INPE-5395-PRE/1753

MARKOVIAN FIELDS AS LIMITS OF MARKOV CHAINS

Oscar H. Bustos
Alejandro C. Frery

INPE
São José dos Campos
1992

SECRETARIA DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-5395-PRE/1753

MARKOVIAN FIELDS AS LIMITS OF MARKOV CHAINS

Oscar H. Bustos*
Alejandro C. Frery

Procedings of the International Conference on
use of Computers in Scientific and Technical Research, 2.,
Argentina, 1991, p.85-97

*IMPA

INPE
São José dos Campos
1992

CDU: 621.376.5

Key Words: Digital image modelling; Gibbs sampler; Graph teory; Ising model; Markovian fiels; Pseudo-likehood estimation

Markovian Fields as Limits of Markov Chains

OSCAR H. BUSTOS
IMPA
Estrada Dona Castorina, 110
CEP 22460 RJ Brazil
E-Mail: obuv@lncc.bitnet

ALEJANDRO C. FRERY
INPE — DPI
Avenida dos Astronautas, 1758
CEP 12201 SP Brazil
E-Mail: inpedpi@brfapesp.bitnet

Abstract

Markovian fields are useful stochastic models for the statistical treatment of digitized images, among other applications. In this paper we present some simulation results about the convergence properties of one of the available methods to simulate Markovian fields: the Gibbs sampler. We use the observed values of the mean magnetization, the pseudo-likehood estimates and the observed short and long range correlations to study the problems associated with the stopping time of the algorithm for different values of the parameter.

Key Words: digital image modelling, Gibbs sampler, graph teory, Ising model, Markovian fields, pseudo-likehood estimation.

1 Introduction

The theory, the acquisition and the practical treatment of discretized and digitized images has classically been a discipline of engineering. Many good solutions to problems have been found, and are still being developed, without the use of statistical tools, but statistical ideas are used more and more.

The key element, when working within the statistical framework, is that a model has to be stated before anything is done. Markovian fields are a promising class of models where spatial interaction is allowed. They were popularized by BESAG (1974) and since then they are receiving more and more research attention from different fields.

It is quite important to obtain samples of Markovian fields. Several reconstruction and classification techniques that use simulation have been proposed in the last years and they seem to be, at least, as good as the classical methods used since the advent of image processing. These techniques are related to the idea of simulated annealing, an all-pourpose method of stochastic relaxation to solve large combinatorial problems.

The simulation of Markovian fields poses several interesting problems. Since only iterative techniques seem adequate to obtain samples from this class of distributions, one of the important questions to be answered is the stopping time of the used algorithm.

In this work we illustrate the convergence properties of one of the most popular algorithms, due to Geman and Geman: the Gibbs sampler.

The paper is divided into four main parts. The first contains some useful elements of graph theory, the definitions of Markovian fields and of Gibbs distributions. The second part is devoted to one of the main simulation techniques already available to generate Markovian fields: the Gemans' algorithm. In the third part the theory of pseudo-likehood estimation is presented and illustrated by examples. The fourth part presents some non conclusive results on the determination of the stopping time for the simulation of the finite isotropic free-boundary Ising model; the aim of this paper is to draw the attention on the problems that may arise, rather than formally obtaining definitive results. The relevant listings are provided under request.

2 The theory of Markovian fields

Our main interest is in modelling digital discretized images, so we consider that the pixels of such an image are the pairs (site of a graph, observed value in that site), and we present some unified notation. Markovian fields, that is, vectors of random variables indexed by the sites of a graph, are then introduced and, finally, the relation between Markovian fields and Gibbs distributions is recalled.

2.1 Elements of graph theory

To every site s we associate a random variable whose conditional distribution, given all the other variables is, for Markovian fields, determined by the values taken by pixels in a suitable subset of the whole image: those pixels that are "neighbours" of the pixel s.

Let S be a finite set and $\preceq$ a total order relation in S; let o and q be elements of S. The set $\mathcal{L} \subset S \times S$ is an **edge set** for S if (i) $o \in S \Rightarrow \exists q \in S: (o,q) \in \mathcal{L}$ or $(q,o) \in \mathcal{L}$, and (ii) $\mathcal{L} \subset \{(o,q): o \preceq q\}$.

