)\colon \mu \in \mathrm{core}(T), v \in \Xi\}.$$

For two capacity functionals T_1 and T_2, it is easy to check that

$$|\hat{\Delta}(T_1 \mid \Xi) - \hat{\Delta}(T_2 \mid \Xi)| \leqslant \rho_\Delta(\mathrm{core}(T_1), \mathrm{core}(T_2)).$$

Thus we have

Corollary 3.7. Under the conditions of Theorem 3.6, we have for any $\varnothing \neq \Xi \subset \mathscr{M}(Y)$, $\lim_n \hat{\Delta}(T_\omega^{(n)} \mid \Xi) = \hat{\Delta}(T \mid \Xi)$ almost surely. Moreover, for any $\epsilon > 0$, there exist n_ϵ and $L_\epsilon > 0$ such that

$$P(\omega \in \Omega : |\hat{\Delta}(T_\omega^{(n)} \mid \Xi) - \hat{\Delta}(T \mid \Xi)| > \epsilon) \leqslant e^{-nL_\epsilon}$$

for all $n \geqslant n_\epsilon$.

We remark that Corollary 3.7 has been proved by Schreiber[11] in the case that Y consists of finitely many points. This result could be used to test certain assumptions about the coarsening mechanism, when we consider the random set S as a coarsening of some random vector X with $P(X \in S) = 1$. For example, if we have a model $\Xi \subset \mathscr{M}(Y)$ for the distribution of X, then under this model assumption we must have for the capacity functional T of S

$$\hat{\Delta}(T \mid \Xi) = 0.$$

Then we could use Corollary 3.7 to test this hypothesis.

As suggested by one of the referees, the central limit theorem can be used to give a further characterization of the convergence rate of empirical capacity functionals. Let $\mathbb{M}(Y)$ denote the dual of $C(Y)$. The metric ρ_Δ can be passed over to $\mathbb{M}(Y)$ by

$$\rho_\Delta(G_1, G_2) = \sum_{i=1}^{\infty} \frac{|(G_1 - G_2)(f_i)|}{2^i \, \|f_i\|},$$

where f_i is given as in (3.1). This metric generates the weak-star topology on $\mathbb{M}(Y)$. The measure space $\mathscr{M}(Y)$ is just a convex subspace of $\mathbb{M}(Y)$. Combining the central limit theorem and Theorem 3.5, we have

Theorem 3.8. Under the conditions of Theorem 3.6, for every $\mu \in \operatorname{core}(T)$ and every $\epsilon > 0$, there exist N and n_0 such that with probability greater than $1 - \epsilon$, for every $n > n_0$ we can find $\mu_n \in \operatorname{core}(T_\omega^{(n)})$ with

$$\rho_\Delta(\sqrt{n}(\mu_n - \mu), 0) < N.$$

The proof of this theorem will be given in Section 7. As a direct corollary we have

Corollary 3.9. Under the conditions of Theorem 3.6, for every $\mu \in \operatorname{core}(T)$, every $\alpha < 1/2$, every $\epsilon > 0$ and every $\delta > 0$, there exist n_0 such that with probability greater than $1 - \epsilon$, for every $n > n_0$ we can find $\mu_n \in \operatorname{core}(T_\omega^{(n)})$ with

$$\rho_\Delta(n^\alpha(\mu_n - \mu), 0) < \delta.$$

4. SOME PROPERTIES OF CAPACITY FUNCTIONALS

In this section, we present some necessary properties of $\mathscr{A}(Y)$ and capacity functionals. Most of them are already known.

Lemma 4.1.

(i) If $\{E_n\}$ is an increasing sequence of sets in $\mathscr{A}(Y)$, then $\lim_n E_n \in \mathscr{A}(Y)$.

(ii) $E \in \mathscr{A}(Y)$ for any $E \in \mathscr{K}$.

(iii) $E \in \mathscr{A}(Y)$ for each open subset E of Y.

Proof. Note that if $\{E_n\}$ is an increasing sequence of sets in $\mathscr{A}(Y)$, then

$$\{K \in \mathscr{K} : K \cap \lim_n E_n \neq \varnothing\} = \lim_n \{K \in \mathscr{K} : K \cap E_n \neq \varnothing\}, \tag{4.1}$$

from which (i) follows. For (ii) and (iii), see Sections 1.1, 1.2, and 2.1 of Matheron.[6] □

Now suppose T is the capacity functional of a random set S. We begin with an elementary lemma.

Lemma 4.2.

(i) If $\{E_n\}$ is an increasing sequence of sets in $\mathscr{A}(Y)$, then $T(\lim_n E_n) = \lim_n T(E_n)$.

(ii) If $\{E_n\}$ is a decreasing sequence of sets in $\mathscr{K}$, then $T(\lim_n E_n) = \lim_n T(E_n)$.

Proof. To see (i), note that if $\{E_n\}$ is an increasing sequence of sets in $\mathscr{A}(Y)$,

$$\begin{aligned} T(\lim_n E_n) &= P(S^{-1}(\{K \in \mathscr{K} : K \cap \lim_n E_n \neq \varnothing\})) \\ &= P\left(\bigcup_n S^{-1}(\{K \in \mathscr{K} : K \cap E_n \neq \varnothing\})\right) \\ &= \lim_n P(S^{-1}(\{K \in \mathscr{K} : K \cap E_n \neq \varnothing\})) \\ &= \lim_n T(E_n). \end{aligned}$$

For (ii), see Sections 1.1 and 1.2 of Molchanov.[7]

As a corollary of Lemma 4.2, we have

Corollary 4.3. For any $E \in \mathscr{K}$, we have

$$\lim_{\epsilon \to 0} T(\overline{B_\epsilon(E)}) = \lim_{\epsilon \to 0} T(B_\epsilon(E)) = T(E),$$

where $\overline{B_\epsilon(E)}$ denotes the closure of $B_\epsilon(E)$.

