MARKOV CHAINS IN MANY DIMENSIONS

by

Dimitris N. Politis
Purdue University

Technical Report #92-02

Department of Statistics
Purdue University

January 1992

◇

Markov chains in many dimensions

Dimitris N. Politis
Department of Statistics
Purdue University
W. Lafayette, IN 47907

Abstract

A generalization of the notion of a stationary Markov chain in more than one dimensions is proposed, and is found to be a special class of homogeneous Markov random fields. Stationary Markov chains in many dimensions are shown to possess a maximum entropy property, analogous to the corresponding property for Markov chains in one dimension. In addition, a representation of Markov chains in many dimensions is provided, together with a method for their generation that converges to their stationary distribution.

AMS 1980 subject classifications. Primary 62M20; secondary 62M10.

Key words and phrases. Markov chains, random fields, random sequences.

1. Introduction

Stationary Markov chains have long been among the most favorite models for sequences $\{X(i), i \in \mathbf{Z}\}$ of random variables. Stated in words, a stationary sequence is said to be Markov if the conditional distribution of $X(0)$ given the values of $X(i)$ with $i < 0$, depends only on the value of $X(-1)$.

Supposing that, for any $m \in \mathbf{N}$, the random variables $X(1), \ldots, X(m)$ possess a density f with respect to some σ-finite measure, the Markov chain property for stationary sequences reads: for any $m \in \mathbf{N}$,

$$f(X(0)|X(i), -m < i < 0) = f(X(0)|X(-1)) \tag{1}$$

where $f(X(0)|X(i), -m < i < 0)$ denotes the conditional density of $X(0)$ given the values of $X(-1), \ldots, X(-m+1)$. If instead of (1) the sequence satisfies $f(X(0)|X(i), -m < i < 0) = f(X(0)|X(i), -r \leq i < 0)$, for any $m > r$, then it referred to as a Markov chain of order r.

For homogeneous random fields in the plane $\{X(i,j), i \in \mathbf{Z}, j \in \mathbf{Z}\}$, a slightly different definition applies. In particular, a homogeneous random field is said to be Markov if the conditional distribution of $X(0,0)$ given the values of $X(i,j)$ with $(i,j) \neq (0,0)$, depends only on the immediate neighbors of point $(0,0)$, where the set of 'immediate neighbors' is the set of points closest to $(0,0)$ in some metric. For the rigorous definition and extension to higher order Markov random fields, see Dobrushin (1968).

Again assuming that the $X(i,j)$ random variables possess densities, the Markov property for homogeneous random fields reads: for any $m \in \mathbf{N}$,

$$f(X(0,0)|X(i,j), |i| < m, |j| < m, (i,j) \neq (0,0)) = f(X(0,0)|X(-1,0), X(0,1), X(1,0), X(0,-1)) \tag{2}$$

where Euclidean distance was used to find the points closest to $(0,0)$. An analogous definition of the Markov property holds for general homogeneous fields in n dimensions, i.e. $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$.

For a stationary sequence $X(i)$, (which is a homogeneous random field in one dimension), to be a Markov random field according to Dobrushin's definition it must satisfy

$$f(X(0)|X(i), 0 < |i| < m) = f(X(0)|X(-1), X(1)) \tag{3}$$

for any $m \in \mathbf{N}$. It is apparent that the Markov random field property (3) is weaker than the Markov chain property (1), and a stationary sequence $X(i)$ might satisfy (3), without necessarily being a Markov chain. Consider Dobrushin's example where a (non-stationary) sequence has only two and equiprobable sample paths, namely $(\ldots, 0, 0, 0, 0, 0, 0, \ldots)$ and $(\ldots, 0, 0, 1, 2, 0, 0, \ldots)$. For a stationary example, consider a binary (taking on the values 0 and 1) stationary sequence having only three and equiprobable sample paths, i.e., assume that $\{X(0), X(1), X(2), \ldots\}$ can be $\{0, 1, 1, 0, 1, 1, 0, 1, 1, 0, , \ldots\}$, $\{1, 1, 0, 1, 1, 0, 1, 1, 0, 1, \ldots\}$, or $\{1, 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots\}$ with equal probability. It is immediate that this stationary sequence is a Markov chain of order 2, satisfying (3), without satisfying (1). In contrast, it can be easily checked (using the chain rule of probabilities) that a stationary sequence satisfying the Markov chain property (1), satisfies also the Markov property for random fields (3).

Specifically, the Markov chain property (1) is a 'predictive' property, while (3) is an 'interpolative' property. The situation is also illustrated if we consider the case where $X(t)$ is a Gaussian Markov chain of order r, generated by the linear autoregressive model

$$X(t) + \sum_{i=1}^{r} a_i X(t-i) = Z(t) \tag{4}$$

where $Z(t), t \in \mathbf{Z}$, is a sequence of i.i.d. $N(0,1)$ random variables. Then the optimal (in the Hilbert space sense) predictor of $X(0)$ given the values of $X(i)$ with $i < 0$ is

$$E(X(0)|X(i), i < 0) = E(X(0)|X(i), -r \leq i < 0) = -\sum_{i=1}^{r} a_i X(-i)$$

and the optimal interpolator given the values of $X(i)$ with $i \neq 0$ is

$$E(X(0)|X(i), i \neq 0) = E(X(0)|X(i), 0 < |i| \leq r) = -\sum_{0<|i|\leq r} \tilde{\rho}_i X(-i)$$

where $\tilde{\rho}_i$ is the inverse autocorrelation sequence (see e.g. Politis (1992)). It is then easy to see that the sequence $X(t)$ also satisfies the 'interpolating' Markov equation

$$X(t) + \sum_{0<|i|\leq r} \tilde{\rho}_i X(t-i) = U(t) \tag{5}$$

where $U(t), t \in \mathbf{Z}$, is a sequence of mean zero, normal random variables, independent of the

$X(t)$ sequence, and with autocovariances satisfying

$$\frac{E[U(0)U(t)]}{E[U(0)]^2} = \begin{cases} 1 & \text{for } t = 0 \\ \tilde{\rho}_t & \text{for } 0 < |t| \leq r \\ 0 & \text{for } |t| > r \end{cases}$$

In his pioneering paper, Whittle (1954) considered both the 'predictive' and the 'interpolative' Markov models (4) and (5), and argued philosophically that, because time has a natural 'direction', model (4) is suitable for time series, while a generalization of model (5) to the random field case (in the spirit of (3)) is more suitable for spatial processes. A different philosophical position was expressed by H. G. Wells in his novel 'The Time Machine', where he claimed that the only distinction between time and space is that we can not move in time, whereas we do move in space.