A system $\mathcal{G} = \{S, \mathcal{L}\}$, where $\mathcal{L}$ is an edge set of S is called **a graph**, $s \in S$ will be called **a site** and $\ell \in \mathcal{L}$ will be called **an edge**. From now on we suppose that a graph $\mathcal{G}$ is given. Given

$o \in S$, $q \in S$ we say that o **is a neighbour of** q if $o \neq q$ and $(o,q) \in \mathcal{L}$ or $(q,o) \in \mathcal{L}$; and for every $q \in S$ we write $\partial_q = \{o \in S : q \text{ is neighbour of } o\}$. Note that $o \in \partial_q \iff q \in \partial_o$; for non-empty subsets $A \subset S$ we write $\partial_A = \{q \in S \setminus A : \partial_q \cap A \neq \emptyset\}$. For every non-empty subset $A \subset S$ we call: the **closure** of A the set $\overline{A} = A \cup \partial_A$ and the **interior** of A the set $\overset{\circ}{A} = \{s \in A : s \notin \partial_t \forall t \in S \setminus A\}$. We say that $A \subset S$ is a **clique** if $a \in A$ and $b \in A$ with $a \neq b$ imply that $a \in \partial_b$. The subset $\lambda \subset S$ is called **a connected subgraph** of $\mathcal{G}$ if for every pair of distinct points p, $q \in \lambda$ there exists a sequence $\ell_0, \ldots, \ell_m$, $m \geq 0$, satisfying (i) $\ell_j \in \lambda$ for all $0 \leq j \leq m$ and (ii) $\ell_j \in \partial_{\ell_{j+1}}$ for all $0 \leq j \leq m-1$.

Example 1: For digital images where $S = \{1,\ldots,M\} \times \{1,\ldots,N\} \subset \mathbf{Z}^2$ there are standard definitions of neighbourhoods. Sites $s \in S$, $t \in S$ are considered first-order (second-order) neighbours if $\|s-t\| \leq 1$ ($\sqrt{2}$ respectively), and we write $\partial^{(1)}$ ($\partial^{(2)}$ respectively). There are many "natural" ways to generalize this and to define hierarchies of Markovian fields in increasing order of spatial dependence ("spatial memory"); in the remainder of this text we use the L^∞ norm:

$$\partial_r^{(k)} = \left\{ (v_1, v_2) \in S \setminus \{(r_1, r_2)\} : \sup\{|v_1 - r_1|, |v_2 - r_2|\} \leq k-1 \right\} \quad \forall k > 1.$$

2.2 Markovian fields

We shall now define the probabilistic elements of a Markovian field. The central idea is the conditional independence between components of a random vector; this, when stated in terms of graphs, will be used to define a Markovian field.

Suppose a graph $\mathcal{G}$ is given. For every $s \in S$ let Ξ_s be a non-empty finite set, and $\Xi = \prod_{s \in S} \Xi_s$ the product space with its natural σ-algebra $\mathcal{P}(\Xi)$, the subsets of Ξ. Let $(\Omega, \mathcal{A}, \mathbb{P}\mathrm{r})$ be a probability space and $X : \Omega \longrightarrow \Xi$ a measurable function; also suppose that $\mathbb{P}\mathrm{r}(X = x) = \mathbb{P}\mathrm{r}(x) > 0$ for every $x \in \Xi$.

For non-empty subsets A, B and C of S; we say that A and B are **conditionally independent** given C (with respect to X) if

$$\mathbb{P}\mathrm{r}(X_A = x_A, X_B = x_B \mid X_C = x_C) = \mathbb{P}\mathrm{r}(X_A = x_A | X_C = x_C) \cdot \mathbb{P}\mathrm{r}(X_B = x_B | X_C = x_C) \forall x_A \in \Xi_A, \quad x_B \in \Xi_B, \quad x_C \in \Xi_C,$$

and we write it $A \perp B | C$ or $X_A \perp X_B | X_C$.

Suppose $3 \leq \#S < \infty$; the following **local Markovian conditions** are equivalent:

1. $s \in S, t \in S, s \notin \partial_t \Rightarrow \{s\} \perp \{t\} \mid \{u\}$ resp. $X, \forall u \in S \setminus \{s,t\}$.
2. $\mathbb{P}\mathrm{r}(X_s = x_s | X_{S\setminus\{s\}} = x_{S\setminus\{s\}}) = \mathbb{P}\mathrm{r}(X_s = x_s | X_{\partial_s} = x_{\partial_s})\ \forall x = [x_u]_{u \in S} \in \Xi, \forall s \in S.$
3. For every $x \in \Xi$, every $s \in S$ and every $y \in \Xi$ with $y_{\overline{s}} = x_{\overline{s}}$ holds $\mathbb{P}\mathrm{r}(X_s = x_s \mid X_{S\setminus\{s\}} = x_{S\setminus\{s\}}) = \mathbb{P}\mathrm{r}(X_s = y_s \mid X_{S\setminus\{s\}} = y_{S\setminus\{s\}})$.

We say that $X = [X_s]_{s \in S}$ is a Markovian field if it satisfies any of the above local Markovian conditions.

For A and B non-empty subsets of S the following **global Markovian conditions** are equivalent:

1. $A \perp B | S \setminus (A \cup B)$ if $A \cap \overline{B} = \emptyset$ $(\overline{A} \cap B = \emptyset)$
2. $\mathbb{P}\mathrm{r}(X_A = x_A | X_{S\setminus A} = x_{S\setminus A}) = \mathbb{P}\mathrm{r}(X_A = x_A | X_{\partial_A} = x_{\partial_A}), \forall x \in \Xi$

3. $\Pr(X_A = x_A | X_{S\setminus A} = x_{S\setminus A})$ does not depend on $x_{S\setminus\overline{A}}$, $\forall x \in \Xi$.

