Pergamon

Neural Networks, Vol 7, No. 1, pp 101–111, 1994

0893-6080/94 $6.00 + .00

CONTRIBUTED ARTICLE

Topologically Ordered Competitive Sampling

MICHAEL LEMMON

University of Notre Dame

(*Received* 28 *August* 1991; *revised and accepted* 7 *June* 1993)

Abstract—*An important problem in nonparametric modeling is the selection of samples from an independent and identically distributed observation (IID) process such that the resulting sample ensemble forms a prototypical model of the observation process. This paper discusses a sampling technique that is closely related to topologically ordered competitive learning In the context of a tightly constrained notion of topological ordering, the proposed sampling algorithm uses a randomized winner-take-all rule to modify sample ensembles in a way that generates a reversible Markov chain The stationary density for this reversible chain is explicitly computed and used to characterize the sampling procedure's steady-state statistical behaviour This characterization shows that topologically ordered competitive sampling can generate steady-state sample ensembles that form equi-probable partitions of the observation space More importantly, this paper provides a formal analytical framework that can be used as a model for the analysis of other competitive learning paradigms*

Keywords—Competitve learning, Nonparametric density estimation, Topological ordering.

1. INTRODUCTION

In various applications, a system's status is monitored by randomized sampling of a sequence of observations on the process. Such samples are often called prototypes and the entire monitoring process can be referred to as prototypical sampling. The objective of the sampling process is to obtain an ensemble (collection) of samples that in some sense characterize the system's current state. More precisely, we can consider the sequence of observations as an independent and identically distributed (IID) stochastic process, $\bar{y}(n)$, with underlying observation density, $p(\bar{y})$. The objective is to collect and selectively update a finite collection of N prior observations so that the resulting ensemble extremalizes an assumed performance or error criterion measuring how prototypical the ensemble is of the observation process. This paper presents a specific sampling protocol in which the sample ensemble forms an equiprobable partition of the observation space. This partition will be prototypical in the sense of maximizing the average mutual information between the ensemble samples and the observation process. In analyzing the proposed sampling protocol's performance, this paper provides a framework for the analysis of topologically ordered competitive learning algorithms.

A common approach to prototype selection is to take the last N observations as prototypes (ensemble samples). This approach, which may be referred to as independent sampling, has been used by recent neural network models to solve classification and parameter estimation problems (Specht, 1990). The order statistics of the N sampled observation vectors will have beta distributions so that the network's steady state behaviour can be readily characterized. Unfortunately, the ensemble's order statistics will exhibit significant statistical fluctuations unless the number of prototypes is relatively large. This weakness of independent sampling has led some researchers to use competitive learning algorithms in forming prototypical ensembles (Burrascano, 1991).

Competitive learning (CL) algorithms (Rumelhart & Zipser, 1986; Kohonen, 1982) are neural network training paradigms. In these algorithms, the observation replaces or updates those ensemble samples (i.e., neuron weight vectors) that are closest to the current observation. The neurons can be thought of as competing for the right to be updated by the current observation. The criterion for winning this competition is the proximity of the neuron's weight vector to the observation with respect to an assumed vector norm. If the ith neuron has a weight vector $\bar{w}_i$ and the input is $\bar{y}$, then the update rule for the winning neuron will be given by

Acknowledgement: I would like to acknowledge the partial financial support of the National Science Foundation (IRI91-09298).

Requests for reprints should be sent to the author at Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556

$$\bar{w}_i(n+1) = \bar{w}_i(n) + \eta[\bar{y}(n) - \bar{w}_i(n)] \qquad (1)$$

where the increment parameter η is positive and less than or equal to unity. For simple competitive learning (Rumelhart & Zipser, 1986), η is usually a constant. For networks such as Kohonen's feature map, η is a function of the input and all of the network's weight vectors. Note that if $\eta = 1$, then the winning neuron's weight vector is replaced by the observation. In this case, the network is sampling the observation process and for this reason the process can be called competitive sampling.

Unfortunately, competitive learning algorithms have proven notoriously difficult to analyze. For simple competitive learning (Rumelhart & Zipser, 1986), it can be shown that the procedure implements (on the average) a gradient descent on a functional that penalizes poor clusterings of inputs and weight vectors. For simple competitive learning, this functional is usually nonconvex and hence it is difficult to determine whether or not the procedure converges in some appropriate sense. Convergence results for Kohonen's feature map, however, are more mature (Ritter & Schulten, 1986, 1988). But these analyses are somewhat limited in their ability to provide quantitative characterizations of the Kohonen network in any but the simplest cases. In spite of these limitations, the existing analyses of the Kohonen network are some of the most rigorous and complete analytical results existing for competitive learning. This paper conjectures that the reason for this completeness is due in large measure to a key difference between Kohonen's network and simple competitive learning. This difference is the Kohonen network's introduction of a topological ordering over the neurons.

Neurons are often thought of as being arranged on a two-dimensional sheet called the cortical space or cortex. The relationship between neurons in this cortical space establishes a fixed ordering of the neurons. Therefore, topological ordering refers to the physical relationship of one neuron to another in this cortical space. Simple competitive learning does not use this notion of ordering. For the Kohonen network, however, a topological ordering of the neurons is generated by assigning a vector $\bar{r}_i$ to the ith neuron. This vector, $\bar{r}_i$, represents the ith neuron's position in cortical space. This location is assumed to be fixed. The competitive adjustment equation [eqn (1)] is then modified to take this ordering into account by making the increment η a function of $\bar{r}_i$. For the Kohonen network, the increment takes the form $\eta(\|\bar{r}_i - \bar{r}_{i^*}\|)$, where $\eta(x)$ decays monotonically to zero as $\|x\|$ goes to infinity and i^* is the neuron whose weight vector is closest to the applied input (i.e., $\|\bar{w}_{i^*} - \bar{y}\| \leq \|\bar{w}_i - \bar{y}\|$ for all i).

The introduction of a fixed topological ordering may be one reason why prior analyses were successful in characterizing the behaviour of simple Kohonen networks. By introducing a topological ordering, the number of configurations that the network can assume will be severely limited. If the ordering can be appropriately defined, then it is quite possible that the resulting learning procedure will generate a Markov chain with a single ergodic class, thereby allowing a quantitative evaluation of the learning procedure's steady-state behaviour.

Motivated by the preceding arguments, this paper proposes a competitive sampling algorithm that is closely related to the Kohonen algorithm in that a precisely defined notion of topological ordering is used. While Kohonen's algorithm arose from heuristic conjectures about the functional nature of biological neural assemblies, the algorithm proposed and analyzed in this paper arose from a conscious effort to develop a computational procedure that generated an easily characterized Markov chain. Consequently, the analyses of this paper will be significant in that they allow quantitative evaluation of competitive sampling to an extent that is not possible with other algorithms.

The specific notion of topological ordering used in this paper will be defined in Section 2. The competitive sampling procedure is proposed in Section 3. This procedure is analyzed in Section 4 under the assumption of a uniformly distributed observation process. The analysis is extended to nonuniform observation processes in Section 5. These sections will show that the proposed competitive sampling procedure will generate a reversible Markov chain whose stationary density implies a steady-state ensemble forming an equi-probable partition of the observation space. The significance of these results and analyses will be summarized in Section 6.

2. TOPOLOGICAL ORDERING

This section precisely defines the notion of topological ordering used in this paper. In particular, a topological ordering for a sample ensemble is defined to be a mapping from the sample ensemble onto a collection of sets that form a topology for the observation space. A specific topological ordering is considered in which the topology is generated by a basis set consisting of disjoint simplices. A recursive algorithm is then presented for generating this basis set of simplices.