Of course, a more pragmatic approach is to just consider model (4) as more restrictive than model (5) and study its properties. In fact, Tjøstheim (1978, 1983) studied the extension of the 'predictive' (unilateral) autoregressive model (4) to the case of random fields in many dimensions, and obtained many important results concerning estimation of parameters and prediction, that are in parallel to the one-dimensional case. Since autoregressive processes are special cases of Markov processes, an extension of the general Markov chain property (1) to the case of random fields in many dimensions might also be worth-while, taking into account the popularity of Markov chain models for sequences. This is the subject of the following sections.

2. Some definitions

Suppose $\{X(\mathbf{t}), \mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbf{Z}^n\}$ is a homogeneous (stationary, shift invariant) random field in n dimensions, with $n \in \mathbf{N}$, i.e. a collection of random variables $X(\mathbf{t})$ taking values in the set $S \subset \mathbf{R}$, defined on a probability space $(\Omega, \mathcal{A}, P)$, and indexed by the variable $\mathbf{t} \in \mathbf{Z}^n$. In the particular cases where $n = 1$ or 2, we might equivalently denote the random field by $\{X(i), i \in \mathbf{Z}\}$, and $\{X(i,j), i \in \mathbf{Z}, j \in \mathbf{Z}\}$ respectively.

It will be assumed that the marginal distribution function of $X(\mathbf{t})$ possesses a density f with respect to a measure ν on $\mathbf{R}$. In general, it will be assumed that for any finite set $A \subset \mathbf{Z}^n$, with cardinality m, the marginal distribution function of $\{X(\mathbf{t}), \mathbf{t} \in A\}$ possesses a density with respect to the product measure ν^m on $\mathbf{R}^m$, and the generic notation for this density will be f.

We will now recursively define a certain notion of the 'past' of $X(\mathbf{t})$ in the direction j (where $1 \leq j \leq n$) in many dimensions. So, let $\mathbf{u} = (u_1, \ldots, u_n)$, and for $n = 1$ (and hence $j = 1$ too) define

$$Past_1^{(1)}(X(\mathbf{t})) = \{X(\mathbf{u}) : u_1 < t_1\} \tag{6}$$

For a general $n > 1$ define

$$Past_j^{(n)}(X(\mathbf{t})) = \{X(\mathbf{u}) : u_j < t_j\} \cup Past_{[j-1]^\star}^{(n-1)}(X(\mathbf{t}|_{t_j = u_j})) \tag{7}$$

where $(X(\mathbf{t}|_{t_j=u_j}))$ is the random field in $n-1$ dimensions obtained by observing the random field $\{X(\mathbf{t}\}$ only at points in the hyperplane $t_j = u_j$ in $\mathbf{Z}^n$, and the function $[\cdot]^\star$ is defined by

$$[j-1]^\star = \begin{cases} j-1 & \text{if } 1 < j \leq n \\ n & \text{if } j = 1 \end{cases}$$

For example, in two dimensions ($n = 2$)

$$Past_j^{(2)}(X(\mathbf{t})) = \{X(\mathbf{u}) : u_j < t_j\} \cup \{X(\mathbf{u}) : u_j = t_j, u_k < t_k, k \neq j\} \tag{8}$$

We will also define

$$Future_j^{(n)}(X(\mathbf{t})) = \{X(\mathbf{u}) : X(\mathbf{u}) \notin Past_j^{(n)}(X(\mathbf{t}))\} - \{X(\mathbf{t})\} \tag{9}$$

to be the 'future' of $X(\mathbf{t})$ in the direction j.

The general 'past' of point $X(\mathbf{t})$ in n dimensions is defined to be

$$Past^{(n)}(X(\mathbf{t})) = \cap_{j=1}^{n} Past_j^{(n)}(X(\mathbf{t})) \tag{10}$$

In all cases, if the dimension n is apparent, the superscript (n) will be omitted.

The reason behind the above definition of $Past_j^{(n)}(X(\mathbf{t}))$ is to construct a particular ordering of the elements of a cube of points in $\mathbf{Z}^n$.

Lemma 1 *Let N be any positive integer, and let C_N the cube of points $\mathbf{t} \in \mathbf{Z}^n$ whose coordinates satisfy $|t_k| \leq N$. If $\mathbf{t}, \mathbf{u}$ are two points in C_N, (with $\mathbf{t} \neq \mathbf{u}$), then*

$$|Past_j^{(n)}(X(\mathbf{t})) \cap C_N| \neq |Past_j^{(n)}(X(\mathbf{u})) \cap C_N|$$

for any 'direction' $1 \leq j \leq n$, where $|\cdot|$ denotes cardinality. In particular, $|Past_j^{(n)}(X(\mathbf{u})) \cap C_N| < |Past_j^{(n)}(X(\mathbf{t})) \cap C_N|$ if and only if $X(\mathbf{u}) \in Past_j^{(n)}(X(\mathbf{t})) \cap C_N$.

Proof. (Induction.) For $n = 1$ (or 2) the Lemma is obviously true. Suppose it is true for $n = k - 1 > 1$. From equation (7) it follows that if $u_j < t_j$, then $Past_j^{(k)}(X(\mathbf{u}))$ is strictly included in $Past_j^{(k)}(X(\mathbf{t}))$. Similarly, if $u_j > t_j$, $Past_j^{(k)}(X(\mathbf{t}))$ is strictly included in $Past_j^{(k)}(X(\mathbf{u}))$. If finally $u_j = t_j$, then $|Past_j^{(k)}(X(\mathbf{t})) \cap C_N| \neq |Past_j^{(k)}(X(\mathbf{u})) \cap C_N|$ by the induction hypothesis. □

The above Lemma shows that the notion of 'past in the direction j' can be used to construct an ordering of the elements in C_N. The following chain rule of probabilities using this ordering will be the main tool in dealing with Markov chains in many dimensions.