Proof. It suffices to show $\lim_{n\to\infty} T(\overline{B_{\frac{1}{n}}(E)}) = T(E)$. To see this, note that $\{\overline{B_{\frac{1}{n}}(E)}\}$ is a decreasing sequence of compact sets in Y with $\lim_n \overline{B_{\frac{1}{n}}(E)} = E$. It follows from Lemma 4.2 that $\lim_n T(\overline{B_{\frac{1}{n}}(E)}) = T(E)$. This completes the proof. □

The following fact should be mentioned.

Proposition 4.4. If $\mu \in \mathrm{core}(T)$, then for any $E \in \mathscr{B}(Y)$, there exists $F \subseteq E$ with $F \in \mathscr{A}(Y)$ such that $\mu(E) \leqslant T(F)$.

Proof. Since Y is compact, μ is a Radon measure. Thus for any $E \in \mathscr{B}(Y)$, there exists an increasing sequence of compact sets $K_n \subset E$ with $\mu(E) = \lim_n \mu(K_n) = \mu(\lim_n K_n)$. Denote by $K = \lim_n K_n$. By Lemma 4.1, we have $K \in \mathscr{A}(Y)$. By Lemma 4.2, we have

$$T(K) = \lim_n T(K_n) \geqslant \lim_n \mu(K_n) = \mu(E). \qquad \square$$

Proposition 4.5. $\Lambda(T, T') = 0 \Leftrightarrow T(K) = T'(K), \forall K \in \mathscr{K}$.

Proof. First we prove "⇐." Assume $T(K) = T'(K)$ for all $K \in \mathscr{K}$. For any $W \in \mathcal{O}_n$, there exists an increasing sequence of sets $K_i \in \mathscr{K}$ such that $W = \lim_i K_i$. By Lemma 4.2,

$$T(W) = \lim_i T(K_i) = \lim_i T'(K_i) = T'(W),$$

from which $\Lambda(T, T') = 0$ follows.

Now we prove "⇒." Assume $\Lambda(T, T') = 0$, that is, $T(W) = T'(W)$ for any n and $W \in \mathcal{O}_n$. For any $K \in \mathscr{K}$ define $E_n = B_{\frac{1}{n}}(K) \cap H_n$ and $W_n = B_{\frac{1}{n}}(E_n)$. It can be checked that $W_n \in \mathcal{O}_n$ and $W_n \supset K$. Moreover the sequences $\{\overline{W_n}\}$ and $\{W_n\}$ are decreasing and they converge to K. Therefore

$$T(K) = \lim_n T(\overline{W_n}) = \lim_n T(W) = \lim_n T'(W) = \lim_n T'(\overline{W_n}) = T'(K).$$

This finishes the proof. □

5. A USEFUL PROPOSITION

The main result of this section is the following proposition.

Proposition 5.1. Suppose $(\Omega, \mathscr{A}, P)$ is a probability space, and $S\colon \Omega \to \mathscr{K}$, $y\colon \Omega \to Y$ are $\mathscr{A}$–$\mathscr{B}(\mathscr{K})$ and $\mathscr{A}$–$\mathscr{B}(Y)$ measurable respectively. Assume there exists $\epsilon > 0$ such that

$$d(y(\omega), S(\omega)) < \epsilon, \qquad \forall \omega \in \Omega.$$

Then there exists a $\mathscr{A}$–$\mathscr{B}(Y)$ measurable map x from Ω to Y such that

$$x(\omega) \in S(\omega) \quad \text{and} \quad d(x(\omega), y(\omega)) < 6\epsilon, \qquad \forall \omega \in \Omega.$$

We shall develop a technical tool to prove the above result. This tool will also be applied later in the proof of Theorem 3.3. Also, at the end of this section we shall use it to provide a new proof of the non-emptiness of core(T).

Let us first introduce some notations. For any $n \geqslant 1$, let H_n be the finite $\frac{1}{n}$-spanning set of Y defined as in Section 3. Define $\pi_n: 2^Y \to 2^{H_n}$ by

$$\pi_n(E) = \left\{ x \in H_n : \exists y \in E \text{ with } d(x, y) < \frac{1}{n} \right\}, \qquad \forall E \subset Y. \tag{5.1}$$

The set $\pi_n(E)$ may be considered as the projection of E onto H_n. Since H_n is a finite set, there exists a map θ_n from 2^{H_n} (the class of all subsets of H_n) to H_n such that $\theta_n(E) \in E$ for any $\varnothing \neq E \subset H_n$.

Lemma 5.2. For each $n \geqslant 1$,

(i) $\pi_n|_{\mathscr{K}}$ (the restriction of π_n on $\mathscr{K}$) is measurable from $\mathscr{K}$ to H_n.

(ii) θ_n is continuous from 2^{H_n} to H_n.

(iii) For any $E \subset H_n$ and $\epsilon > 0$, the set $\{x \in Y : \pi_n(\overline{B_\epsilon(x)}) = E\}$ is a measurable subset of Y.