We say that $X = [X_s]_{s\in S}$ is a Markovian field if any of the above global Markovian conditions hold. The non-trivial part of the proof of the equivalence between global and local Markovian conditions may be found in MØLLER (1988).

The joint distribution of a Markovian field is uniquely determined by its local characteristics 2. See BESAG (1974) and, in a more detailed version, MØLLER (1988). This is the characterization that we mainly chose to work with, since it is practical for pourposes of simulation and of estimation.

2.3 Gibbs distributions

The next main ingredient in the theory is the concept of a Gibbs distribution. We shall define it using the concepts of *potential* and of *energy*, both borrowed from the framework of statistical mechanics.

We call a **potential function** (associated to the graph $\mathcal{G}$) every function $V: \mathcal{C}(S) \times \Xi \longrightarrow \mathbb{R}$, where $\mathcal{C}(S)$ is the familiy of non-empty cliques of S, such that $V(A,x) = V(A,y)$ if $x_A = y_A$. We use the standard notation $V(A,x) = V_A(x)$. Given a potential function V we call the **energy** associated to V the function $U_V: \Xi \longrightarrow \mathbb{R}$ defined as $U_V(x) = \sum_{A\in\mathcal{C}(S)} V_A(x)$. Let $X: \Omega \longrightarrow \Xi$ be a measurable function, V a potential function and U_V the energy associated to V. We say that X has a **Gibbs distribution** with potential V if

$$\Pr(X = x) = \frac{1}{Z_V} \exp\Big\{-U_V(x)\Big\},$$

where $Z_V = \sum_{y\in\Xi} \exp\{-U_V(y)\}$ is called the **partition function.**

The addition of the *temperature* parameter is sometimes convenient:

$$\Pr{}_T(X = x) = \frac{1}{Z_{V,T}} \exp\Big\{-\frac{1}{T} U_V(x)\Big\},$$

with $Z_{V,T} = \sum_{y\in\Xi} \exp\{-\frac{1}{T} U_V(y)\}$.

One of the main theorems related to this theory is due to Hammersley and Clifford. It proves that X is a Markovian field if and only if X has a Gibbs distribution. There exist several proofs of this theorem, from the —unpublished— original to a quite simple proof in BESAG (1974); we suggest the reader to see the elegant one in MØLLER (1988), essentially due to SPEED (1978).

Note that if the Gibbsian form is known it is easy to obtain the local characteristics; but knowing the local characteristics little helps to calculate the joint form of the distribution due to the awkward form of the partition function: it involves a summation over all elements of Ξ.

Remark: this theorem (and most of the material exposed here) is valid only if $\#S < \infty$. The interest in this theorem lies in the facts that

1. it is easier to specify a Markovian field in local characteristic form than in Gibbsian form but, once the (suitable) local characteristics are specified, we know the form of the joint distribution. Recall that, outside this context, the conditional distributions of the components of a random vector do not determine uniquely the joint distribution;
2. the Gibbs distributions have received a lot of attention in physics, since they represent the equilibrium properties of large scale interacting systems well (ideal gases, crystals, metal alloys, ferromagnets, etc.);

3. often it is interesting to evaluate the quotients $\mathbb{P}\mathrm{r}(x) \,/\, \mathbb{P}\mathrm{r}(y)$ for $x, y \in \Xi$; in this case the normalizing constants cancel out leaving only a summation of potential functions over cliques;
4. Gibbs measures have the property that they maximize the entropy for constant mean energy;
5. as a theoretical tool it is useful for some proofs to be able to move from one representation to the other.

Let us introduce some terminology that will be useful when discussing Markovian fields. Assume that $S = \mathbf{Z} \times \mathbf{Z}$, and that all the sites have similar neighbourhood systems, i. e. $\partial_s = s + \partial_0$ for all $s \in S$. A Markovian field X is called:

- **stationary** if its local characteristics are location-invariant, i. e.

$$\mathbb{P}\mathrm{r}(X_s = x_s \,|\, X_t = x_t \,\forall t \in \partial_s) = \mathbb{P}\mathrm{r}(X_0 = x_s \mid X_t = x_t \,\forall t \in \partial_0);$$

- **isotropic** if its local characteristics are rotation-invariant i) in 0 and ii) in s, i. e.

$$\text{i)} \qquad \mathbb{P}\mathrm{r}(X_0 = x_0 \,|\, X_t = x_t \,\forall t \in \partial_0) = \mathbb{P}\mathrm{r}(X_0 = x_0 \mid X_t = x_{\varrho_k(t)} \,\forall t \in \partial_0) \quad k = 0,1,2,3$$

$$\text{ii)} \qquad \mathbb{P}\mathrm{r}(X_s = x_s \,|\, X_t = x_t \,\forall t \in \partial_s) = \mathbb{P}\mathrm{r}(X_s = x_s \,|\, X_t = x_{\varrho_{s,k}(t)} \,\forall t \in \partial_s),$$

  where $\varrho_{s,k}(t) = s + \varrho_k(t - s)$ for all $t \in \partial_s$ and ϱ_k is the $k\frac{\pi}{2}$-counterclockwise rotation operator defined over S (in this case we have assumed that $\partial_0 = \varrho_k(\partial_0)$ for all k, that is, the set ∂_0 is ϱ_k-invariant).