Lemma 2 (Chain rule 'in the direction j'.)

The joint density of the random variables $X(\mathbf{t}), \mathbf{t} \in C_N$ can be expanded as

$$f(X(\mathbf{t}), \mathbf{t} \in C_N) = \prod_{\mathbf{t} \in C_N} f(X(\mathbf{t}) | C_N \cap Past_j(X(\mathbf{t}))) \tag{11}$$

The ordering 'in the direction j' and the associated chain rule in two dimensions ($n = 2$) have been used previously (cf. Anastassiou and Sakrison (1982)), in the context of the computation of the entropy rate of a random field in the plane (see Section 4 for a definition).

We will also make use of the l_p norm and distance in $\mathbf{Z}^n$, and for any two points $\mathbf{t}$ and $\mathbf{u}$ in $\mathbf{Z}^n$, define for $1 \leq p < \infty$

$$d_p(\mathbf{t}, \mathbf{u}) = (\sum_j |t_j - u_j|^p)^{1/p} \tag{12}$$

and

$$d_\infty(\mathbf{t}, \mathbf{u}) = \sup_j |t_j - u_j| \tag{13}$$

and for two sets $\mathbf{E}_1, \mathbf{E}_2$ in $\mathbf{Z}^n$, define for $p \in [1, \infty]$

$$d_p(\mathbf{E}_1, \mathbf{E}_2) = \inf\{d_p(\mathbf{t}, \mathbf{u}) : \mathbf{t} \in \mathbf{E}_1, \mathbf{u} \in \mathbf{E}_2\} \tag{14}$$

Let

$$\Pi_p^{(r)}(X(\mathbf{t})) = \{X(\mathbf{u}) : 0 < d_p(\mathbf{t}, \mathbf{u}) \leq r\} \tag{15}$$

be the set of r-close neighbors, and $\Pi_p(X(\mathbf{t})) = \Pi_p^{(1)}(X(\mathbf{t}))$ the set of immediate neighbors of $X(\mathbf{t})$. Also let

$$L_p^{(r)}(X(\mathbf{t})) = \Pi_p^{(r)}(X(\mathbf{t})) \cap Past^{(n)}(X(\mathbf{t})) \tag{16}$$

be the set of r-close neighbors contained in the past of $X(\mathbf{t})$, and $L_p(X(\mathbf{t})) = L_p^{(1)}(X(\mathbf{t}))$.

We will say that the homogeneous random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is a *Markov chain in the direction* j if

$$f(X(\mathbf{t})|Past_j(X(\mathbf{t}))) = f(X(\mathbf{t})|L_p(X(\mathbf{t}))) \tag{17}$$

and a *Markov chain of order* r *in the direction* j if

$$f(X(\mathbf{t})|Past_j(X(\mathbf{t}))) = f(X(\mathbf{t})|L_p^{(r)}(X(\mathbf{t}))) \tag{18}$$

We will also say that $\{X(\mathbf{t})\}$ is a *Markov chain* if it is a Markov chain in all directions, and a *Markov chain of order* r it is a Markov chain of order r in all directions. Note that the Markov neighborhoods depend on the norm l_p chosen, so that a more accurate statement would include the value of p if it is not apparent.

We close this section with an example of a Markov chain in the plane (more examples will be given in Section 5). For any fixed i, let $\{X(i,j), j \in \mathbf{Z}\}$ be a stationary Markov chain, and let the Markov chains $\{X(i,\cdot), i \in \mathbf{Z}\}$ be independent and identically distributed. Then $X(i,j)$ can be verified to be a Markov chain in two dimensions (with respect to Euclidean distance).

3. Markov chains and Markov random fields

We now show that if the homogeneous random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is a Markov chain, then it is also a Markov random field in the sense of Dobrushin (1968), for any $n \in \mathbf{N}$.

Theorem 1 *Suppose the homogeneous random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is a Markov chain of order r. Let N be any positive integer with $N > r$, and let C_N the cube of points $\mathbf{t} \in \mathbf{Z}^n$ whose coordinates satisfy $|t_k| \leq N$. Then it is true that*

$$f(X(\mathbf{0})|X(\mathbf{u}), \mathbf{u} \in C_N - \{\mathbf{0}\}) = f(X(\mathbf{0})|\Pi_p^{(r)}(X(\mathbf{0}))) \tag{19}$$

Proof. We will repeatedly make use of the chain rule 'in the direction j' expansion of probabilities that was given in Lemma 2. Let $M = N + r$, and write

$$f(X(\mathbf{0})|X(\mathbf{u}), \mathbf{u} \in C_M - \{\mathbf{0}\}) = \frac{f(X(\mathbf{u}), \mathbf{u} \in C_M)}{f(X(\mathbf{u}), \mathbf{u} \in C_M - \{\mathbf{0}\})} \tag{20}$$

$$= \frac{f(X(\mathbf{u}), \mathbf{u} \in C_N|S_P)}{\int f(X(\mathbf{0}) = x, X(\mathbf{u}), \mathbf{u} \in C_N - \{\mathbf{0}\}|S_P)d\nu(x)} \tag{21}$$

$$= [\int \frac{f(X(\mathbf{0}) = x, X(\mathbf{u}), \mathbf{u} \in C_N - \{\mathbf{0}\}|S_P)}{f(X(\mathbf{u}), \mathbf{u} \in C_N|S_P)} d\nu(x)]^{-1} \tag{22}$$

where $S_P = \{X(\mathbf{u}), \mathbf{u} \in C_M - C_N\}$, and we have assumed without loss of generality that $f(X(\mathbf{u}), \mathbf{u} \in C_M - \{\mathbf{0}\}) > 0$.