Proof. To see part (i), note that for each nonempty subset $E \subset H_n$,

$$\pi_n|_{\mathscr{K}}(K) = E \Leftrightarrow d(x, K) < \frac{1}{n} \quad \text{for any} \quad x \in E \quad \text{and}$$

$$d(y, K) \geqslant \frac{1}{n} \quad \text{for any} \quad y \in H_n \setminus E.$$

Therefore

$$(\pi_n|_{\mathscr{K}})^{-1}(E) = \left(\bigcap_{x \in E} \left\{ K \in \mathscr{K} : d(x, K) < \frac{1}{n} \right\} \right)$$
$$\cap \left(\bigcap_{y \in H_n \setminus E} \left\{ K \in \mathscr{K} : d(y, K) \geqslant \frac{1}{n} \right\} \right).$$

Note that for each $x \in H_n$, the set $\{K \in \mathscr{K} : d(x, K) < \frac{1}{n}\}$ is an open set in $\mathscr{K}$, while $\{K \in \mathscr{K} : d(x, K) \geqslant \frac{1}{n}\}$ is closed in $\mathscr{K}$. Thus $(\pi_n|_{\mathscr{K}})^{-1}(E)$ is $\mathscr{B}(\mathscr{K})$ measurable.

Part(ii) is obvious. The proof of part (iii) is similar to that of part (i). To be more precise, note that

$$\{x \in Y : \pi_n(\overline{B_\epsilon(x)}) = E\}$$
$$= \left(\bigcap_{y \in E} \left\{x \in Y : d(y, \overline{B_\epsilon(x)}) < \frac{1}{n}\right\}\right) \cap \left(\bigcap_{z \in H_n \setminus E} \left\{x \in Y : d(z, \overline{B_\epsilon(x)}) \geqslant \frac{1}{n}\right\}\right).$$

Observing that $\{x \in Y : d(y, \overline{B_\epsilon(x)}) < \frac{1}{n}\}$ is open in Y and $\{x \in Y : d(z, \overline{B_\epsilon(x)}) \geqslant \frac{1}{n}\}$ are closed in Y, we obtain the desired result. □

Proof of Proposition 5.1. Let k_0 be the integer so that $\epsilon \leqslant 2^{-k_0} < 2\epsilon$. We construct a sequence of maps $\{q_k\}_{k \geqslant k_0}$ from Ω to Y by

$$q_{k_0}(\omega) = y(\omega),$$
$$q_{k_0+1}(\omega) = \theta_{2^{k_0+1}}(\pi_{2^{k_0+1}}(S(\omega) \cap \pi_{2^{k_0+1}}(\overline{B_{2^{-k_0}}(q_{k_0}(\omega))}))),$$

and

$$q_k(\omega) = \theta_{2^k}(\pi_{2^k}(S(\omega) \cap \pi_{2^k}(\overline{B_{2^{-(k-1)}}(q_{k-1}(\omega))})))$$

for any $k \geqslant k_0 + 1$. It can be checked that

$$d(q_{k+1}(\omega), q_k(\omega)) \leqslant 3 \cdot 2^{-(k+1)}, \qquad d(q_k(\omega), S(\omega)) \leqslant 2^{-k} \tag{5.2}$$

for any $k \geqslant k_0$.

Take $q(\omega) = \lim_k q_k(\omega)$. By (5.2), $q(\omega) \in S(\omega)$ and

$$d(q(\omega), y(\omega)) \leqslant \sum_{k \geqslant k_0} d(q_{k+1}(\omega), q_k(\omega)) \leqslant 3 \cdot 2^{-k_0} < 6\epsilon.$$

By Lemma 5.2, q_k is $\mathscr{A}$–$\mathscr{B}(Y)$ measurable for any $k \geqslant 1$, which implies that q is $\mathscr{A}$–$\mathscr{B}(Y)$ measurable. □

By using Lemma 5.2 we give a short proof of the following known result (see, Molchanov,[7] p. 102):

Proposition 5.3. Let Y be an arbitrary compact metric space. Then

(i) there exists a map q: $\mathscr{K} \to Y$ such that q is $\mathscr{B}(\mathscr{K})$–$\mathscr{B}(Y)$ measurable and

$$q(K) \in K, \qquad \forall K \in \mathscr{K};$$

(ii) core(T) is a nonempty compact convex subset of $\mathscr{M}(Y)$ for each capacity functional T on $\mathscr{A}(Y)$.

Proof. Define a sequence of maps $q_k: \mathscr{K} \to Y$ by

$$q_k(K) = \theta_{2^k}(\pi_{2^k}(K) \cap \pi_{2^k}(\overline{B_{2^{-(k-1)}}(q_{k-1}(K))})), \qquad k \geqslant 2.$$

Letting $q = \lim_{k \to \infty} q_k$, we obtain (i). The proof of part (ii) is divided into the following two steps:

Step 1. We prove that core(T) is compact and convex. The convexity is trivial. To see the compactness, suppose $\mu_n \in \text{core}(T)$ and μ_n converges to μ. It suffices to show $\mu \in \text{core}(T)$. Note that for each $K \in \mathscr{K}$ and $\epsilon > 0$

$$\mu(K) \leqslant \mu(B_\epsilon(K)) \leqslant \limsup_n \mu_n(B_\epsilon(K)) \leqslant \limsup_n \mu_n(\overline{B_\epsilon(K)}) \leqslant T(\overline{B_\epsilon(K)}).$$

Letting $\epsilon \downarrow 0$, by Corollary 4.3 we have $\mu(K) \leqslant T(K)$. Thus $\mu \in \text{core}(T)$.