Consider a discrete-time IID m-dimensional process, $\bar{y}(n)$ with underlying probability density function $p(\bar{y})$. This process is called the observation process and the vector $\bar{y}(n) \in \Re^m$ is the observation at time n. Assume that the observation density, $p(\bar{y})$, has its support [set over which $p(\bar{y})$ is nonzero] on a set Ω that is a convex polyhedron in $\Re^m$.

Since the support set Ω is a convex polyhedron, there exists a finite collection of vectors in $\Re^m$ whose convex hull equals Ω. This set of vectors will be called the set of *boundary prototypes* Consider a finite collection of observations from within Ω. These vectors are called

interior prototypes. The entire collection of prototypes (boundary and interior) will be assumed to be in general position (no subset of $m + 1$ points lie in the same hyperplane). The resulting collection of N vectors, $\mathcal{X} = \{\bar{x}_1, \ldots, \bar{x}_N\}$, will be called an Ω-*ensemble* or just an ensemble. An individual vector $\bar{x}_i \in \Re^m$ will be called a *prototype.* As noted above, each prototype will either be a boundary or interior prototype.

Assume that there exists a mapping from the Ω-ensemble $\mathcal{X}$ onto a collection of M open subsets $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_M\}$ where $\tau_i \subset \Re^m$ for $i = 1, \ldots, M$. This family of sets will be called a *collection.* If $\mathcal{T}$ includes the null set and Ω, is closed under set intersection, and includes $\cup_i \tau_i$, then this collection will be called the *topology* for set Ω generated by Ω-ensemble $\mathcal{X}$. Note that the topology is a partially ordered set with respect to set inclusion. Therefore, if $\mathcal{T}$ is a topology for Ω, it can be said that the ensemble $\mathcal{X}$ is *topologically ordered.* Further note that because Ω is contained in the topology that the ordered pair $(\Omega, \mathcal{T})$ is a topological space. A topological ordering of $\mathcal{X}$ is, therefore, a specific mapping from $\mathcal{X}$ onto a collection of sets that form a topology, $\mathcal{T}$, for Ω. This mapping will be denoted as the topological ordering, $\mathcal{T}(\mathcal{X})$. The range of $\mathcal{T}(\mathcal{X})$ is a topology for Ω.

The topological ordering, $\mathcal{T}(\mathcal{X})$, can be generated in a variety of ways. In this paper, a specific class of mappings is studied in which simplices form a basis for $\mathcal{T}(\mathcal{X})$. In particular, consider the ensemble $\mathcal{X} = \{\bar{x}_1, \ldots, \bar{x}_N\}$. Associate K subsets of $\mathcal{X}$, $\mathcal{F}_j(i)$ for $j = 1, \ldots, K$, with the ith prototype, $\bar{x}_i$. Each subset, $\mathcal{F}_j(i)$, consists of exactly m prototypes. The convex hull of these vectors $\mathrm{Co}[\mathcal{F}_j(i)]$ will be called the jth *facet* associated with the ith prototype. The facet is a simplex in $\Re^{m-1}$ [also called an $(m-1)$-simplex]. Note that the union of $\mathcal{F}_j(i)$ with the prototype vector $\mathbf{x}_i$ yields another subset whose convex hull will be an m-simplex in $\Re^m$. The jth simplex associated with the ith observation can therefore be written as $S_j(i) = \mathrm{Co}[\mathcal{F}_j(i) \cup \{\bar{x}_i\}]$. We then define a collection of sets, $\mathcal{T}(\mathcal{X})$, that consists of the null set and all sets generated by unions and intersections of the simplices, $S_j(i)$. This collection will be called the *simplicial collection* generated by the collection of simplices $\mathcal{S} = \{S_j(i)\}$.

It will be convenient in the following discussion to define a variety of other sets generated by the simplices, $S_j(i)$. The union of all K simplices associated with the ith prototype will be called the ith *polygon cell,* P_i:

$$P_i = \bigcup_{j=1}^{K} S_j(i). \tag{2}$$

Note that the ith cell complex, P_i, will generally not be convex. For various reasons to be seen below, it will be convenient to consider a convex set within P_i. Note that the prototypes in each facet $\mathcal{F}_j(i)$ lie in a unique hyperplane. This hyperplane separates $\Re^m$ into two halfspaces. The prototype $\bar{x}_i$ belongs to only one of these halfspaces (assuming all points are in general position). Denote the halfspace containing $\bar{x}_i$ as $H_j(i)$. The ith prototype's *neighborhood cell,* C_i, will be defined to be the intersection of all these halfspaces:

$$C_i = \bigcap_{j=1}^{J} H_j(i). \tag{3}$$

These sets define m-dimensional cells in $\Re^m$ that can be interpreted as representing the neighborhood structure of each prototype. The polygon cell represents the neighborhood directly in terms of the simplices. The neighborhood cell, C_i, is simply a convex body within P_i. The reason for introducing these sets is that they will be used by the competitive sampling algorithm to ensure that the updated ensemble's topological ordering is not violated. Figure 1 illustrates the relationships between C_i, P_i, and $S_j(i)$ for a specific topologically ordered Ω-ensemble.

The following theorem provides a sufficient condition for a simplicial collection to be a topology for Ω.

THEOREM 1. *Let $\mathcal{X}$ be an Ω-ensemble and let $\mathcal{T}(\mathcal{X})$ be the simplicial collection generated by the simplices $\{S_j(i)\}$. If $\Omega \in \mathcal{T}(\mathcal{X})$ and all simplices $S_j(i)$ form a disjoint partition with possible repetitions of Ω, then $\mathcal{T}(\mathcal{X})$ is a topological ordering for Ω*

Proof. By the definition of simplicial collection, $\mathcal{T}(\mathcal{X})$ contains the null set. By assumption the interiors of all simplices are disjoint; therefore, all elements in $\mathcal{T}(\mathcal{X})$ can be represented as the interior of the union of various simplices $S_j(i)$. The intersection of any two sets in $\mathcal{T}(\mathcal{X})$ will clearly be another union of simplices and so the collection is closed under set intersection. By assumption, $\Omega \in \mathcal{T}(\mathcal{X})$; therefore, the simplicial collection must be a topological ordering for $\mathcal{X}$. ■

The topology $\mathcal{T}(\mathcal{X})$ generated by simplices $S_j(i)$ will also be called a *simplicial topology* To construct a specific simplicial topology for Ω requires that either the facets, $\mathcal{F}_j(i)$, or the simplices, $S_j(i)$, be known. The following theorem defines a special class of simplicial topologies that can be recursively generated using a

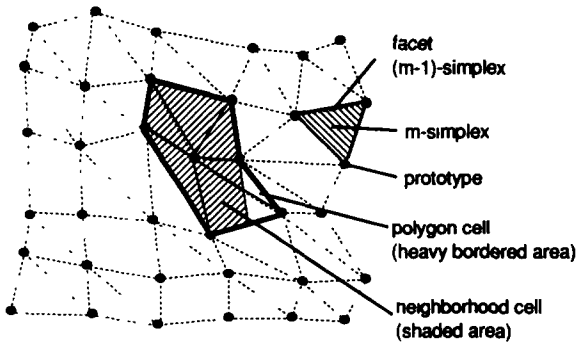

FIGURE 1. Topological ordering: facets, simplex, polygon cell, neighborhood cell.

multiresolution procedure. These topological orderings will therefore be called *multiresolution simplicial topologies*

THEOREM 2. *Assume Ω is an m-simplex and let $\mathcal{X}$ be an Ω-ensemble Let a simplicial collection $\mathcal{T}$ be formed by the following procedure.*