The reason for adding this extra conditioning on S_P is to have all points in C_N have full L_p neighborhoods in their (limited by C_N) 'past'. Define $Past_j^N(X(\mathbf{t})) = C_N \cap Past_j(X(\mathbf{t}))$, and $Future_j^N(X(\mathbf{t})) = C_N \cap Future_j(X(\mathbf{t}))$. Then, using the chain rule 'in the direction j' to expand the joint (conditional) density

$$f(X(\mathbf{u}), \mathbf{u} \in C_N|S_P) = \prod_{\mathbf{u} \in C_N} f(X(\mathbf{u})|S_P \cup Past_j^N(X(\mathbf{u})))$$

in both numerator and denominator in (22), and after cancellations between numerator and denominator, it is immediate that the quantity in (22) is equal to

$$[\int \frac{f(X(\mathbf{0}) = x|S_P \cup Past_j^N(X(\mathbf{0}))) \prod_{\mathbf{t} \in Future_j^N(X(\mathbf{0}))} f(X(\mathbf{t})|S_P \cup Past_j^N(X(\mathbf{t})), X(\mathbf{0}) = x)}{f(X(\mathbf{0})|S_P \cup Past_j^N(X(\mathbf{0}))) \prod_{\mathbf{t} \in Future_j^N(X(\mathbf{0}))} f(X(\mathbf{t})|S_P \cup Past_j^N(X(\mathbf{t})))} d\nu(x)]^{-1} \tag{23}$$

where $Past_j^N(X(\mathbf{t})) = C_N \cap Past_j(X(\mathbf{t}))$, and $Future_j^N(X(\mathbf{t})) = C_N \cap Future_j(X(\mathbf{t}))$.

Observe that points $\mathbf{t} \in Future_j^N(X(\mathbf{0}))$ are of three kinds:

(a) $S_a = \{\mathbf{t} \in Future_j^N(X(\mathbf{0})) : L_p^{(r)}(X(\mathbf{t})) \not\subset C_N\}$, i.e. points that are r-close to the 'side' (boundary) of C_N that is parallel to the direction j;

(b) $S_b = S_a^c \cap \{\mathbf{t} \in Future_j^N(X(\mathbf{0})) : X(\mathbf{0}) \notin L_p^{(r)}(X(\mathbf{t}))\}$, where $S_a^c = \{\mathbf{t} \in Future_j^N(X(\mathbf{0})) : L_p(X(\mathbf{t}) \subset C_N\}$, i.e. points that are not r-close to the boundary, and whose $L_p^{(r)}$ neighborhood does not involve $X(\mathbf{0})$;

(c) $S_c = S_a^c \cap \{\mathbf{t} \in Future_j^N(X(\mathbf{0})) : X(\mathbf{0}) \in L_p^{(r)}(X(\mathbf{t}))\}$, i.e. points that are not r-close to the boundary, and whose $L_p^{(r)}$ neighborhood does involve $X(\mathbf{0})$.

Using the Markov chain property 'in the direction j' it is obvious that contributions from points in S_b cancel out between the numerator and denominator in (23). The points in S_a are a nuisance, but note that their contributions also cancel out between the numerator and denominator, because of the added conditioning on S_P. Thus, the quantity in (22) is a function of *only* the variables in $S_c \cup \{X(\mathbf{0})\}$ and their respective L_p neighborhoods, that notably are either in $\Pi_p(X(\mathbf{0})) \cup \{X(\mathbf{0})\}$, or can be obtained by points in $\Pi_p(X(\mathbf{0})) \cup \{X(\mathbf{0})\}$ by 'moving' a distance of at most r, in a direction *other than* j.

Now repeating the above procedure by taking chain rule expansions in all directions $1 \leq j \leq n$, (and using the symmetries of $\Pi_p^{(r)}(X(\mathbf{0}))$ around any hyperplane defined by setting one coordinate of $\mathbf{Z}^n$ to zero), it is shown that the quantity in (22) is a function of only the variables in $\Pi_p(X(\mathbf{0})) \cup \{X(\mathbf{0})\}$. This implies that the extra conditioning on S_P is immaterial, and that

$$f(X(\mathbf{0})|X(\mathbf{u}), \mathbf{u} \in C_N - \{\mathbf{0}\}) = f(X(\mathbf{0})|X(\mathbf{u}), \mathbf{u} \in C_M - \{\mathbf{0}\}) = f(X(\mathbf{0})|\Pi_p^{(r)}(X(\mathbf{0})) \quad (24)$$

and the theorem is proven.□

As an illustration, consider the case $n = 2, r = 1$, and $p = \infty$, in which case the Π_p neighborhood is of square form (see Figure 1). Then the chain rule 'in the direction 2' is associated with the ordering 'in the direction 2'

$$(-N, -N), (-N, -N+1), \ldots, (-N, N),$$

$$(-N+1, -N), (-N+1, -N+1), \ldots, (-N+1, N)$$

....

$$(N,-N),(N,-N+1),\ldots,(N,N)$$

and by a chain rule 'in the direction j', (with $j=1$ or 2), we see that the quantity in (22) is a function of only the variables in $\Pi_\infty(X(\mathbf{0}))\cup\{X(\mathbf{0})\}$.

To further elaborate, consider also the case where $n=2, r=1$, and $p<\infty$, in which case the Π_p neighborhood is of diamond shape, (it is essentially a l_1 neighborhood). This case is of practical significance because it leads to a dependence structure associated with the well-known Ising model for random fields.

In the same fashion, using the chain rule expansion 'in the direction 1', it is obtained that the quantity in (22) is a function of only the variables $\Pi_1(X(\mathbf{0}))\cup\{X(0,0),X(1,1),X(-1,-1)\}$. Using the chain rule 'in the direction 2', the quantity in (22) is shown to be a function of only the variables $\Pi_1(X(\mathbf{0}))\cup\{X(0,0),X(1,-1),X(-1,1)\}$. Hence the quantity in (22) is a function of only the variables $\Pi_1(X(\mathbf{0}))\cup\{X(0,0)\}$ as desired.

Remark. Dobrushin's (1968) definition is actually more restrictive than (19), in that the conditional distribution of $\{X(\mathbf{t}),\mathbf{t}\in A\}$ given the values of $\{X(\mathbf{t}),\mathbf{t}\notin A\}$ should depend only on the immediate neighbors of set A, for any finite set A. This stronger property can be seen to follow from a stronger (involving any finite set A) definition of the Markov chain property of equation (17).