Step 2. We prove that core(T) is nonempty. Let q be defined as in part (i). Then $q(S(\omega))$ is a random vector taking values in Y. Denote by μ the distribution of $q(S(\omega))$ on Y, that is

$$\mu(E) = P(\omega: q(S(\omega)) \in E), \qquad \forall E \in \mathscr{B}(Y).$$

By the definition of T one can check directly that $\mu \in \text{core}(T)$. □

6. PERTURBATION PROPERTIES OF core(T)

Let us begin with a result of Norberg.[9]

Proposition 6.1 (Theorem 4.6 of Ref. 9). Let T be the capacity functional of a random set $S: \Omega \to \mathscr{K}$. Then $\mu \in \text{core}(T)$ if and only if there exists a probability space $(\Omega_1, \mathrm{S}_1, P_1)$, a random set $S_1: \Omega_1 \to \mathscr{K}$ and a random vector $x_1: \Omega_1 \to Y$ such that S_1 has the same distribution of S, x_1 has the distribution μ and moreover $x_1 \in S_1$ almost surely.

Using this result we can give a complete characterization of core(T) whenever Y is a finite set.

Proposition 6.2. Suppose Y is a finite set and T is the capacity functional of a random set $S: \Omega \to 2^Y \setminus \{\varnothing\}$. Denote by

$$\phi(E) := \sum_{F \subseteq E} (-1)^{\#(E \setminus F)} (1 - T(Y \setminus F)), \qquad \forall E \in 2^Y \setminus \{\varnothing\}. \tag{6.1}$$

Then $\mu \in \text{core}(T)$ if only if there is a map p: $2^Y \setminus \{\emptyset\} \times Y \to \mathbb{R}^+$ satisfying:

(i) $p(E, x) = 0$ if $x \notin E$;

(ii) $\sum_{x \in E} p(E, x) = 1$ for any $E \in 2^Y \setminus \{\emptyset\}$. such that

$$\mu(\{x\}) = \sum_{E \ni x} p(E, x)\, \phi(E), \qquad \forall x \in Y.$$

Proof. It is not hard to check that $\phi(E) = P(\{\omega\colon S(\omega) = E\})$ for any $E \in 2^Y \setminus \{\emptyset\}$. We divide the proof into the following two steps.

Step 1. Sufficiency. Construct a probability space $(2^Y \setminus \{\emptyset\} \times Y, 2^{2^Y \setminus \{\emptyset\} \times Y}, \nu)$ by

$$\nu((E, x)) = \phi(E)\, p(E, x).$$

Define Z: $2^Y \setminus \{\emptyset\} \times Y \to 2^Y \setminus \{\emptyset\}$ and z: $2^Y \setminus \{\emptyset\} \times Y \to Y$ by

$$Z(E, x) = E \quad \text{and} \quad z(E, x) = x, \qquad \forall (E, x) \in 2^Y \setminus \{\emptyset\} \times Y.$$

One can check directly that the random set Z has the same distribution as S, and z has the distribution μ. Moreover $z \in Z$ almost surely. Thus $\mu \in \text{core}(T)$.

Step 2. Necessity. Suppose $\mu \in \text{core}(T)$. By Proposition 6.1, there exists a probability space $(\Omega_1, \mathbf{S}_1, P_1)$, a random set $S_1\colon \Omega_1 \to \mathcal{K}$, and a random vector $x_1\colon \Omega_1 \to Y$ such that S_1 has the same distribution as S, x_1 has the distribution μ and moreover $x_1 \in S_1$ almost surely. Define a map p: $2^Y \setminus \{\emptyset\} \times Y \to \mathbb{R}^+$ in the following way: set $p(E, x) = 0$ if $x \notin E$; otherwise if $x \in E$ and $\phi(E) > 0$, define

$$p(E, x) = \frac{1}{\phi(E)} P_1(\{\omega_1\colon S_1(\omega_1) = E, x_1(\omega_1) = x\})$$

and if $\phi(E) = 0$, we define $p(E, x) = \frac{1}{\#E}$ for $x \in E$. It is clear that p satisfies (i), (ii) and moreover $\mu(\{x\}) = \sum_{E \ni x} p(E, x)\, \phi(E)$ for any $x \in Y$. □

Now we need to study the perturbation of core(T), i.e., the way in which the core(T) depends on T, in the finite case.

Lemma 6.3. Suppose Y is a finite set. Let T_1 and T_2 be two capacity functionals on 2^Y. Assume

$$|T_1(E) - T_2(E)| < \delta, \qquad \forall E \subset Y. \tag{6.2}$$

Then

$$\rho_\Delta(\text{core}(T_1), \text{core}(T_2)) < \#Y \cdot 4^{\#Y} \cdot \delta.$$

Proof. Define

$$\phi_i(E) = \sum_{F \subseteq E} (-1)^{\#(E \setminus F)} (1 - T_i(Y \setminus F)), \qquad \forall E \in 2^Y \setminus \{\varnothing\}, \quad i = 1, 2.$$

By (6.2),

$$|\phi_1(E) - \phi_2(E)| < \#(2^E) \cdot \delta \leqslant \#(2^Y) \cdot \delta = 2^{\#Y} \cdot \delta.$$