1. *Initialize $\mathcal{X}_B$ to be the set of all boundary prototypes and $\mathcal{X}_I$ be the set of all interior prototypes Initialize the collection of simplices $\mathcal{S}$ to contain Ω.*
2. *Randomly select a single prototype, $\bar{x}$, from $\mathcal{X}_I$ and determine which m-simplex in $\mathcal{S}$ contains $\bar{x}$. Let that simplex be S_i Remove $\bar{x}$ from $\mathcal{X}_I$ and put it in collection $\mathcal{X}_B$*
3. *Let $S_{i,j}^{m-1}$ denote the jth $(m-1)$-simplex formed from $m-1$ vertices of the selected m-simplex, S_i Let $S_{i,j}^{m}$ denote the m-simplex formed by the $m-1$-simplex $S_{i,j}^{m-1}$ and the prototype $\bar{x}$ Replace S_i in collection $\mathcal{S}$ with $S_{i,j}^{m}$ for $j = 1, \ldots, m+1$*
4. *Return to step 2 if $\mathcal{X}_I$ is not empty*
5. *Let $\mathcal{T}$ be the collection formed by the null set and arbitrary unions of elements in $\mathcal{S}$*

Then $\mathcal{T}$ is a simplicial topology for Ω

Proof By construction, $\mathcal{T}$, is clearly a simplicial collection. The construction procedure guarantees that the interiors of all simplices are disjoint since at each iteration, a single element, S_i, of $\mathcal{S}$ is replaced by a disjoint collection of smaller simplices whose union equals S_i. Also Ω must be in $\mathcal{T}$ since at each iteration the union of all elements in $\mathcal{S}$ must equal Ω. Therefore by the preceding theorem, it can be concluded that $\mathcal{T}$ is a simplicial topology for Ω. ■

REMARK 1. At the jth iteration of the algorithm, the set of simplices, denoted as $\mathcal{S}_j$, can be used to construct a simplicial topology for Ω that is denoted as $\mathcal{T}_j$. Furthermore, the simplicial topology at $\mathcal{T}_j$ is properly included in $\mathcal{T}_{j+1}$. Therefore, the sequence of simplicial topologies form a multiresolution approximation of Ω. It is for this reason that topologies generated by Theorem 2 will be called multiresolution simplicial topologies.

REMARK 2. Note that because any convex polyhedron can be partitioned into a set of simplices with a single point in the interior, then this result also applies to any support set Ω that is a convex polyhedron.

REMARK 3. The topology formed by a loop through the algorithm in Theorem 2 can be associated with the index of the interior prototype that was added to $\mathcal{X}_B$ in step 2 of that algorithm. These indices form a sequence that can be denoted as $I(\mathcal{T})$. This sequence of indices represents the order in which interior prototypes were used to generate the simplicial topology $\mathcal{T}$.

REMARK 4. A simplicial topology will be specified in one of two ways. The first way lists the simplices in $\mathcal{T}$. The second way specifies the collection $\mathcal{X}$ and the index set $I(\mathcal{T})$.

3. COMPETITIVE SAMPLING ALGORITHM

The proposed competitive sampling (CS) algorithm will now be formally stated using the constructions of the preceding section. In this algorithm each observation will be tested to determine which prototype's neighborhood cell contains the observation. The observation can lie in several neighborhood cells. In the following procedure, the parameter, M_{lub}, represents a least upper bound on the number of neighborhood cells that an observation can belong to. In the following algorithm, the index n refers to time.

1. Let $n = 1$. Initialize a collection of N prototypes, $\mathcal{X}$, with boundary prototypes and interior prototypes so that the collection is an Ω-ensemble.
2. Define a mapping from the Ω-ensemble $\mathcal{X}$ onto a set of simplices, $\mathcal{S}$, that form the basis for a simplicial topology of Ω.
3. Get the observation, $\bar{y}(n)$.
4. For all interior prototypes, $\bar{x}_i$, tag those prototypes whose neighborhood cells, C_i, contain $\bar{y}$ [i.e., $\bar{y}(n) \in C_i$]. Assume that there are M such prototypes.
5. Decide with probability M/M_{lub} to replace a prototype.
6. If the decision is not to replace a prototype, then increment n and go to step 3.
7. Else with probability $1/M$ select one and only one of the M tagged prototypes and replace it with the observation $\bar{y}(n)$. Increment n and go to step 3.

REMARK 1. The procedure in Theorem 2 can be used to generate the collection of basis simplices, $\mathcal{S}$.

REMARK 2. The CS algorithm's use of the neighborhood cell, C_i, is reminiscent of a recent modification of the Kohonen algorithm (Cherkassky & Lori–Najafi, 1991) in which the topological ordering is also simplicial in nature. An important aspect of Cherkassky's work is that the ordering is not violated by the incremental update algorithm. As will be seen in the following section, the CS algorithm also has this property.

REMARK 3. A key difference between the CS algorithm and other topologically ordered competitive learning paradigms is the use of a randomized replacement rule. This rule essentially identifies a number of prototypes for possible replacement and then makes a probabilistic replacement decision. The motivation behind introducing a randomized replacement rule is that similar rules yield reversible Markov chains. This fact will be used extensively in the following sections.

4. UNIFORM OBSERVATION PROCESSES

In this section the CS algorithm is analyzed under the assumption of uniformly distributed observations (Ω

$= [0, 1]^m$). The analysis starts by showing that the sequence of Ω-ensembles formed by repeated iterations of the CS algorithm will always generate simplicial topologies of Ω. The Ω-ensemble at time n, $\mathcal{X}(n)$, will be shown to be a reversible Markov process whose stationary density $\pi(\mathcal{X})$ implies that all topologically ordered Ω-ensembles are equally likely. For general simplicial topologies, these results are proven in Subsection 4.1.

Unfortunately, explicit evaluation of this stationary density will not generally be possible. For multiresolution simplicial topologies, however, combinatorial arguments can be used to evaluate marginal densities for a vector stochastic process, $V(n)$, whose components are the volumes of the simplices in $\mathcal{S}$. These results for multiresolution simplicial topologies are presented in Subsection 4.2. They show that the CS algorithm, when applied to Ω-ensembles with a multiresolution simplicial topology, will generate sample ensembles forming an equi-probable partition of the observation space, Ω.

Finally, it is important to obtain experimental validation of these analytical predictions. Simulation results supporting this section's analyses are presented in Subsection 4.3.

4.1. General Simplicial Topologies

Repeated application of the CS algorithm will generate a sequence of ensembles. The ensemble at time n will be sometimes called a *configuration* of the ensemble and will be denoted as $\mathcal{X}(n)$. The stochastic process, $\mathcal{X}(n)$, is clearly a Markov chain. The configurations will therefore be adapted to the observation process through a single-step transition probability density, $p[\mathcal{X}(n+1)|\mathcal{X}(n)]$. Because the CS algorithm replaces only one prototype at a time, the configurations at time n and $n+1$ can be written as

$$\mathcal{X}(n) = \{\bar{x}_1, \ldots, \bar{x}_{k-1}, \bar{x}_k, \bar{x}_{k+1}, \ldots, \bar{x}_N\} \quad (4)$$

$$\mathcal{X}(n+1) = \{\bar{x}_1, \ldots, \bar{x}_{k-1}, \bar{y}(n), \bar{x}_{k+1}, \ldots, \bar{x}_N\} \quad (5)$$

where the kth prototype was chosen by the CS algorithm for replacement by the current observation $\bar{y}(n)$.