From the proof of Theorem 1 it is apparent that even if the homogeneous random field $\{X(\mathbf{t}),\mathbf{t}\in\mathbf{Z}^n\}$ is a Markov chain in *at least one* direction, then it is also a Markov random field in the sense of Dobrushin (1968), although the Markov random field dependence neighborhoods are larger and not necessarily symmetrical. The following Corollary can be proved by a simple chain rule expansion in the direction j, analogously to Theorem 1.

Corollary 1 *Suppose the homogeneous random field $\{X(\mathbf{t}),\mathbf{t}\in\mathbf{Z}^n\}$ is a Markov chain of order r in the direction j. Let N be any positive integer with $N>2r$, and let C_N the cube of points $\mathbf{t}\in\mathbf{Z}^n$ whose coordinates satisfy $|t_k|\le N$. Then it is true that*

$$f(X(\mathbf{0})|X(\mathbf{u}),\mathbf{u}\in C_N-\{\mathbf{0}\})=f(X(\mathbf{0})|\Pi_p^{(2r)}(X(\mathbf{0}))) \quad (25)$$

Remark. As a matter of fact, the Markov random field dependence neighborhood will be a subset of $\Pi_p^{(2r)}(X(\mathbf{0}))$, and this subset can be identified given the specifics of the problem. For example, consider the case where $n = 2, r = 1$, and $p = 1$, and the random field $\{X(\mathbf{t})\}$ is a Markov chain in the direction 2. Using the chain rule 'in the direction 2', it was shown that the quantity in (22) is a function of only the set of variables $\Pi_1(X(\mathbf{0})) \cup \{X(0,0), X(1,-1), X(-1,1)\}$. This implies, in the particular case, that

$$f(X(\mathbf{0})|X(\mathbf{u}), \mathbf{u} \in C_N - \{\mathbf{0}\}) = f(X(\mathbf{0})|A)$$

where $A{=}\Pi_1(X(\mathbf{0})) \cup \{X(1,-1), X(-1,1)\}$, which is quite smaller than the set $\Pi_1^{(2)}(X(\mathbf{0}))$.

An analog to Corollary 1 was proved in Abend *et al.* (1965), for a type of binary random field on a finite grid on the plane called a Markov 'mesh'. Their definition of a Markov 'mesh' is related (with a different notion of the 'Past') to the definition of a Markov chain in one (of the two) directions in the plane. That the notion of the 'Past' as defined in equation (7) is a most useful one will be demonstrated in the next section.

4. The maximum entropy property of Markov chains

In the literature of Gibbs measures (see for example the very thorough exposition in Georgii (1988) and the references therein), it is well known that homogeneous Markov random fields (in the sense of Dobrushin) possess a certain maximum entropy property. However, in the case of random fields in one dimension (i.e. sequences), it has been shown (cf. Spitzer (1972)) that it is actually stationary Markov *chains* that possess the maximum entropy property. In this section it is shown that the maximum entropy property is intimately associated with our notion of a Markov chain in many dimensions.

Some definitions are in order. Let A be a finite set in $\mathbf{Z}^n$ with cardinality m. The entropy of the set of variables $\{X(\mathbf{t}), \mathbf{t} \in A\}$ is defined by (cf. Pinsker (1964))

$$H(X(\mathbf{t}), \mathbf{t} \in A) = -\int f(X(\mathbf{t}), \mathbf{t} \in A) \log f(X(\mathbf{t}), \mathbf{t} \in A) d\nu^m \tag{26}$$

Similarly, the conditional entropy of the random variable $X(\mathbf{u})$ given $\{X(\mathbf{t}), \mathbf{t} \in A\}$ is defined by

$$H(X(\mathbf{u})|X(\mathbf{t}), \mathbf{t} \in A) = -\int f(X(\mathbf{u}), X(\mathbf{t}), \mathbf{t} \in A) \log f(X(\mathbf{u})|X(\mathbf{t}), \mathbf{t} \in A) d\nu^{m+1} \tag{27}$$

where $f(X(\mathbf{u})|X(\mathbf{t}), \mathbf{t} \in A)$ is the conditional density of $X(\mathbf{u})$ given the values of $\{X(\mathbf{t}), \mathbf{t} \in A\}$.

The random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is said to have an entropy rate

$$h_X = \lim_{N \to \infty} \frac{H(X(\mathbf{t}), \mathbf{t} \in C_N)}{|C_N|} \tag{28}$$

provided the limit exists, where C_N is the cube defined in Theorem 1, with cardinality $|C_N| = (2N+1)^n$. For homogeneous random fields it can be shown (cf. Georgii (1988)) that the limit exists (in $[-\infty, \infty]$), and can alternatively be calculated as

$$h_X = \lim_{N \to \infty} \frac{H(X(\mathbf{t}), \mathbf{t} \in C_N^+)}{N^n} \tag{29}$$

where C_N^+ is the cube of points $\mathbf{t} \in \mathbf{Z}^n$ whose coordinates satisfy $0 < t_k \leq N$.

Theorem 2 *Suppose the homogeneous random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ has the same $L_p^{(r)}$ marginal density with the homogeneous random field $\{W(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$, i.e. suppose the distribution of the* *set of variables* $\{X(\mathbf{0})\} \cup \{X(\mathbf{t}), \mathbf{t} \in L_p^{(r)}(X(\mathbf{0}))\}$ *is identical to that of* $\{W(\mathbf{0})\} \cup \{W(\mathbf{t}), \mathbf{t} \in L_p^{(r)}(W(\mathbf{0}))\}$. *Assume also that* $-\infty < H(X(\mathbf{0})) < \infty$. *If the field* $\{W(\mathbf{t})\}$ *is a Markov chain of order* r *in the direction* j *(with respect to the neighborhood* $L_p^{(r)}$*), then* $h_X \leq h_W$.

Proof. The proof uses a combination of methods from Anastassiou and Sakrison (1982), Choi and Cover (1984), and Politis (1991).