Let $\mu_1 \in \text{core}(T_1)$. By Proposition 6.2, there is a map p: $2^Y \setminus \{\varnothing\} \times Y \to \mathbb{R}^+$ satisfying

(i) $p(E, x) = 0$ if $x \notin E$;

(ii) $\sum_{x \in E} p(E, x) = 1$ for any $E \in 2^Y \setminus \{\varnothing\}$, such that

$$\mu_1(\{x\}) = \sum_{E \ni x} p(E, x)\, \phi_1(E), \qquad \forall x \in Y.$$

Now define a probability measure μ_2 on Y by

$$\mu_2(\{x\}) = \sum_{E \ni x} p(E, x)\, \phi_2(E), \qquad \forall x \in Y.$$

Using Proposition 6.2 again, we know that $\mu_2 \in \text{core}(T_2)$. Furthermore for any $x \in Y$,

$$|\mu_1(\{x\}) - \mu_2(\{x\})| \leqslant \sum_{E \ni x} |\phi_1(E) - \phi_2(E)| \leqslant 2^{\#Y} \cdot 2^{\#Y} \cdot \delta = 4^{\#Y} \delta.$$

By (3.1),

$$\Delta(\mu_1, \mu_2) \leqslant \sum_{x \in Y} |\mu_1(\{x\}) - \mu_2(\{x\})| \leqslant \#Y \cdot 4^{\#Y} \cdot \delta.$$

This implies that

$$\text{core}(T_1) \subset B_\epsilon(\text{core}(T_2)),$$

where $\epsilon = \#Y \cdot 4^{\#Y} \cdot \delta$. In a similar way, we can prove

$$\text{core}(T_2) \subset B_\epsilon(\text{core}(T_1)).$$

Therefore we have $\rho_\Delta(\text{core}(T_1), \text{core}(T_2)) < \#Y \cdot 4^{\#Y} \cdot \delta$. □

To study the perturbation of core(T) in the case where Y is an arbitrary compact space, we use a technique to approximate core(T) in the following way.

Suppose Y is an arbitrary compact space. T is the capacity functional of a random set $S: \Omega \to \mathscr{K}$. Let H_n, π_n, and θ_n be defined as in the first part of Section 5. Define a capacity $\Theta_n(T)$ on 2^{H_n} as follows:

$$\Theta_n(T)(E) = P\{\omega \in \Omega : \pi_n(S(\omega)) \cap E \neq \varnothing\}, \qquad \forall E \in H_n. \tag{6.3}$$

It is clear that

$$\Theta_n(T)(E) = T(B_{\frac{1}{n}}(E)), \qquad \forall E \in H_n. \tag{6.4}$$

As we know, $\Theta_n(T)$ is the capacity functional of the random set $\pi_n S$, taking values in the finite collection of all subsets of H_n, and core($\Theta_n(T)$) is a set of probability measures on H_n. Since every probability measure on H_n can be viewed as a Borel probability measure on Y, core($\Theta_n(T)$) can be treated as a compact subset of $\mathscr{M}(Y)$. In the following we consider the distance between core($\Theta_n(T)$) and core(T) in the Hausdorff metric ρ_Δ.

Lemma 6.4. Suppose Y is an arbitrary compact space. Let $\{f_i\}$ be defined as in (3.1). T and $\Theta_n(T)$ are given as above. Then

$$\rho_\Delta(\text{core}(\Theta_n(T)), \text{core}(T)) \leqslant \sum_{i=1}^{\infty} \frac{C_{f_i}(\frac{6}{n})}{2^i \|f_i\|}, \tag{6.5}$$

where $C_f(\epsilon) = \sup\{|f(x) - f(y)| : d(x, y) \leqslant \epsilon\}$.

Proof. The proof will be divided into two steps.

Step 1. core(T) $\subset B_\delta$(core($\Theta_n(T)$)) *with* $\delta = \sum_{i=1}^{\infty} \frac{C_{f_i}(\frac{6}{n})}{2^i \|f_i\|}$.

To show this, pick any $\mu \in$ core(T). By Proposition 6.1 there exists a probability space $(\Omega_1, \mathrm{S}_1, P_1)$, a random set $S_1: \Omega_1 \to \mathscr{K}$ and a random vector $x_1: \Omega_1 \to Y$ such that S_1 has the same distribution as S, x_1 has the distribution μ and moreover $x_1 \in S_1$ almost surely. Now define $S_2: \Omega_1 \to 2^{H_n}$ and $x_2: \Omega_1 \to H_n$ by

$$S_2(\omega_1) = \pi_n(S_1(\omega_1)) \qquad \text{and} \qquad x_2(\omega_1) = \theta_n(\pi_n(x_1(\omega_1))).$$

Let μ_2 denote the distribution of x_2. Since $\Theta_n(T)$ is the capacity functional of S_2 and $x_2 \in S_2$ almost surely, $\mu_2 \in$ core($\Theta_n(T)$). By (3.1), we have

$$\Delta(\mu_2, \mu) = \sum_{i=1}^{\infty} \frac{|\int f_i(x_2(\omega_1))\, dP_1(\omega_1) - \int f_i(x_1(\omega_1))\, dP_1(\omega_1)|}{2^i \, \|f_i\|}$$

$$\leqslant \sum_{i=1}^{\infty} \frac{\int |f_i(x_2(\omega_1)) - f_i(x_1(\omega_1))|\, dP_1(\omega_1)}{2^i \, \|f_i\|}$$

$$\leqslant \sum_{i=1}^{\infty} \frac{C_{f_i}(\frac{1}{n})}{2^i \, \|f_i\|},$$

from which we get the desired result.