- **homogeneous** if it is stationary and isotropic.

3 Simulation of Markovian fields

Developing our statistical intuition through the observation of simulated Markovian fields is not the only stimulus to studying appropriate simulation techniques. Several image-related problems can be suitably treated via simulated annealing (AARTS AND LAARHOVEN, 1989), a procedure that heavily relies on adequate simulation algorithms; see, for example, the works by CARNEVALLI ET AL. (1985), GEMAN AND GEMAN (1984) and KELLY ET AL. (1988).

Let $X : \Omega \longrightarrow \Xi$ be a Markovian field; assume that we are given its local characteristics or its potential functions. Our aim is to obtain a sample of size one taken from the distribution of X by simulation. The full evaluation of the state space $\Xi = \{\xi_1, \ldots, \xi_{\#\Xi}\}$, the corresponding probabilities $\{p_1, \ldots, p_{\#\Xi}\}$ (where $p_i = \mathbb{P}\mathrm{r}(X = \xi_i)$ for all i) and, then, sampling from this distribution, implies a computational burden unfeasible with today's tools. Also, we do not have any *special* point where to start the simulation, as we had when working with a Markov chain.

We will construct a (possibly non-homogeneous) Markov chain that converges to the desired distribution $\mathbb{P}\mathrm{r}(x)$, i. e. $(X(t))_{t \geq 0}$ with state space Ξ is such that $\mathbb{P}\mathrm{r}(X(t) = x) \longrightarrow \mathbb{P}\mathrm{r}(x)$ as $t \to \infty$. The three main algorithms already available to obtain this convergence are the Metropolis, the Gemans' and the Swandsen-Wang. For a survey see RIPLEY AND KIRKLAND (1990).

3.1 Comparison of methods

Prof. Brian Ripley has suggested that, when the goal is to study the behaviour of some characteristic quantity $\mathbb{E}\psi(X)$, X being a Markovian field, through a measure $\widehat{\mathbb{E}}\psi(X) = n^{-1}\sum_{j=1}^{n}\psi(X_j)$ over the n Markovian fields $X_1, \ldots, X_n$, the three simulating techniques must be used for different values of the parameter vector.

There are no conclusive studies in this direction, but the Swendsen-Wang dynamic seems to be much better than the other two when the desired field, an Ising model, has parameter values corresponding to the supercritical region in the infinite-lattice case.

In this paper we study some problems associated to the stopping time of the Gemans' method: the so-called *Gibbs sampler*.

3.2 The Gibbs Sampler algorithm

This a way to obtain outcomes of Markovian fields, and the structure of the algorithm is suitable for parallel processing. The idea is very natural, and detailed proofs of convergence can be found in GEMAN AND GEMAN (1984). It has been said (DUBES AND JAIN, 1989) that the main drawback of this technique is the fact that time wasting calculations must be performed for every visited pixel (in particular, for binary fields a quotient of exponentials has to be evaluated). This is not the case when every transition probability is calculated and stored in a matrix as soon as the parameters are known. The renewal process has only to access these values and randomly choose the new states.

For binary Markovian fields, we have implemented the following simplified version of this algorithm: for notational ease consider an ordering of the set $S = \{1, \ldots, N^2\}$ and let $X(0):\Omega \longrightarrow \Xi$ be a random vector defined on Ω with values in Ξ such that $[X(0)_s]_{s\in S}$ are independent random variables uniformly distributed on Ξ_s for each $s \in S$. Consider for all x and all w in Ξ:

$$q_1(x \mid w) = \begin{cases} 0 & \text{if } x_{S\setminus 1} \neq w_{S\setminus 1} \\ \mathbb{P}\mathrm{r}(X_1 = x_1 \mid X_{\partial_1} = w_{\partial_1}) & \text{if } x_{S\setminus 1} = w_{S\setminus 1}. \end{cases}$$

Now define the random vector $X(1):\Omega \longrightarrow \Xi$ such that $\mathbb{P}\mathrm{r}(X(1) = x \mid X(0) = w) = q_1(x \mid w)$ thus, $X(1)$ is defined from $X(0)$ by changing only the value in the pixel 1. Now we define $X(2)$ from $X(1)$ in similar way modifying only the pixel 2: $\mathbb{P}\mathrm{r}(X(2) = x \mid X(1) = w) = q_2(x \mid w)$ thus, $X(2)$ is defined from $X(1)$ by changing only the value in the pixel 2. These are called *renewals* and the kth *iteration* is $X(N^2k)$.