Let an integer $N > r$, and consider the density $f(X(\mathbf{u}), \mathbf{u} \in C_N^+)$, and expand it using a chain rule 'in the direction j' (cf. Lemma 2). This would yield the following chain rule for entropies

$$H(X(\mathbf{t}), \mathbf{t} \in C_N^+) = \sum_{\mathbf{t} \in C_N^+} H(X(\mathbf{t}) | C_N^+ \cap Past_j(X(\mathbf{t}))) \tag{30}$$

However, using the fact that $H(A|B,C) \leq H(A|B)$ for any three random elements A, B, C ('conditioning reduces entropy', cf. Pinsker (1964)), we have that

$$\sum_{\mathbf{t} \in C_N^+} H(X(\mathbf{t}) | C_N^+ \cap Past_j(X(\mathbf{t}))) \leq \sum_{\mathbf{t} \in C_N^+} H(X(\mathbf{t}) | C_N^+ \cap L_p^{(r)}(X(\mathbf{t}))) \tag{31}$$

Note that the computation of the RHS of (31) requires knowledge only of the distribution of the set of variables $\{X(\mathbf{0})\} \cup \{X(\mathbf{t}), \mathbf{t} \in L_p^{(r)}(X(\mathbf{0}))\}$, which is (by assumption) identical to that of $\{W(\mathbf{0})\} \cup \{W(\mathbf{t}), \mathbf{t} \in L_p^{(r)}(W(\mathbf{0}))\}$. It follows that

$$\sum_{\mathbf{t} \in C_N^+} H(X(\mathbf{t}) | C_N^+ \cap L_p^{(r)}(X(\mathbf{t}))) = \sum_{\mathbf{t} \in C_N^+} H(W(\mathbf{t}) | C_N^+ \cap L_p^{(r)}(W(\mathbf{t}))) \tag{32}$$

To complete the proof, observe that by a chain rule 'in the direction j' expansion of the density $f(W(\mathbf{u}), \mathbf{u} \in C_N^+)$ one obtains

$$H(W(\mathbf{t}), \mathbf{t} \in C_N^+) = \sum_{\mathbf{t} \in C_N^+} H(W(\mathbf{t}) | C_N^+ \cap Past_j(W(\mathbf{t})) \tag{33}$$

Note that, due to the Markov assumption, the sum in the RHS of (33) differs from that in the RHS of (32) only in the contributions of the rN^{n-1} points that are r-close to the sides (boundary) of C_N that are parallel to the direction j, i.e. the points whose $L_p^{(r)}$ neighborhood is not completely included in C_N^+; (see Figure 2 for a two-dimensional illustration). Because of the assumed boundedness of the entropy we can then write

$$H(X(\mathbf{t}), \mathbf{t} \in C_N^+) \le H(W(\mathbf{t}), \mathbf{t} \in C_N^+) + O(rN^{n-1}) \tag{34}$$

Dividing by N^n and taking limits in the above asymptotic inequality the theorem is proven. □

The interpretation of Theorem 2 is that a stationary Markov chain $\{W(\mathbf{t})\}$ in n dimensions has maximum entropy among all homogeneous random fields that share with $\{W(\mathbf{t})\}$ the same 'local characteristics', with respect to the neighborhood $L_p^{(r)}$.

5. Stationary distributions and recursive generation of Markov chains

A natural question to ask regarding a Markov chain is whether it has a stationary distribution. Although, under regularity conditions, Markov sequences have unique stationary distributions, the situation for Markov random fields on the plane (or higher dimensions) is quite different (cf. Kindermann and Snell (1980)).

For example, consider a Markov Ising (that is, satisfying equation (2)) binary random field on the square C_N in the plane ($n = 2$). Under some conditions, the dependence of the value $X(0,0)$ on the initially chosen boundary conditions (on the outside of C_N) does not diminish as $N \to \infty$. Intuitively, this is due to the fact that the boundary increases in size as $N \to \infty$, as well as to the 'interpolative' character of equation (2).

The situation is a bit more straightforward for Markov chains in many dimensions, essentially because the chains can be generated in a certain 'direction' (the main diagonal), and are not interpolated.

Consider a binary (taking on the values 0 and 1) homogeneous Markov chain $X(i,j)$ on the plane that satisfies

$$f(X(0,0)|Past_j(X(0,0)) = f(X(0,0)|X(-1,0), X(0,-1)) \tag{35}$$

for $j = 1$ or 2. Note that, by Theorem 1, the Markov chain $X(i,j)$ is *also* an Ising model, i.e. it also satisfies (2).

Let $\mathbf{Y}_k$ be the sequence $\{X(i,j) : i+j = k\}$, i.e. an infinite 'strip' of $X(i,j)$ points, running along a line of 135^o angle with respect to the first axis. Also let Y_k be a real number in [0,2] that has as a binary expansion the sequence $\mathbf{Y}_k$; for instance, let $Y_k = \sum_{m=0}^{\infty} c_m^{(k)} 2^{-m}$, where we define

$$c_m^{(k)} = \begin{cases} X(-m/2, k+m/2) & \text{if } m \text{ is even} \\ X(\frac{m+1}{2}, k - \frac{m+1}{2}) & \text{if } m \text{ is odd} \end{cases}$$

Obviously, the sequences $\{\mathbf{Y}_k, k \in \mathbf{Z}\}$ and $\{Y_k, k \in \mathbf{Z}\}$ are in one-to-one correspondence with each other, and they are both stationary.

Observe that, given the value of $\mathbf{Y}_0$, the values of $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ can be recursively generated. As a matter of fact, the sequence $\{\mathbf{Y}_k, k \in \mathbf{Z}\}$, is a Markov sequence, and hence so is the sequence $\{Y_k, k \in \mathbf{Z}\}$. But, under regularity conditions (cf. Doob (1953)), the real-valued Markov sequence $\{Y_k, k \in \mathbf{Z}\}$ possesses a unique stationary distribution. Hence, so does the sequence $\{\mathbf{Y}_k, k \in \mathbf{Z}\}$, as well as the homogeneous Markov chain $X(i,j)$ that is generated recursively in the above manner.