Step 2. $\mathrm{core}(\Theta_n(T)) \subset B_\delta(\mathrm{core}(T))$ *with* $\delta = \sum_{i=1}^{\infty} \frac{C_{f_i}(\frac{1}{n})}{2^i \|f_i\|}$.

To show this, assume $\mu \in \mathrm{core}(\Theta_n(T)))$. Define ϕ by

$$\phi(E) = P\{\omega \in \Omega : \pi_n(S(\omega)) = E\}, \qquad \forall E \subset H_n.$$

By Proposition 6.2, there exists a map p: $2^{H_n} \setminus \{\varnothing\} \times H_n \to \mathbb{R}^+$ satisfying

(i) $p(E, x) = 0$ if $x \notin E$;

(ii) $\sum_{x \in E} p(E, x) = 1$ for any $E \in 2^{H_n} \setminus \{\varnothing\}$, such that

$$\mu(\{x\}) = \sum_{E \ni x} p(E, x)\, \phi(E), \qquad \forall x \in H_n.$$

For any $\varnothing \neq E \subset H_n$, define

$$\Omega_E = \{\omega \in \Omega : \pi_n(S(\omega)) = E\}.$$

By Lemma 5.2, $\Omega_E \in \mathscr{F}$. Construct

$$\Omega_1 = \bigcup_{\varnothing \neq E \subset H_n} \Omega_E \times \{(E, x) \colon x \in E\}.$$

Define a σ-algebra $\mathbf{S}_1$ such that each element of $\mathbf{S}_1$ is the finite union of elements of following form:

$$A_E \times \{(E, x)\}, \qquad E \subset H_n,\; x \in E,\; A_E \subset \Omega_E,\; A_E \in \mathscr{F}.$$

By Kolmogorov's consistency theorem, there is a unique probability measure on the measurable space $(\Omega_1, \mathbf{S}_1)$ such that

$$P_1(A_E \times \{(E, x)\}) = P(A_E)\, p(E, x), \qquad E \subset H_n,\; x \in E,\; A_E \subset \Omega_E,\; A_E \in \mathscr{F}.$$

Define $S_1 \colon \Omega_1 \to \mathscr{K}$, $S_2 \colon \Omega_1 \to 2^{H_n}$, and $x_2 \colon \Omega_1 \to H_n$ respectively by

$$S_1(\omega, E, x) = S(\omega), \qquad S_2(\omega, E, x) = E, \qquad x_2(\omega, E, x) = x.$$

One can check that S_1 has the same distribution as S, S_2 has the same distribution as $\pi_n S$ and that S_2 induces the capacity $\Theta_n(T)$. Furthermore, x_2 has the distribution μ. Moreover for each $(\omega, E, x) \in \Omega_1$,

$$x \in E = \pi_n(S(\omega)),$$

hence $d(x_2(\omega, E, x), S_1(\omega, E, x)) \leqslant \frac{1}{n}$. By Proposition 5.1, there exists a S_1–$\mathscr{B}(Y)$ measurable map $y \colon \Omega_1 \to Y$ such that

$$y(\omega, E, x) \in S_1(\omega, E, x), \qquad d(y(\omega, E, x), x_2(\omega, E, x)) \leqslant \frac{6}{n}$$

for any $(\omega, E, x) \in \Omega_1$. Let μ_1 denote the distribution of y, then $\mu_1 \in \operatorname{core}(T)$. Furthermore

$$\begin{aligned}
\Delta(\mu_1, \mu) &= \sum_{i=1}^{\infty} \frac{\left|\int f_i(x_2(\omega, E, x))\, dP_1 - \int f_i(y(\omega, E, x))\, dP_1\right|}{2^i\, \|f_i\|} \\
&\leqslant \sum_{i=1}^{\infty} \frac{\int |f_i(x_2(\omega, E, x)) - f_i(y(\omega, E, x))|\, dP_1}{2^i\, \|f_i\|} \\
&\leqslant \sum_{i=1}^{\infty} \frac{C_{f_i}(\frac{6}{n})}{2^i\, \|f_i\|},
\end{aligned}$$

from which we get the desired result.

Proofs of Theorem 3.3. Fix an integer n. Let $\Theta_n(T_1)$ and $\Theta_n(T_2)$ be defined as in (6.4). By Lemma 6.4, we have

$$\rho_\Delta(\operatorname{core}(\Theta_n(T_j)), \operatorname{core}(T_j)) \leqslant \sum_{i=1}^{\infty} \frac{C_{f_i}(\frac{6}{n})}{2^i\, \|f_i\|}, \qquad j = 1, 2. \tag{6.6}$$

In Lemma 6.3, replacing Y, T_1, T_2 respectively by H_n, $\Theta_n(T_1)$, and $\Theta_n(T_2)$ we obtain that

$$\begin{aligned}
\rho_\Delta(\operatorname{core}(\Theta_n(T_1)), \operatorname{core}(\Theta_n(T_2))) &\leqslant \#H_n \cdot 4^{\#H_n} \cdot \max_{E \subset H_n} |\Theta_n(T_1)(E) - \Theta_n(T_2)(E)| \\
&= \#H_n \cdot 4^{\#H_n} \cdot \max_{W \in \mathcal{O}_n} |T_1(W) - T_2(W)|.
\end{aligned}$$