The transition probability density can be obtained as follows. Let the configuration at time n be $\mathcal{X}(n)$. The probability that the kth prototype will be replaced equals the probability of an observation lying in the kth neighborhood cell C_k divided by M_{lub}. Because the observations are uniformly distributed, this implies that the conditional density will be constant over the neighborhood cell C_k. Therefore the transition density will be

$$p[\mathcal{X}(n+1)|\mathcal{X}(n)] = \frac{I_{C_k}[\bar{y}(n)]}{M_{\text{lub}}\text{vol }\Omega} \quad (6)$$

where I_{C_k} is an indicator function over the kth neighborhood cell, vol Ω is the volume of Ω, and $\bar{y}(n)$ is the kth element of configuration $\mathcal{X}(n+1)$ (i.e., the observation vector at time n). In this paper, it will be assumed that $\Omega = [0, 1]^m$ has unit volume so that vol Ω can be dropped from the preceding equation.

The CS algorithm possesses certain properties that allow it to preserve the topological ordering of the ensemble. These properties are summarized in the following lemma.

LEMMA 1. *Let $\mathcal{X}(n)$ be an Ω-ensemble and let $\mathcal{T}(\mathcal{X}(n))$ be a simplicial topology for Ω generated by $\mathcal{X}(n)$. If the CS algorithm replaces the ith prototype by $\bar{y}(n)$, then*

1. *The ith prototype's neighborhood cell at time n, $C_i(n)$, equals the ith prototype's neighborhood cell, $C_i(n+1)$, at time $n+1$.*
2. *$\mathcal{T}[\mathcal{X}(n+1)]$ is a simplicial topology for Ω generated by $\mathcal{X}(n+1)$,*

Proof Note that the set $C_i(n)$ is a convex set determined solely by the prototypes of the ith prototype's neighborhood cell, $\mathcal{F}_J(i)$. By definition, the neighborhood cells are independent of i. Therefore, the neighborhood cell at time $n+1$ will be unchanged and so $C_i(n+1)$ must equal $C_i(n)$. The first conclusion of the Lemma is therefore proven.

Since this is a simplicial topology at time n, all simplices are mutually disjoint. Assume that a single application of the CS algorithm results in at least two different simplices with nontrivial intersections. Consider the intersecting simplices in pairs. One of these simplices (denote as S_1) must have a vertex that is the updated prototype, $\bar{x}_i$. The other simplex (denote as S_2) may or may not have an updated vertex. If it does have an updated vertex, then both simplices clearly belong to the same polygon cell. If it does not have an updated vertex, then S_2 must intersect the ith polygon cell at time n, $P_i(n)$. The preceding paragraph's discussion implies that $P_i(n-1) = P_i(n)$. Therefore, S_2 must intersect $P_i(n-1)$. By assumption, however, these simplices were disjoint so that S_2 must be inside $P_i(n-1)$. It can therefore be concluded that both intersecting simplices belong to the same polygon cell.

Now assume that the simplices, $S_l(i)$ and $S_k(\iota)$, intersect. Because the simplices are convex sets, any point in the simplex $S_l(i)$ is a convex combination of $\bar{x}_i$ and some point $\bar{x}_{B1}$ on a face of the simplex {i.e., $\bar{x}_{B1} \in \text{Co}[\mathcal{F}_J(i)]$ for some $J = 1, \ldots, K$}. Because of the nonempty intersection we can therefore find two points, $\bar{x}_{B1}$ and $\bar{x}_{B2}$, such that

$$\mu\bar{x}_i(n) + (1-\mu)\bar{x}_{B1} = \lambda\bar{x}_i(n) + (1-\lambda)\bar{x}_{B2} \quad (7)$$

where $0 < \mu < 1$ and $0 < \lambda < 1$. Rearranging terms yields

$$\bar{x}_i(n) = \frac{1-\lambda}{\mu-\lambda}\bar{x}_{B2} - \frac{1-\mu}{\mu-\lambda}\bar{x}_{B1} \quad (8)$$

$$= \gamma_1\bar{x}_{B2} + (1-\gamma_1)\bar{x}_{B1}. \quad (9)$$

From the constraints on μ and λ it is clear that the two simplices intersect iff $\gamma_1 < 0$ or $\gamma_1 > 1$. These values for γ_1 coupled with eqn (9) imply that $\bar{x}_i(n+1)$ does not lie in one of the halfspaces $H_j(i)$ associated with the neighborhood cell. This, of course, generates a contradiction because the CS algorithm requires that all $\bar{x}_i$ lie in the neighborhood cell. As a result of this contradiction, all simplices in the updated ensemble must be disjoint and the first conclusion of the Lemma is proven. ■

An immediate consequence of Lemma 1 is that the forward and backward single-step transition probabilities must be equal. This result is stated and proved in the following lemma.

LEMMA 2. *If $\mathcal{X}(n)$ is the Markov chain of ensembles that is generated by the CS algorithm with respect to a uniformly distributed observation process, then*

$$p[\mathcal{X}(n+1)|\mathcal{X}(n)] = p[\mathcal{X}(n)|\mathcal{X}(n+1)]. \quad (10)$$

Proof The configurations $\mathcal{X}(n+1)$ and $\mathcal{X}(n)$ are identical except for a single prototype, say $\bar{x}_i$, in the ensemble's configuration. This vector, $\bar{x}_i$, will be replaced only if the observation vector $\bar{y}(n)$ lies in the ith neighborhood cell, $C_i(n)$ (note the cell's dependence on time n). The single-step forward transition density can therefore be written as $p[\mathcal{X}(n+1)|\mathcal{X}(n)] = I_{C_i(n)}[\bar{y}(n)]/M_{\text{lub}}$. The probability of the reverse transition can therefore be written as $p[\mathcal{X}(n+1)|\mathcal{X}(n)] = I_{C_i(n+1)}[\bar{y}(n)]/M_{\text{lub}}$. By Lemma 1, $C_i(n) = C_i(n+1)$, so it can be concluded that the forward and backward transition densities are equal, by eqn (6). ■

As will be shown a little further on, the process $\mathcal{X}(n)$ will be a reversible Markov chain. Reversible Markov chains possess a number of important properties that are now summarized. A stochastic process $\mathcal{X}(n)$ is said to be *reversible* (Kelly, 1979) if $[\mathcal{X}(n_1), \mathcal{X}(n_2), \ldots, \mathcal{X}(n_m)]$ has the same density has $[\mathcal{X}(k-n_1), \mathcal{X}(k-n_2), \ldots, \mathcal{X}(k-n_m)]$ for all $n_1, n_2, \ldots, n_m$ and k. Basically this means that if the direction of time is reversed, then the process's behaviour is unchanged. An important property of reversible processes is that they are stationary (Kelly, 1979). This can be easily seen by the fact that for k equals 0, both the forward and time-reversed processes $\mathcal{X}(n)$ and $\mathcal{X}(-n)$ have the same density. For Markov chains, Kolmogorov's criteria (Kelly, 1979) provides a convenient method of determining whether or not the chain is reversible.

THEOREM 3. *Kolomogorov's criteria. A Markov chain is reversible iff its transition densities satisfy*

$$p(\mathcal{X}_2|\mathcal{X}_1)p(\mathcal{X}_3|\mathcal{X}_2) \cdots p(\mathcal{X}_{m-1}|\mathcal{X}_m)p(\mathcal{X}_m|\mathcal{X}_1) = p(\mathcal{X}_1|\mathcal{X}_m)p(\mathcal{X}_m|\mathcal{X}_{m-1}) \cdots p(\mathcal{X}_3|\mathcal{X}_2)p(\mathcal{X}_2|\mathcal{X}_1) \quad (11)$$

for any finite sequence of states $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m$

Once the reversibility of the chain has been established then the prior discussion would imply that the chain is stationary. In fact, the stationary density must satisfy the so-called *detailed balance* conditions.