Now consider another binary homogeneous Markov chain $X(i,j)$ on the plane that satisfies (for $j = 1$ or 2)

$$f(X(0,0)|Past_j(X(0,0)) = f(X(0,0)|X(-1,0), X(0,-1), X(-1,-1)) \tag{36}$$

Defining the sequences $\{\mathbf{Y}_k, k \in \mathbf{Z}\}$ and $\{Y_k, k \in \mathbf{Z}\}$ as before, it is immediate that these two sequences are now Markov of order 2. Thus, given the value of $\mathbf{Y}_0$ and $\mathbf{Y}_1$, the values of $\mathbf{Y}_2, \mathbf{Y}_3, \ldots$ can be recursively generated, and (under some regularity conditions) the sequence $\{\mathbf{Y}_k, k \in \mathbf{Z}\}$, as well as the field $X(i,j)$, will tend to the unique stationary distribution.

Both examples suggest that Markov chains in many dimensions can be generated recursively along the main diagonal in $\mathbf{Z}^n$, using the values of the chain on hyperplanes perpendicular to the main diagonal. This property of directional generation along the main diagonal makes l_1 the 'natural' choice for distance on $\mathbf{Z}^n$. Of course, if a homogeneous field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is a Markov chain of order r' with respect to distance l_p, then there is a smallest $r \geq r'$ such that the field is also a Markov chain of order r with respect to distance l_1. By the above discussion we have proven the following theorem.

Theorem 3 *Let $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ be a homogeneous random field. If $\{X(\mathbf{t})\}$ is Markov chain of order r with respect to distance l_1, then it can be put in one-to-one correspondence with a Markov (of order r) sequence $\{\mathbf{Y}_k, k \in \mathbf{Z}\}$. The value of $\mathbf{Y}_k$ is the set of values $\{X(\mathbf{t}) : t_1 + t_2 + \ldots + t_n = k\}$, i.e., values of the $\{X(\mathbf{t})\}$ field found on a hyperplane perpendicular to the main diagonal. Hence, $\{X(\mathbf{t})\}$ can be recursively generated along the main diagonal, and questions regarding the stationary distribution of $\{X(\mathbf{t})\}$ can be answered by investigating the stationary distribution of $\{\mathbf{Y}_k, k \in \mathbf{Z}\}$. In particular, if $\{X(\mathbf{t})\}$ takes on a finite set of values, then $\{\mathbf{Y}_k\}$ is essentially real-valued, and it will have a unique stationary distribution under the usual regularity conditions, (Doeblin's condition, or an equivalent weak dependence condition, cf. Doob (1953)).*

It might seem at first that this method of recursive generation of the $\{X(\mathbf{t})\}$ field is not practical because $\mathbf{Y}_k$ is an *infinite* 'strip' of points. However, to generate the $\{X(\mathbf{t})\}$ field on a *finite* grid, only a *finite* number of initial values are required. For example, consider the two dimensional field satisfying (35). To generate it on the positive cube C_N^+, (for any $N \in \mathbf{N}$), only the values of $\{X(i,j) : i + j = 0, |i| \leq N\}$ are required (see Figure 3).

As a final example, consider the unilateral Gaussian autoregression (in the plane) of Whittle (1954). That is, let $\{Z(i,j), (i,j) \in \mathbf{Z}^2\}$ be a random field of i.i.d. normal $N(0,1)$ random variables, and define $X(i,j)$ by the linear autoregression

$$X(i,j) = aX(i-1,j) + bX(i,j-1) + Z(i,j) \tag{37}$$

with $|a| + |b| < 1$.

It can be shown (cf. Woods (1972)) that the $X(i,j)$ field also satisfies the 'interpolating' equation

$$X(i,j) = -\tilde{\rho}_{01}[X(i,j+1) + X(i,j-1)] - \tilde{\rho}_{10}[X(i+1,j) + X(i-1,j)] + U(i,j) \tag{38}$$

where $U(i,j), (i,j) \in \mathbf{Z}^2$, is a field of mean zero, normal random variables, independent of the $\{X(i,j)\}$ field, and satisfying

$$\frac{E[U(0,0)U(i,j)]}{E[U(0,0)]^2} = \begin{cases} 1 & \text{for } (i,j) = (0,0) \\ \tilde{\rho}_{01} & \text{for } (i,j) = (0,\pm 1) \\ \tilde{\rho}_{10} & \text{for } (i,j) = (\pm 1, 0) \\ 0 & \text{otherwise} \end{cases}$$

It is apparent that the $X(i,j)$ field satisfies (35), and thus can be generated recursively using the $\mathbf{Y}_k$ sequence which is defined as before. Existence (and uniqueness) of a stationary distribution for the $X(i,j)$ field is now shown by Whittle's 'solution' of equation (37) in the infinite moving-average unilateral form

$$X(i,j) = \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{(m+k)!}{m!k!} a^m b^k Z(i-m, j-k) \tag{39}$$

Acknowledgements

The author is grateful to Tom Cover, Steve Lalley, and Herman Rubin, for many helpful discussions, and to D. Gatzouras for suggesting the example of a Markov chain in the plane given at the end of Section 2.

References

[1] Abend, K., Harley, T.J., and Kanal, L.N. (1965), Classification of binary random patterns, *IEEE Trans. Inform. Theory*, 11, 538-544.

[2] Anastassiou, D. and Sakrison, D. J. (1982), Some results regarding the entropy rate of random fields, *IEEE Trans. Inform. Theory*, 28, 340-343.

[3] Choi, B.S. and Cover, T.M. (1984), An Information-Theoretic Proof of Burg's Maximum Entropy Spectrum, *Proc. IEEE*, vol. 72, 8, 1094-1095.

[4] Dobrushin, R.L. (1968), The description of a random field by means of conditional probabilities and conditions of its regularity, *Theor. Prob. Appl.*, 13, 197-224.

[5] Doob, J.L. (1953), *Stochastic Processes*, John Wiley, New York.

[6] Georgii, H.-O. (1988), *Gibbs measures and phase transitions,* DeGruyter Studies in Mathematics, vol. 9, Berlin.