Proceeding in this manner we define inductively the random sequence $(X(t))_{t\geq 0}$, by modifying only the pixel labeled $(t-1)\mathrm{mod}\ N^2$. It is easy to see that $(X(t))$ is a Markov chain and Geman and Geman proved that $X(t)$ converges in distribution to X as $t \to \infty$, that is, $\mathbb{P}\mathrm{r}(X(t) = x) \longrightarrow \mathbb{P}\mathrm{r}(X = x)$ for all $x \in \Xi$.

Since only a finite number of iterations may be performed in practice, the resulting simulated field is only approximately the desired Markovian field. Among the convergence criteria that could be used, we worked with the ones that depend on the statiblity from one iteration to the next of the following quantities (to be defined ahead):

1. magnetization;
2. short range correlation;
3. long range correlation, and

4. parameter estimates using maximum pseudo-likehood estimators.

Other sensitive quantities are, for example, percentual variation of pixel values and the radius of the (approximate) circumference that encloses a fixed percent of power in the frequency space of the sampled image (c. f. FRERY, 1990).

3.3 Model under study

We implemented the Gemans' algorithm in the C programming language using a generalized shift register pseudorandom number generator (BUSTOS, 1990) to obtain 128 x 128 first order neighbourhood binary Markovian fields: the celebrated Ising model.

This model was introduced in 1925 by the german physicist Ernst Ising. He was trying to explain certain empirically observed facts about ferromagnetic materials using an stochastic modelling.

In this work we consider the model $S = \{1,\ldots,N\} \times \{1,\ldots,N\}$, $N = 128$, $\Xi_s = \{-1,1\}$ for all $s \in S$, $\theta = (\alpha,\beta) \in \Theta = \mathbb{R}^2$. Define the local characteristics as

$$\mathbb{P}\mathrm{r}(X_s = x_s \mid X_{\partial_s} = x_{\partial_s}) = \frac{\exp\left\{\frac{1+x_s}{2}[\alpha\Upsilon_s + \beta\Psi_s]\right\}}{1+\exp\left\{\alpha\Upsilon_s + \beta\Psi_s\right\}},$$

where

Υ_s = sum in the row-neighbouring pixels of s

Ψ_s = sum in the column-neighbouring pixels of s.

3.4 Quantities for the convergence criteria

One of the central features of the Ising model is that it is capable of explaining the physical phenomena associated with *spontaneous magnetization*. For values of α sufficiently high, even if the spins are random to begin with, they will tend to move to a state of lower energy: mostly up or down, forming a magnet. In terms of random variables, the variables begin being locally dependent and there suddenly appear extremely long range dependences. PICKARD (1987), based on classical works of statistical mechanics, says that these values are $\alpha > \alpha_c = \sinh^{-1}(1) = \ln(1+\sqrt{2})$.

In order to understand the concepts of short and long range correlations we introduce some notation. We call **correlation** of order k the quantity $c(k) = M(k)^{-1}\sum_{(s,t)\in D(k)} X_s X_t$, where $D(k) = \{(s,t) \in S^2 : \|s-t\| = k\}$ and $M(k)$ is a certain normalizing constant (see details in the listing of the program). $c(1)$ is the *short* range correlation, and we consider $c(10)$ as *long* range correlation.

The behaviour of $c(10)$ as a function of α can be seen in Figure A (solid line), as proved by PICKARD (1987) for the $\#S = \infty$ case. Nothing is known analitically for the finite case, but we expect a strong (though smooth) resemblance, as can be seen in the same Figure in dotted line.

These are the ideas that we propose to use as stopping rules for the iterative simulation algorithm generating a Markovian field: criteria 1., 2. and 3. indicate to stop the process whenever the observed characteristics of the current iteration are close *enough* to the theoretical values provided by PICKARD (1987) as a formula for the magnetization and graphics for correlations. For the former consider the random variable $M_\alpha = (\#S)^{-1}|\sum_{s\in S} X_s|$, the net absolute magnetization

of the isotropic binary field X (our case with $\alpha = \beta$). It can be seen in HUANG (1963) that its mean value is given by:

$$\lim_{\#S\to\infty} \mathbb{E}(M_\alpha) = \begin{cases} 0 & \text{if } \alpha < \alpha_c \\ \sqrt{\dfrac{\sqrt{\left[1+\exp\{-2\alpha\}\right]\sqrt{1-6\exp\{-2\alpha\}+\exp\{-4\alpha\}}}}{1-\exp\{-2\alpha\}}} & \text{if } \alpha \geq \alpha_c, \end{cases}$$

and this is the goal theoretical value for the magnetization for the field with parameter α.