Combining this with (6.6), we have

$$\begin{aligned}\rho_\Delta(\text{core}(T_1), \text{core}(T_2)) &\leqslant \sum_{j=1}^{2} \rho_\Delta(\text{core}(\Theta_n(T_j)), \text{core}(T_j)) \\ &\quad + \rho_\Delta(\text{core}(\Theta_n(T_1)), \text{core}(\Theta_n(T_2))) \\ &\leqslant \#H_n \cdot 4^{\#H_n} \cdot \max_{W \in \mathcal{O}_n} |T_1(W) - T_2(W)| + 2 \sum_{i=1}^{\infty} \frac{C_{f_i}(\frac{6}{n})}{2^i \, \|f_i\|}.\end{aligned}$$

This finishes the proof. □

7. PROOFS OF THEOREMS 3.6 AND 3.8

We first prove the following lemma using the classical Cramer Principle in large deviation theory (cf. Dupuis and Ellis[(1)]).

Lemma 7.1. Fix an integer $m \geqslant 1$. For any $\delta > 0$, there exists N_δ and $R_\delta > 0$ such that

$$P\left\{\omega \in \Omega : \sum_{W \in \mathcal{O}_m} |T_\omega^{(n)}(W) - T(W)|^2 > \delta^2\right\} \leqslant e^{-nR_\delta}$$

for all $n \geqslant N_\delta$.

Proof. For any $i \geqslant 1$, let $\chi_i(\omega, W)$ be defined as in (3.3). Denote by $Y_i(\omega)$ the $\#\mathcal{O}_m$ dimensional vector $(x_i(\omega, W))_{W \in \mathcal{O}_m}$ indexed by $W \in \mathcal{O}_m$. It is clear that $\{Y_i\}$ is a sequence of i.i.d random vector taking values in $\mathbb{R}^{\#\mathcal{O}_m}$, with a distribution supported on finitely many points. Note that

$$\frac{1}{n} \sum_{i=1}^{n} Y_i(\omega) = (T_\omega^{(n)}(W))_{W \in \mathcal{O}_m}, \qquad E(Y_1) = (T(W))_{W \in \mathcal{O}_m}.$$

Applying the classical Cramer Principle to the i.i.d random vector $\{Y_i\}$ (see, e.g., Theorem 3.5.1 of Dupuis and Ellis,[(1)] p. 87), we get the desired result. □

Proof of Theorem 3.6. Fix $\epsilon > 0$. Choose a large integer m such that

$$\sum_{i=1}^{\infty} \frac{C_{f_i}(\frac{6}{m})}{2^i \, \|f_i\|} < \frac{\epsilon}{4}.$$

Set $\delta = \frac{\epsilon}{2\#H_m 4^{\#H_m}}$. By Lemma 7.1, there exists n_δ and $R_\delta > 0$ such that

$$P\left\{\omega \in \Omega : \sum_{W \in \mathcal{O}_m} |T_\omega^{(n)}(W) - T(W)|^2 > \delta^2\right\} \leqslant e^{-nR_\delta}$$

for all $n \geqslant n_\delta$.

By Theorem 3.3,

$$\rho_\Delta(\text{core}(T_\omega^{(n)}, \text{core}(T)) > \epsilon \Rightarrow \max_{W \in \mathcal{O}_m} |T_\omega^{(n)}(W) - T(W)| \geqslant \delta.$$

Therefore

$$P\{\omega \in \Omega : \rho_\Delta(\text{core}(T_\omega^{(n)}, \text{core}(T)) > \epsilon\} \leqslant e^{-nR_\delta}$$

for all $n \geqslant n_\delta$. Define $n_\epsilon = N_\delta$ and $L_\epsilon = R_\delta$. Then the above fact implies the desired result. □

Proof of Theorem 3.8. Let $\mu \in \text{core}(T)$. By Proposition 6.1, there exist a probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, a random set $\bar{S}: \bar{\Omega} \to \mathcal{K}$ and a random vector $\bar{X}: \bar{\Omega} \to Y$ such that $\bar{S}$ has the same distribution of S, $\bar{X}$ has distribution μ and $\bar{P}(\bar{X} \in \bar{S}) = 1$. Let $\{\bar{S}_i\}$ be a sequence of independent copies of $\bar{S}$, $\{\bar{X}_i\}$ a sequence of independent copies of $\bar{X}$ with $\bar{P}(\bar{X}_i \in \bar{S}_i) = 1$. Let $\bar{T}_{\bar{\omega}}^{(n)}$ be the empirical capacity functional based on $\{\bar{S}_i(\bar{\omega})\}_{i=1}^n$, and $\bar{\mu}_{\bar{\omega}}^{(n)}$ be the empirical measure based on $\{\bar{X}_i(\bar{\omega})\}_{i=1}^n$ given by

$$\bar{\mu}_{\bar{\omega}}^{(n)}(K) = \frac{1}{n} \#\{1 \leqslant i \leqslant n, \bar{X}_i(\bar{\omega}) \in K\}.$$

By $\bar{P}(\bar{X}_i \in \bar{S}_i) = 1$ and Proposition 6.1, we have $\bar{\mu}_{\bar{\omega}}^{(n)}(K) \leqslant \bar{T}_{\bar{\omega}}^{(n)}(K)$ a.s. for any given $K \in \mathcal{K}$. Note that $\bar{T}^{(n)}$ has the same distribution as $T^{(n)}$, and thus the set of probability measures $\text{core}(\bar{T}^{(n)})$ is the same set as $\text{core}(T^{(n)})$. Therefore we only need to prove the approximation result for $\sqrt{n}(\bar{\mu}_{\bar{\omega}}^{(n)} - \mu)$.