THEOREM 4. (*Kelly*, 1979) *The stationary density $\pi(X)$ of a reversible Markov chain satisfies the detailed balance conditions*

$$\pi(\mathcal{X}_1)p(\mathcal{X}_2|\mathcal{X}_1) = \pi(\mathcal{X}_2)p(\mathcal{X}_1|\mathcal{X}_2) \quad (12)$$

where $\mathcal{X}_1$ and $\mathcal{X}_2$ are any two states in the chain

The following Lemma proves that the Ω-ensembles form a reversible Markov chain.

LEMMA 3. *If $\mathcal{X}(n)$ is the Markov chain of configurations generated by the CS algorithm with respect to a uniformly distributed observation process, then the process is reversible*

Proof Let $\mathcal{Z}_1$ be any Ω-ensemble generated by the CS algorithm from an initial Ω-ensemble, $\mathcal{Z}$. This means that there is a sequence of observations that (through the CS algorithm) generates a sequence of ensembles that begins in $\mathcal{Z}$ and terminates in $\mathcal{Z}_1$. Denote this sequence as $\mathcal{Z}, \mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m, \mathcal{Z}_1$. From Lemma 2 it is known that $p(\mathcal{X}_{i+1}|\mathcal{X}_i) = p(\mathcal{X}_i|\mathcal{X}_{i+1})$ so that the forward and backward strings of probability densities are equal. This conclusion holds for any ensemble $\mathcal{Z}_i$ that can be reached from the initial configuration $\mathcal{Z}$. Note that if $\mathcal{Z}_i$ and $\mathcal{Z}_j$ are reachable from $\mathcal{Z}$, then they are reachable from each other and it can be concluded that $p(\mathcal{Z}_i|\mathcal{Z}_j) = p(\mathcal{Z}_j|\mathcal{Z}_i)$. We can now consider any sequence of configurations, $\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_m$, that are reachable from $\mathcal{Z}$, then the equality of the forward and backward transition densities implies that

$$p(\mathcal{Z}_2|\mathcal{Z}_1)p(\mathcal{Z}_3|\mathcal{Z}_2) \cdots p(\mathcal{Z}_m|\mathcal{Z}_{m-1})p(\mathcal{Z}_1|\mathcal{Z}_m) = p(\mathcal{Z}_1|\mathcal{Z}_2)p(\mathcal{Z}_2|\mathcal{Z}_3) \cdots p(\mathcal{Z}_{m-1}|\mathcal{Z}_m)p(\mathcal{Z}_m|\mathcal{Z}_1). \quad (13)$$

This equation satisfies Kolomogorov's criteria and therefore the Markov chain $\mathcal{X}(n)$ is reversible. ■

The reversibility of the chain $\mathcal{X}(n)$ implies that the process is stationary. Furthermore, because $\mathcal{X}(n)$ is Markov, this implies that detailed balance is satisfied. These facts are summarized in the following theorem.

THEOREM 5. *If $\mathcal{X}(n)$ is the Markov chain generated by the CS algorithm over the set of topologically ordered ensembles with respect to a uniformly distributed observation process, then the process's stationary density $\pi(\mathcal{X})$ satisfies the equality $\pi(\mathcal{X}) = \pi(\mathcal{Z})$ for all ensembles $\mathcal{X}$ and $\mathcal{Z}$ reachable from $\mathcal{X}(0)$ by the CS algorithm.*

Proof By Lemma 3 the chain $\mathcal{X}(n)$ is reversible and by Theorem 4 it satisfies the detailed balance conditions, $\pi(\mathcal{Z})p(\mathcal{X}|\mathcal{Z}) = \pi(\mathcal{X})p(\mathcal{Z}|\mathcal{X})$, where $\mathcal{Z}$ and $\mathcal{X}$ are any ensembles reachable by the CS algorithm from an initial ensemble. From the proof of Lemma 3, it is clear that these transition densities are equal

[i.e., $p(\mathcal{X}\,|\,\mathcal{Z}) = p(\mathcal{Z}\,|\,\mathcal{X})$]. Therefore, the stationary density must satisfy $\pi(\mathcal{Z}) = \pi(\mathcal{X})$ for any $\mathcal{X}$ and $\mathcal{Z}$ that are reachable from the initial ensemble $\mathcal{X}(0)$. ■

4.2. Multiresolution Simplicial Topologies

Theorem 5 is significant for it asserts that all topologically ordered ensembles that are reachable from $\mathcal{X}(0)$ are equally likely. This result is valid for any ensemble with a simplicial topology. The theorem, however, falls short of actually computing the stationary density [i.e., the constant value of $\pi(\mathcal{X})$]. Performing this computation involves computing the number of ensembles that can be reached from $\mathcal{X}(0)$. For arbitrary simplicial topologies, this computation is nontrivial.

For multiresolution simplicial topologies, however, the computation can be done by examining an associated stochastic process $V(n)$ that is derived from the original Markov chain, $\mathcal{X}(n)$. This associated process consists of the volumes of all simplices $S_j(i)$ in $\mathcal{S}$ generated at the end of the recursive procedure in Theorem 2. In this section, it is shown that if $\mathcal{T}(\mathcal{X})$ is a multiresolution simplicial topology, then the process $V(n)$ is a reversible Markov process whose stationary density implies a steady-state ensemble in which all simplices have equal volumes (on the average). The mean value of this steady-state volume can be explicitly computed through combinatorial arguments. The results show that the CS algorithm for Ω-ensembles with multiresolution simplicial topologies will form equi-probable partitions of the observation space.

Consider a collection of m-simplices, $\mathcal{S} = \{S_i\}$, forming the basis of a multiresolution simplicial topology. Assume there are N of these simplices. Because the ensemble is topologically ordered, these simplices partition Ω. Let v_i be the volume of the ith simplex

$$v_i = \text{vol } S_i = \int_{\bar{y}\in S_i} d\bar{y}. \tag{14}$$

The simplex S_i is determined by $m+1$ prototypes, $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_m$, in $\Re^m$. A well-known formula for the volume of this simplex is

$$v_i = \text{vol } S_i = \frac{1}{m!}\left|\det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \bar{x}_0 & \bar{x}_1 & \cdots & \bar{x}_m \end{pmatrix}\right|. \tag{15}$$

The collection of all these volumes at time n, $V(n) = \{v_1(n), \ldots, v_N(n)\}$, forms the ensemble's volume or V-configuration at time n. The process $V(n)$ will be called the V process.

The V-configuration at time n and $n+1$ will be denoted as $V(n)$ and $V(n+1)$, respectively. $V(n+1)$ will be generated from $V(n)$ by a single application of the CS algorithm. Because the CS algorithm only updates a single prototype $\bar{x}_i$ at a time, then only those L simplices associated with the kth polygon cell, P_k, will have their volumes changed. Let these modified simplices be denoted as $S_{k_1}, S_{k_2}, \ldots, S_{k_L}$. If the V-configuration at time n is $V(n) = \{v_1, \ldots, v_N\}$, then the V-configuration at time $n+1$ after the ith prototype has been updated will be

$$V(n+1) = \{v_1, \ldots, v_{k_1-1}, \tilde{v}_{k_1}, v_{k_1+1}, \ldots, v_{k_L-1}, \tilde{v}_{k_L}, v_{k_L+1}, \ldots, v_N\} \tag{16}$$

where $\tilde{v}_{k_j}$ is the updated volume of simplex S_{k_j}, the jth simplex of polygon cell P_k.

For arbitrary simplicial topologies, the process $V(n)$ is not necessarily Markovian. Consider two different configurations, $\mathcal{X}$ and $\mathcal{Z}$. These two configurations need not have the same neighborhood cells and because the volume of the neighborhood cell determine the transition densities, it is no longer apparent that knowledge of $V(n)$ is sufficient to determine the probability of the transition to $V(n+1)$. Consequently, the process $V(n)$ will not generally be a Markov process.