The criterion 4., of the maximum pseudo-likehood estimation, is a bit different and we postpone its discussion until next section.

4 Parameter Estimation in Markovian fields

The classical estimation procedure, maximum likehood estimation, is not computationally feasible for every Markovian field with today's knowlegde and tools. The problem is, again, that the partition function is a function of the unknown parameters; since it is not computed —with a few exceptions—, it is impossible to obtain and to solve the classical maximum likehood equations. The reader interested in the problems posed by the computation of the partition function is referred to KINDERMAN AND SNELL (1980).

The idea of pseudo-likehood is that the observed process X_S is just a subset of a bigger process $X_{\mathcal{S}}$, and that estimation must be carried out in a proper window $W \subset S \subset \mathcal{S}$ such that $X_W \perp X_{\mathcal{S}\setminus S} \mid X_{S\setminus W}$. We used the biggest possible window, which is determined by the neighbourhood system.

The pseudo-likehood estimator is defined by

$$\mathrm{PL}_W(x,\widehat{\theta}) = \max_{\theta\in\Theta} \mathrm{PL}_W(x,\theta), \quad \text{where } \mathrm{PL}_W(x,\theta) = \prod_{w\in W} \mathbb{P}\mathrm{r}_\theta(X_w = x_w \mid X_{\partial_w} = x_{\partial_w}),$$

and where, in the practice, $\mathrm{pl}_W(x,\theta) = \ln \mathrm{PL}_W(x,\theta)$ is evaluated.

There is a theorem (c. f. JENSEN AND MØLLER, 1989) that proves the consistency of this class of estimators.

Consider the Markovian field previously defined; here the set S is $\{1,\ldots,N\}\times\{1,\ldots,N\}$ and the local characteristics are defined as

$$\mathbb{P}\mathrm{r}_{\alpha,\beta}(X_{i,j} = x_{i,j} \mid X_{S\setminus(i,j)} = x_{S\setminus(i,j)}) = \frac{\exp\left\{[\alpha(x_{i,j-1}+x_{i,j+1})+\beta(x_{i-1,j}+x_{i+1,j})]\dfrac{1+x_{i,j}}{2}\right\}}{1+\exp\left\{\alpha(x_{i,j-1}+x_{i,j+1})+\beta(x_{i-1,j}+x_{i+1,j})\right\}},$$

for every site $(i,j) \in \{2,\ldots,N-1\}\times\{2,\ldots,N-1\}$, with the proper redefinitions for sides and for corners. This first-order neighbourhood induces the maximum window $W_{\max} = \{(i,j) \in \{2,\ldots,N-1\}\times\{2,\ldots,N-1\}\}$ and some calculations give as maximum pseudo-likehood estimators for (α,β) those values $(\widehat{\alpha},\widehat{\beta}) \in \mathbb{R}^2$ that satisfy

$$\begin{aligned} &F_2^1 + C_2^1 - (F_2^{-1} + C_2^{-1}) + 2\Big[P_{-2}^2\,\Phi(2(\widehat{\alpha}-\widehat{\beta})) + P_{-2}^{-2}\,\Phi(-2(\widehat{\alpha}+\widehat{\beta})) - \\ &P_2^2\,\Phi(2(\widehat{\alpha}+\widehat{\beta})) - P_2^{-2}\,\Phi(2(-\widehat{\alpha}+\widehat{\beta}))\Big] + P_0^2(\,\Phi(2\widehat{\alpha}) + \Phi(2\widehat{\beta})) + \\ &P_0^{-2}(\,\Phi(-2\widehat{\alpha}) + \Phi(-2\widehat{\beta})) = 0, \end{aligned}$$

where

$$
\begin{aligned}
F_\ell^k &= \#\{x_{i,j},(i,j)\in W: x_{i,j}=k,\ x_{i,j-1}+x_{i,j+1}=\ell\},\\
C_\ell^k &= \#\{x_{i,j},(i,j)\in W: x_{i,j}=k,\ x_{i-1,j}+x_{i+1,j}=\ell\},\\
P_\ell^k &= \#\{x_{i,j},(i,j)\in W: x_{i,j-1}+x_{i,j+1}=k,\ x_{i-1,j}+x_{i+1,j}=\ell\} \text{ and}\\
\Phi(\nu) &= \frac{\exp\{\nu\}}{1+\exp\{\nu\}}.
\end{aligned}
$$