[7] Kindermann, R. and Snell, J.L. (1980), *Markov random fields and their applications*, Contemporary Mathematics vol. 1, American Mathematical Society, Providence, Rhode Island.

[8] Politis, D.N. (1991), Non-Parametric Maximum Entropy, Technical Report No. 91-17, Department of Statistics, Purdue University.

[9] Politis, D.N. (1992), Moving Average Processes and Maximum Entropy, *IEEE Trans. Inform. Theory.*

[10] Pinsker, M.S. (1964), *Information and Information Stability of Random Variables and Processes*, Holden-Day.

[11] Spitzer, F. (1972), A variational characterization of finite Markov chains, *Ann. Math. Statist.*, 43, 303-307.

[12] Tjøstheim, D. (1978), Statistical spatial series modelling, *Adv. Applied Prob.*, 10, 130-154.

[13] Tjøstheim, D. (1983), Statistical spatial series modelling II: some further results on unilateral lattice processes, *Adv. Applied Prob.*, 15, 562-584.

[14] Whittle, P. (1954), On stationary processes in the plane, *Biometrika*, 434-449.

[15] Woods, J.W. (1972), Two dimensional discrete Markov random fields, *IEEE Trans. Inform. Theory*, 18, 232-240.

Figure 1. The point $\mathbf{t}$ with its past 'in the direction 2', its general past, and its $L_\infty(X(\mathbf{t}))$ neighborhood.

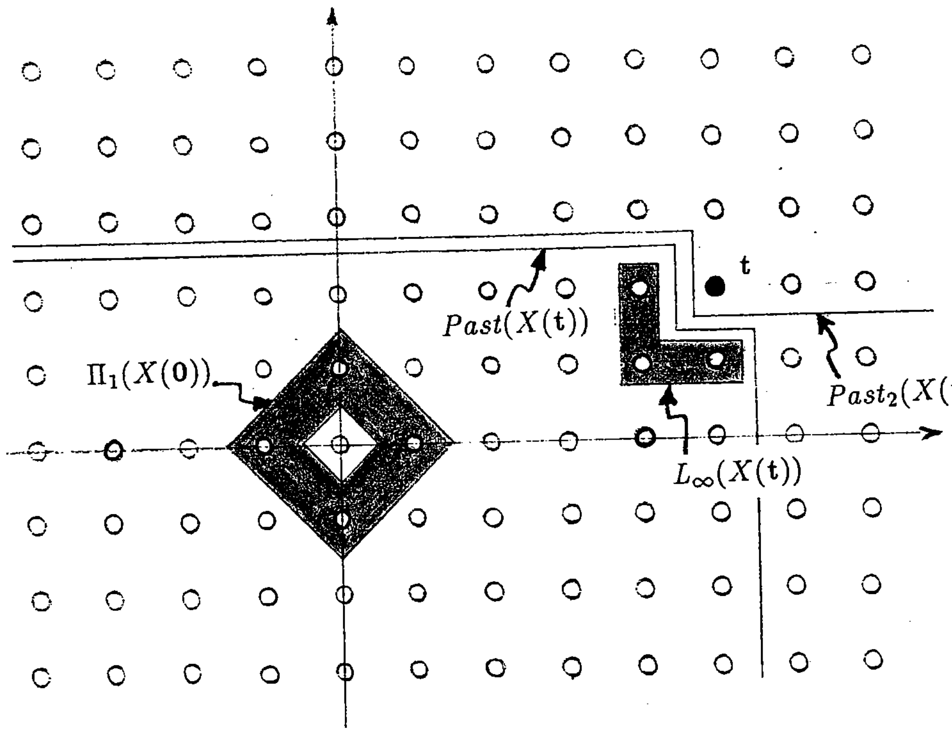

Figure 2. The positive cube C_N^+, and the rN^{n-1}, (with $n = 2$), points that are r-close to the (left) boundary.

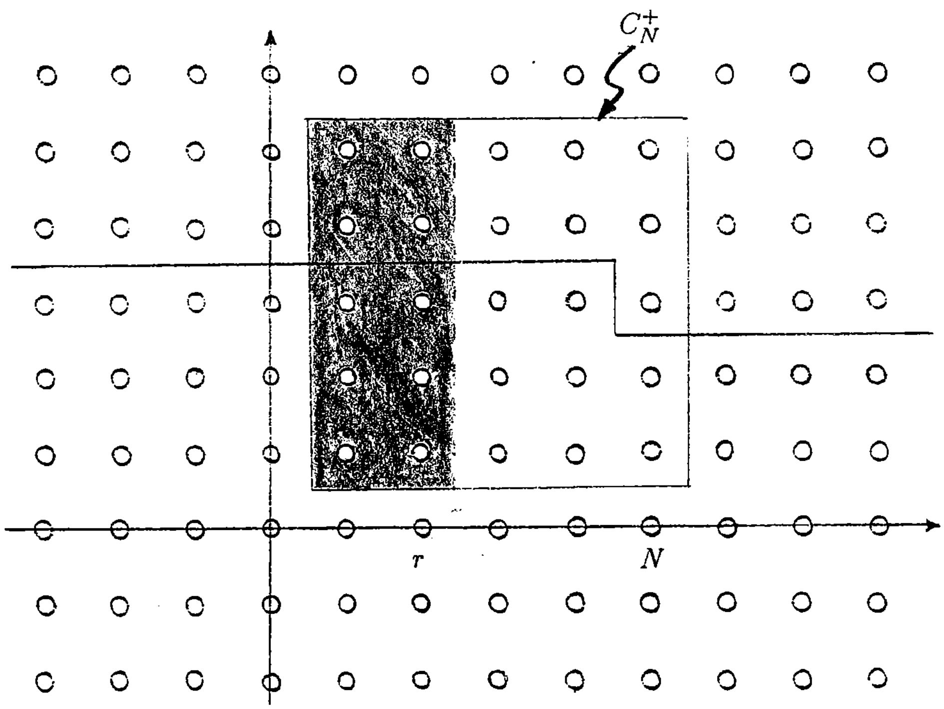

Figure 3. The positive cube C_N^+, and the values of $\{\mathbf{Y}_0\}$ required to generate it.