Note, however, that the reason why $V(n)$ is not Markov is that the specification of the simplex volumes does not uniquely determine the volume of the neighborhood cell, C_i. It does, however, specify the volume of the polygon cell, P_i. If P_i is always convex, then it equals C_i and the simplex volumes at time n would uniquely determine the probability of transitioning to a different V-configuration. It could therefore be concluded that for ensembles in which most polygon cells are convex, then the process $V(n)$ would be *approximately* Markov.

Although $V(n)$ is not Markov for arbitrary simplicial topologies, it can be shown to be Markov for multiresolution simplicial topologies. The following lemmas and theorem form the basis for this conclusion.

LEMMA 4. *Let S be an m-simplex in $\Re^m$ and let S_j denote the jth m-simplex formed by a facet of S and a vecgtor $\bar{x}_c$ within S If the volumes, v_j, of the simplices S_j and the vertices of S are known, then $\bar{x}_c$ is uniquely determined.*

Proof. Equation (15) yields an expression for the volume of simplex S_j

$$v_j = \text{vol } S_j = \frac{1}{m!}\left|\det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \bar{x}_c & \bar{x}_1 & \cdots & \bar{x}_m \end{pmatrix}\right| \tag{17}$$

where $\bar{x}_i$ for $i = 1$ to m are vertices of S forming a facet [i.e., $(m-1)$-simplex] of S. Because the vertices of S are known, then the determinant can be rewritten in terms of the components of the unknown vector $\hat{x} = \{1\,|\,\bar{x}_c^t\}^t$. Let X denote the matrix of vertices in eqn (17), then v_j can be written as

$$v_j = \frac{1}{m!}\left|\sum_k (-1)^{k+l}\hat{x}_l |X_{kl}|\right| \tag{18}$$

where $|X_{kl}|$ is the determinant of the klth cofactor of X and $\hat{x}_l$ is the lth component of vector $\hat{x}$. By appropriate permutation of the rows of X, the absolute value

in eqn (18) can be removed. Therefore, eqn (18) ($j = 1, \ldots, m+1$) forms an $m+1$ system of simultaneous linear equations. Note that there can be no vector $\hat{x}$ such that all simplices would have zero volume v_j. The system therefore has full rank and the vector $\hat{x}$ can be uniquely determined once the simplex volumes and vertices are known. ■

LEMMA 5. *Let $\mathcal{X}$ be an Ω-ensemble and let $\mathcal{T}(\mathcal{X})$ be a multiresolution simplicial topology for Ω generated by $\mathcal{X}$. Let V be a vector whose components are the volumes of the simplices forming the basis for $\mathcal{T}(\mathcal{X})$. If the boundary prototypes of Ω are known and if the index set, $I(\mathcal{T})$, for the simplicial topology is known, then each V specifies a unique configuration $\mathcal{X}$*

Proof. By assumption $\mathcal{T}(\mathcal{X})$ is generated using the procedure in Theorem 2. Because $\mathcal{T}$ is multiresolution, there exists an increasing sequence of topologies ($\{\mathcal{T}_i\}$ such that $\mathcal{T}_i \subset \mathcal{T}_{i+1}$). Because the boundary prototypes are known and the volumes of the simplices are known, the volumes of all simplices in each $\mathcal{T}_i$ can be determined. Now consider $\mathcal{T}_1$. The simplicies forming this topology are the $m+1$ m-simplices obtained by taking a facet of $\mathcal{T}_0$ with an interior prototype. Because the volumes of these simplices are known and because the vertices of $\mathcal{T}_0$ are boundary prototypes (also known), the preceding lemma implies that the interior prototype is uniquely determined. Similarly, if the interior prototype is known, then the volumes of the simplices forming $\mathcal{T}_1$ are uniquely determined. The procedure is now extended through induction to conclude that there is a one-to-one correspondence between interior prototypes of $\mathcal{X}$ and the volumes in V. ■

The following lemma uses the fact that there is a one-to-one correspondence between volume vectors and prototype ensembles to prove that the volume process is Markov and reversible.

THEOREM 6. *Let $\mathcal{X}(n)$ be the Markov chain of ensembles generated by the CS algorithm with respect to a uniformly distributed observation process and let $\mathcal{T}[\mathcal{X}(n)]$ be a multiresolution simplicial topology for Ω generated by $\mathcal{X}(n)$. Then the associated V-process, $V(n)$, is a reversible Markov chain*

Proof If, at time n, the volume vector is $V(n)$, then Lemmas 4 and 5 show that the information needed to determine $\mathcal{X}(n)$ from the volumes consists of the boundary prototypes of $\mathcal{X}$ and the index set of the simplicial topology, $I(\mathcal{T})$. The boundary prototypes are unchanged by the CS algorithm. By Lemma 1 the CS algorithm preserves topological ordering, so that the index set $I(\mathcal{T})$ is independent of n. Therefore, for any two volume vectors, V_1 and V_2, there exists a uniquely associated ensemble $\mathcal{X}_1$ and $\mathcal{X}_2$ that depends only on the current volumes. Both of these ensembles, however, are specific instances of the process $\mathcal{X}(n)$ that is already known to be Markov and reversible. Therefore, $V(n)$ must also be a reversible Markov process. ■

The following theorem shows that because the volume process is Markov and reversible with a uniformly distributed observation process, then the process must have a stationary density that implies all steady-state admissible volume vectors are equally likely.

THEOREM 7. *Let $\mathcal{X}(n)$ be the Markov chain of ensembles generated by the CS algorithm with respect to a uniformly distributed observation process and let $\mathcal{T}[\mathcal{X}(n)]$ be a multiresolution simplicial topology for Ω generated by $\mathcal{X}(n)$. Then the associated V-process, $V(n)$, has a stationary density in which $\pi(V) = \pi(W)$ for any two V-configurations.*

Proof. Because $V(n)$ is reversible and Markov, it has a stationary density $\pi(V)$ that satisfies detailed balance. This stationary distribution will be centered on the set of volume configurations that are reachable from an initial configuration. However, because there is a one-to-one correspondence between the V-process and the $\mathcal{X}$-process, and because the $\mathcal{X}$-process is ergodic, it is known that all possible V-configurations are reachable from any initial V-configuration. Because the observation process is uniformly distributed, the forward and backward transition probabilities are equal and therefore the stationary density $\pi(V)$ is constant over the whole set of possible V-configurations. ■

The following theorem computes the marginal densities for the volume process's stationary density. The significance of this result is that it provides a quantitative evaluation of the sample ensemble's statistical behaviour under the assumption of uniformly distributed observations.