This equation can be solved using numerical tools. If the isotropic model is considered (our case where $\alpha = \beta$) the maximum pseudo-likehood estimator is the value $\widehat{\alpha} \in \mathbb{R}$ that satisfies the objective equation

$$
\begin{aligned}
&2(C_4^1 - C_{-4}^1) + (C_2^1 - C_{-2}^1) + 2(C_{-4}^{-1} + C_4^1)\frac{\exp\{-4\widehat{\alpha}\}}{1+\exp\{-4\widehat{\alpha}\}} +\\
&(C_{-2}^{-1} + C_2^1)\frac{\exp\{-2\widehat{\alpha}\}}{1+\exp\{-2\widehat{\alpha}\}} - (C_2^{-1} + C_2^1)\frac{\exp\{2\widehat{\alpha}\}}{1+\exp\{2\widehat{\alpha}\}} -\\
&2(C_4^{-1} + C_4^1)\frac{\exp\{4\widehat{\alpha}\}}{1+\exp\{4\widehat{\alpha}\}} = 0,
\end{aligned}
$$

where we have written $C_k^\ell = \#\{x_w: w\in W \mid x_w = \ell,\ \Sigma_w = k\}$, and $\Sigma_w = \sum_{t\in\partial_w} x_t$.

The quality of this estimation technique is good enough to ensure their success when used in a variety of contexts, mainly in image-segmentation algorithms. COHEN AND COOPER (1987) present some results in this direction.

4.1 Criterion 4. as stopping rule

Our aim is finding an integer $k^\star$ such that the simulated field in stage $k^\star$, that is $X(k^\star)$, is *close*, in some "senses", to the desired Markovian field X. Three of these senses are criteria 1., 2. and 3.

We know that X, Ising model, is characterized by the value of the parameter α and that we have a good rule of estimation of α, just the zero of the objective equation. In order to simulate a Markovian field we have to fix the desired value for the parameter. It appears sensible to define, as a criterion for the stopping rule of the simulation, the $k^\star$ such that $\widehat{\alpha}(k^\star)$ is close to the imposed α, where $\widehat{\alpha}(k)$ is the solution of the objective equation obtained using $X(k)$ in place of X.

These four criteria jointly define the stopping time. In this work we checked that for $\alpha < \alpha_c$ this stopping time is about 100 iterations, and that for $\alpha > \alpha_c$ it is about 50000 iterations.

5 The results

We worked with the values of $\alpha = \beta$ in the ranges $0,(0.1),0.8$ and $0.85,(0.05),1.2$; note that these values span both the sub and the supercritical regions, i. e., values of the parameter $\alpha < \alpha_c$ and $\alpha > \alpha_c$.

We noted that simulations with different values of the parameter, but belonging to the same sub or supercritical region, do not exhibit major differences in the convergence behaviour. So, in this work, we present the results for a simulation with $\alpha = 0.5$ and other with $\alpha = 1$.

The main results are presented in Figures 1 to 8. For $\alpha = 0.5$ ($\alpha = 1$ respectively) see Figures 1 to 4 (5 to 8 respectively), corresponding, in order, to: mean magnetization vs. iteration, $\widehat{\alpha}$ vs. iteration, $c(1)$ vs. iteration and $c(10)$ vs. iteration.

5.1 Comments on the graphics

From the theoretical results presented in PICKARD (1987) we may expect that the desired convergence has been achieved for k^* if for all $k \geq k^*$ the values of the mean magnetization, $\hat{\alpha}$, $c(1)$ and $c(10)$ oscillate around their correct mean theoretical values given above. That is, for $\alpha = 0.5$ they are 0, 0.5, 0.28 approximately and 0 respectively; for $\alpha = 1$ they are 0.91, 1, 0.88 aproximately and 0.85 aproximately, repectively.

Looking at Figures 1 to 8 we can conclude that k^* could be chosen as a value smaller than 100, for the subcritical case and smaller than 50000 for the supercritical case.

Detailed computational information will be sent by the authors under request.

6 References

—AARTS, E. H. L. AND LAARHOVEN, P. J. M. (1989): Simulated Annealing: An Introduction. *Statistica Neerlandica* **43**; 31–52.

—BESAG, J. E. (1974): Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion). *J. Roy. Statist. Soc.* **B 2**; 192–236.

—BUSTOS, O. H. (1990): A Turbo C implementation of a GFSR Uniform [0, 1] pseudorandom number algorithm. XIII CNMAC; Águas de Lindóia.

—CARNEVALLI, P.,sCOLETTI, L. AND PATARNELLO, S. (1985): Image processing by simulated annealing. *IBM Jr. Res. Develop.* **6**; 569–579.

—COHEN, F. S. AND COOPER, D. B. (1987): Simple Parallel Hierarchical and Relaxation Algorithms for Segmenting Noncausal Markovian Random Fields. *I. E. E. E. Trans.* **PAMI-9**; 195–219.

—DUBES, R. C. AND JAIN A. K. (1989): Random Field Models in Image Analysis. *J. Appl. Statist.* **16–2**; 131–164.

—FRERY, A. C. (1990): *Estatística na Análise e na Síntese de Imagens.* Master's thesis in Applied Mathematics (Statistics), Tese **D-34/90**, Informes de Matemática, IMPA (Instituto de Matemática Pura e Aplicada). Rio de Janeiro.