Now for any $f \in C(Y)$, using the central limit theorem to the i.i.d. random variables $f(\bar{X}_i(\bar{\omega}))$ we obtain

$$\lim_{n \to \infty} \bar{P}\left(-z \leqslant \frac{\sum_{i=1}^n f(\bar{X}_i(\bar{\omega})) - n \int f(\bar{X})\, d\bar{P}}{\sqrt{n}\sqrt{\int f^2(\bar{X})\, d\bar{P} - (\int f(\bar{X})\, d\bar{P})^2}} < z\right)$$

$$= \frac{1}{2\pi} \int_{-z}^{z} e^{-x^2/2}\, dx, \qquad \forall z > 0.$$

Denote $G_{\bar{\omega}}^{(n)} := \sqrt{n}(\bar{\mu}_{\bar{\omega}}^{(n)} - \mu)$. Then

$$\sum_{i=1}^{n} f(\bar{X}_i(\bar{\omega})) - n \int f(\bar{X})\, d\bar{P} = \sqrt{n}\, G_{\bar{\omega}}^{(n)}(f).$$

Since $\sqrt{\int f^2(\bar{X})\, d\bar{P} - (\int f(\bar{X})\, d\bar{P})^2} \leqslant \|f\|$, we have

$$\lim_{n\to\infty} \bar{P}(|G_{\bar{\omega}}^{(n)}(f)| \leqslant z\,\|f\|) \geqslant \frac{1}{\sqrt{2\pi}} \int_{-z}^{z} e^{-x^2/2}\, dx, \qquad \forall z > 0. \tag{7.1}$$

For any $\epsilon > 0$, choose an integer k such that $2^{-k} < \epsilon/2$, and pick a large N such

$$\frac{1}{\sqrt{2\pi}} \int_{-N}^{N} e^{-x^2/2}\, dx > 1 - \frac{\epsilon}{2k}.$$

Let $\{f_i\}$ be given as in (3.1). Then by (7.1),

$$\lim_{n\to\infty} \bar{P}(|G_{\bar{\omega}}^{(n)}(f_i)| \leqslant N\,\|f_i\|) \geqslant 1 - \frac{\epsilon}{2k}, \qquad i = 1,\ldots,k.$$

Note that

$$\{\bar{\omega}\colon \rho_\Delta(G_{\bar{\omega}}^{(n)}, 0) > N\} \subset \bigcup_{i=1}^{k} \{\bar{\omega}\colon |G_{\bar{\omega}}^{(n)}(f_i)| \geqslant N\,\|f_i\|\}.$$

We have

$$\lim_{n\to\infty} \bar{P}(\rho_\Delta(G_{\bar{\omega}}^{(n)}, 0) > N) \leqslant \epsilon.$$

Thus

$$\lim_{n\to\infty} \bar{P}(\rho_\Delta(G_{\bar{\omega}}^{(n)}, 0) \leqslant N) \geqslant 1 - \epsilon.$$

This finishes the proof. □

ACKNOWLEDGMENTS

We thank Professors Hung T. Nguyen and Tonghui Wang for suggesting the problem to us and for their helpful discussions, and three referees for many valuable suggestions. De-Jun Feng was partially supported by the Special Funds for Major State Basic Research Projects in China.

REFERENCES

1. Dupuis, P., and Ellis, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley.
2. Falconer, K. J. (1986). *The Geometry of Fractal Sets*, Cambridge University Press.
3. Gill, R. D., van der Laan, M. J., and Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures, counter-examples. In Lin, D. Y., and Fleming, T. R. (eds.), *Proceedings of the First Seattle Symposium in Biostatistics*, Lecture Notes in Statistics, Vol. 123, Springer-Verlag, New York, pp. 255–294.
4. Hartigan, J. A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82**(357), 267–270.
5. Heitjan, D. F., and Rubin, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **23**, 774–786.
6. Matheron, G. (1975). *Random Sets and Integral Geometry*, Wiley.
7. Molchanov, I. S. (1993). *Limit Theorems for Unions of Random Closed Sets*, Lecture Notes in Mathematics, Vol. 1561, Springer-Verlag.
8. Molchanov, I. S. (1997). *Statistics of the Boolean Model for Practitioners and Mathematicians*, Wiley.
9. Norberg, T. (1992). On the existence of ordered couplings of random sets—with applications. *Israel J. Math.* **77**, 241–264.
10. Polonik, W. (1995). Density estimation under qualitative assumptions in higher dimensions. *J. Multi. Anal.* **55**, 61–81.
11. Schreiber, T. (2000). Statistical inference from set-valued observations. *Prob. Math. Statist.* **20**, 223–235.
12. Shapley, L. S. (1971). Cores of convex games. *Internat. J. Game Theory* **1**, 11–26.
13. van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence of Empirical Processes*, Springer, New York.
14. van der Vaart, A. W., and Wellner, J. A. (2000). Preservation theorems for Glivanko–Cantelli and uniform Glivanko–Cantelli classes. In Gine, E., Mason, D., and Wellner, J. A. (eds.), *High Dimensional Probability II*, Birkhäuser, Boston, pp. 113–132.
15. Walters, P. (1982). *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, Vol. 79, Springer-Verlag.