THEOREM 8. *Let $\mathcal{X}(n)$ be the Markov chain of ensembles generated by the CS algorithm with respect to a uniformly distributed observation process and let $\mathcal{T}[\mathcal{X}(n)]$ be a multiresolution simplicial topology for Ω generated by $\mathcal{X}(n)$ The associated V-process's ith marginal density, $\pi_m^{(i)}$ will be*

$$\pi_m^{(i)}(v_i) = \int \cdots \int \int \cdots \int \pi(v_1, \ldots, v_N) \times dv_1 \ldots dv_{i-1} dv_{i+1} \ldots dv_N \tag{19}$$

$$= (N-1)(1-v_i)^{N-2}. \tag{20}$$

Proof. In this proof, the continuum Ω is being replaced by a discrete approximation. Assume that we have partitioned $[0, 1]^M$ into M volume elements where M is large ($M \gg N$). Each component of the V-configuration is a volume holding some number of volume elements such that the total number of volume elements summed over the entire V-configuration is clearly M. The number of admissible V-configurations will therefore be the number of ways M nondistinct objects can be divided

amongst N bins. This is a simple combinatorial problem (Liu, 1968) whose solution is

$$C(N-1, M) = \frac{(N+M-1)!}{(N-1)!M!}. \tag{21}$$

Now assume that a given simplex, S_i, has K volume elements. The number of V-configurations where S_i has K volume elements equals the number of ways $M - K$ nondistinct objects are divided amongst $N - 1$ distinct bins. This combinatorial problem has the solution

$$C(N-2, M-K) = \frac{(N+M-K-2)!}{(N-2)!(M-K)!}. \tag{22}$$

Equations (21) and (22) are combined to give the probability that a given simplex, S_i, has volume K/M

$$Pr(K/M) = \frac{C(N-2, M-K)}{C(N-1, M)}. \tag{23}$$

It is now necessary to take the limit as M gets large. Applying Stirling's formula to eqn (23) yields

$$Pr(K/M) = (N-1)\left(\frac{\{[N+M-1)-(K+1)]!M!\}}{(N+M-1)!(M-K)!}\right) \tag{24}$$

$$\approx (N-1)\left(\frac{\sqrt{1-\frac{K_1}{M_1}}\left(1-\frac{K_1}{M_1}\right)^{M_1-K_1}\left(\frac{e}{M_1}\right)^{K_1}}{\sqrt{1-\frac{K}{M}}\left(1-\frac{K}{M}\right)^{M-K}\left(\frac{e}{M}\right)^{K}}\right) \tag{25}$$

where $K_1 = K + 1$ and $M_1 = N + M - 1$. For large $M \gg N$, it is assumed that $K_1/M_1 \approx K/M$. Therefore, eqn (25) is simplified to

$$Pr(K/M) \approx (N-1)\left(1-\frac{K}{M}\right)^{N-2}. \tag{26}$$

Replacing K/M by x, the probability density of obtaining an V-configuration whose ith component has volume x is

$$\pi_m^{(i)}(x) = (N-1)(1-x)^{N-2}. \tag{27}$$

This completes the proof. ■

COROLLARY 1. *Under the assumptions of Theorem* 8, *the first and second moments of the marginal density* $\pi_m^{(i)}$ *are* $1/N$ *and* $2/(N^2+N)$, *respectively.*

Proof. This result follows from a direct calculation of the moments. ■

4.3. Simulation Results

This subsection presents simulation results validating the quantitative conclusions of the preceding subsection. A two-dimensional observation space was assumed with a uniform observation density. The CS algorithm was simulated for an ensemble with a multiresolution simplicial topology generated by a random collection of observation vectors. The number of simplices (size of ensemble) was varied between 50 and 200. The simplex volumes and fluctuations (standard deviation) were estimated by computing sample averages over 25,000 iterations of the CS algorithm. The results are shown in Figures 2 and 3. Figure 2 shows the predicted simplex volume (Corollary 1) plotted against ensemble size (solid line). The circles in the figure denote the average volume over all simplices in the ensemble. The error bars indicate the standard deviation in the estimated mean volumes over the entire ensemble. As can be clearly seen, the figure shows close agreement between the predicted and observed simplex volumes. Figure 3 shows the predicted fluctuations (standard deviations) in simplex volumes (Corollary 1) plotted against ensemble size. The circles indicate the average fluctuations and the error bars denote the standard deviation in the estimated fluctuations over the entire ensemble. Once again, the simulation results appear to corroborate the preceding predictions.

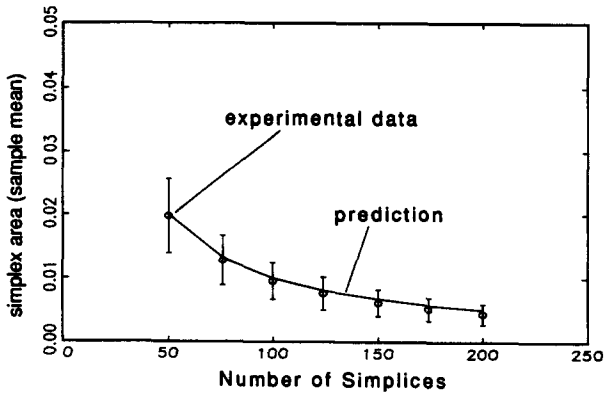

FIGURE 2. Simplex volume (mean) versus ensemble size.

5. NONUNIFORM OBSERVATION PROCESSES

This section indicates how the results of Section 4 can be extended to general observation processes on [0,

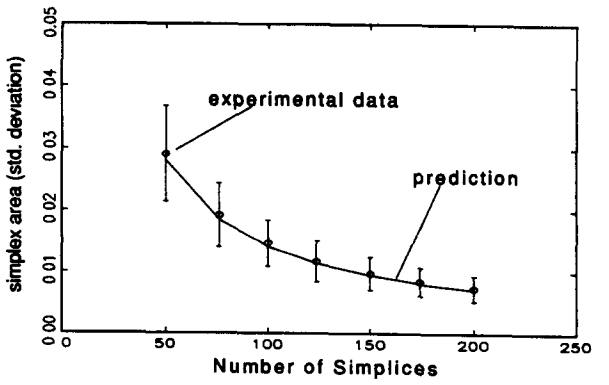

FIGURE 3. Simplex volume (fluctuation) versus ensemble size.

$1]^m$ This extension indicates that the CS algorithm will generate a steady-state sample ensemble, $\mathcal{X}$, whose simplices form an equiprobable partition of Ω.

Consider an m-dimensional IID observation process, $\bar{y}(n)$, with continuously differentiable probability density, $p(\bar{y})$, whose support set is $[0, 1]^m$. Assume that there exists a set of prototypes, $\mathcal{X}$, that are ordered by a multiresolution simplicial topology, $\mathcal{T}(\mathcal{X})$. Let S_i be the ith simplex generating this topology. Assume that there exists a continuously differentiable and invertible coordinate transformation, $w: [0, 1]^m \rightarrow [0, 1]^m$, denoted as $\bar{w} = w(\bar{y})$ such that

$$p(\bar{y}) - \det J_{\bar{w}} = 0 \tag{28}$$

where $J_{\bar{w}}$ is the Jacobian matrix whose (i, j)th component equals $\partial w_i / \partial y_j$.

With the above definitions, the probability of an observation being in simplex S_i will be given by

$$Pr(\bar{y} \in S_i) = \int_{S_i} p(\bar{y}) d\bar{y} \tag{29}$$

$$= \int_{w(S_i)} p[w^{-1}(\bar{w})] |\det J_w^{-1}(\bar{w})| \, d\bar{w} \tag{30}$$

This follows from a simple replacement of variables in the original integral (Spivak, 1965). In light of the assumption on the coordinate transformation it is apparent that this equation reduces to

$$\int_{S_i} p(\bar{y}) d\bar{y} = \int_{w(S_i)} d\bar{w}. \tag{31}$$

The set $w(S_i)$ is the w-image of the simple S_i. Equation (31) says that the probability of observation $\bar{y}$ being contained within simplex S_i will equal the volume of set $w(S_i)$.

Now consider the transformed observation process $\bar{w}(n) = w(\bar{y}(n))$. By construction, the vector $\bar{w}(n)$ must be uniformly distributed. The w-transformed prototypes will generate another process $\mathcal{W}(n)$. If it can be assumed that the w-images of the simplices are also approximately simplices (this should be approximately true if the size of the ensemble is large), then the transformed configuration process $\mathcal{W}(n)$ is identical to Markov chain studied in Subsections 4.1 and 4.2. The statistics for this process were computed in Subsection 4.2. In particular, these results indicate that even for nonuniform observation processes that the steady-state sample ensemble can again form an equi-probable partition of the observation space.

Another way of looking at the sample ensemble is that it can be used to obtain a nonparametric estimate of the observation density, $p(\bar{y})$. Let $v_i = E\{\int_{S_i} d\bar{y}\}$ be the expected volume of the ith simplex. Taking the expected value of eqn (31) with respect to all S_i yields the following equation

$$\frac{1}{N} = p(\bar{y}^*) v_i \tag{32}$$

where $\bar{y}^*$ is some vector in S_i. Rearranging terms yields $p(\bar{y}^*) = 1/(Nv_i)$, which can be used to define the following estimate for the observation density,

$$\hat{p}(\bar{y}) = \frac{1}{Nv_i} \tag{33}$$

where $\bar{y} \in S_i$. Equation (33) is a volumetric probability density estimate.

An important question concerns the sense in which the volumetric density estimate is optimal (i.e., prototypical). This question can be answered in the following manner. Let $\bar{y}$ denote the observation process. Let $i(n)$ denote the label of the simplex that contains the observation $\bar{y}(n)$ at time n. The sequence of simplex labels, i, therefore forms another stochastic process. Assuming N simplices, the average mutual information between the simplex process i and the observation process $\bar{y}$ will be given as

$$I(i, \bar{y}) = \sum_{i=1}^{N} \int_{\Omega} p(i, \bar{y}) \log \frac{p(i, \bar{y})}{q_i p(\bar{y})} d\bar{y} \tag{34}$$

$$= \sum_{i=1}^{N} \int_{\Omega} p(\bar{y}) P(i|\bar{y}) \log \frac{P(i|\bar{y})}{q_i} d\bar{y} \tag{35}$$

where q_i is the probability of the ith simplex, $p(\bar{y})$ is the observation density, $p(i, \bar{y})$ is the joint density of the ith simplex and the observation $\bar{y}$, and $P(i|\bar{y})$ is the conditional probability of simplex i given observation $\bar{y}$. From our prior results, it is clear that $q_i = 1/N$ and $P(i|\bar{y}) = I_i(\bar{y})$. The function $I_i(\bar{y})$ is an indicator function for the ith simplex (i.e., it is unity if $\bar{y}$ is in S_i and is zero otherwise). Using the fact that each simplex has a probability mass (volume) of $1/N$, we can therefore rewrite the average mutual information as

$$I(i; \bar{y}) = \sum_{i=1}^{N} \int_{\Omega} p(\bar{y}) I_i(\bar{y}) \log N I_i(\bar{y}) d\bar{y} \tag{36}$$

$$= \sum_{i=1}^{N} \frac{1}{N} \log N \tag{37}$$

$$= \log N. \tag{38}$$

From the divergence inequality, however, it is known that $I(i; \bar{y}) \leq \log N$. Therefore, the steady-state partition learned by the competitive sampling algorithm is optimal with regard to maximizing the average mutual information between the simplex process and the observation process.

6. SUMMARY

This paper has proposed an algorithm for competitive sampling of an IID process. The proposed algorithm uses a carefully defined notion of topological ordering and uses a randomized winner-take-all rule for updating prototypes. In that the sample ensemble has a topo-

logical ordering, the scheme is closely related to other topologically ordered competitive learning schemes. The principal contributions of this paper are a precise definition of the notion of topologically ordered ensembles and a rigorous analysis of the sampling procedure that yields quantitative predictions about the sampling process under the assumption of uniform observation densities. This paper also outlined a method for extending the analysis to nonuniform observation densities. In that a rigorous analysis of a competitive learning scheme was possible for topologically ordered competitive sampling, this work serves as a model for the design and analysis of other competitive learning schemes.

The significance of the preceding results can be assessed in terms of the significance of the future work suggested by these results. These future directions include a more rigorous examination of the sense in which competitive sampling produces prototypical ensembles, the relation between competitive and independent sampling in terms of space and sample complexity, and the comparison of the learning abilities of various topological orderings. Each of these directions is summarized below.

Competitive sampling is only valuable if it provides a collection of samples which are prototypical of the observation process. Intuitively, we expect a prototypical ensemble to form a model of the observation process that is optimal with regard to some well-defined performance measure. Prior results in Section 5 showed that the resulting ensemble is optimal in the sense of maximizing the average mutual information between the observation process and the set of activated simplices. Competitive sampling can therefore produce ensembles that preserve as much information as possible about the observation process given a fixed number of topologically ordered ensembles. There is, however, a classic tradeoff between this information (i.e., rate) and the distance between prototypes and observations (i.e., distortion). Although competitive sampling maximizes information rate, it does not minimize distortion. An interesting avenue for future research would involve competitive sampling procedures in which average distortion could be bounded while guaranteeing maximization of mutual information rate. The analysis and formalisms of this paper provide a starting point for that research.

The relationship between competitive sampling and independent sampling is an area for future inquiry that has not been fully addressed by this paper. As mentioned in the Introduction, independent sampling will generally require a large number of prototypes in order to control statistical fluctuations. According to prior results, competitive sampling produces an optimal ensemble for a fixed number of prototypes. This optimal ensemble, however, is only obtained after sorting through enough observations so that a consistent estimate of the simplex volumes, v_i, can be obtained. It can be easily argued, using the VC-inequality (Vapnik & Chervonenkis, 1971), that the sample complexity required to obtain a consistent estimate of v_i may be extremely large. In other words, it is quite possible that competitive sampling is buying reduced space complexity (i.e., the size of the ensemble) at the expense of increased sample complexity. How to control this tradeoff between independent and competitive sampling is an important area for future research. The analyses presented in this paper provide a set of quantitative results that should allow this tradeoff.

Finally, it should be noted that the success of this analysis relied heavily on the notion of topological ordering. The use of a topologically ordered ensemble was originally motivated by the realization that the constraints imposed by the ordering can be used to insure that the system has a single ergodic class. The results of this paper for arbitrary and multiresolution simplicial topologies demonstrate that the choice of topological ordering is crucial in the design of competitive learning algorithms whose behaviour can be analytically characterized. The particular ordering (multiresolution simplicial topology) used in this paper is just one topology and there is no reason not to expect other equally useful topologies to exist. The analysis and methods employed by this paper therefore provide a model for the design and evaluation of other topological-ordered competitive learning procedures.

REFERENCES

Burrascano, P (1991). Learning vector quantization for the probabilistic neural network. *IEEE Transactions on Neural Networks*, **2**, 458–461.

Cherkassky, V., & Lari-Najafi, H (1991) Constrained topological mapping for nonparametric regression analysis. *Neural Networks*, **4**, 27–40.

Kohonen, T (1982). Self-organized formation of topologically correct feature maps *Biological Cybernetics*, **43**, 59–69

Kelly, F. P. (1979). *Reversibility and stochastic networks* Chichester. John Wiley

Liu, C L (1968). *Introduction to combinatorial mathematics* New York McGraw-Hill Book Co.

Ritter, H , & Schulten, K. (1986) On the stationary state of Kohonen's self-organizing sensory mapping. *Biological Cybernetics*, **54**, 99–106.

Ritter, H., & Schulten, K. (1988). Convergence properties of Kohonens topology conserving maps: Fluctuations, stability, and dimension selection. *Biological Cybernetics*, **60**, 59–71

Rumelhart, D., & Zipser, D. (1986) Feature discovery by competitive learning. In D Rumelhart & J McClelland (Eds), *Parallel distributed processing Explorations in the microstructure of cognition* Cambridge. MIT Press.

Specht, D F (1990). Probabilistic neural networks. *Neural Networks*, **3**, 109–118.

Spivak, M. (1965) *Calculus on manifolds* Menlo Part· W A. Benjamin Inc.

Vapnik, V N , & Chervonenkis, A. Y. (1971) On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, **16**, 264–280.