—GEMAN, S. AND GEMAN, D. (1984): Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Degradated Images. *I. E. E. E. Trans.* **PAMI-6**; 721–741.

—HUANG, K. (1963): *Statistical Mechanics.* John Wiley & Sons, Inc.

—JENSEN, J. L. AND MØLLER, J. (1989): Pseudolikehood for Exponential Family Models of Spatial Processes. Research Report; Department of Theoretical Statistics, Århus University.

—KELLY, P. A., DERIN, H. AND HARTT, K. D. (1988): Adaptive Segmentation of Speckled Images Using a Hierarchical Random Field Model. *I. E. E. E. Trans.* **ASSP 36**; 1628–1641.

—KINDERMAN, R. AND SNELL, J. L. (1980): *Markov Random Fields and their Applications*, Contemporary Mathematics **1** ; American Mathematical Society.

—MØLLER, J. (1988): *Lectures on Markov Random Fields.* Department of Theoretical Statistics, Århus University.

—PICKARD, D. K. (1987): Inference for Discrete Markov Fields: The Simplest Nontrivial Case. *Jr. Amer. Statist. Assoc.* **82**; 90–96.

—RIPLEY, B. D. AND KIRKLAND, M. D. (1990): Iterative Simulation Methods. In *Random Numbers and Simulation.* Comp. Appl. Math., Lambrecht.

—SPEED, T. P. (1978): Relations between models for Spatial data, Contingency tables and Markov fields on graphs. *Supplement Adv. Appl. Probab.* **10**; 111–122.

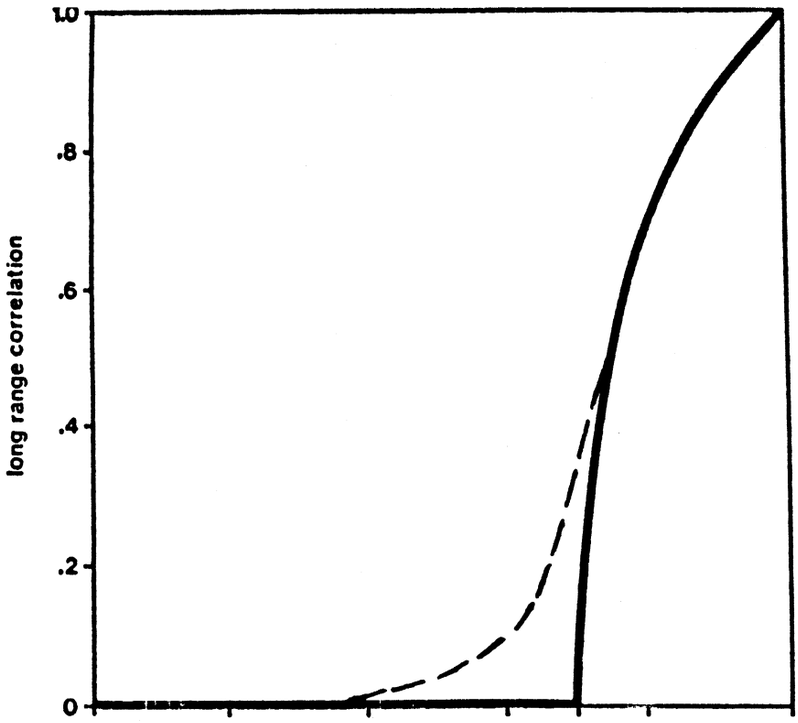

Figure A

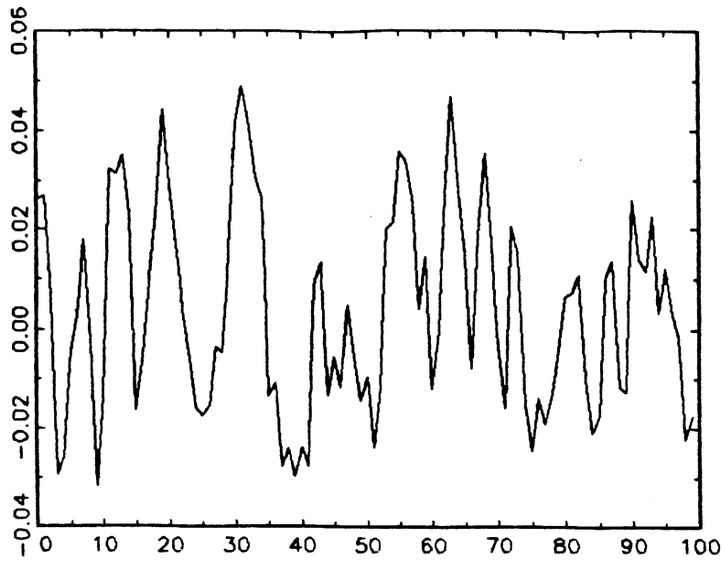

Figure 1

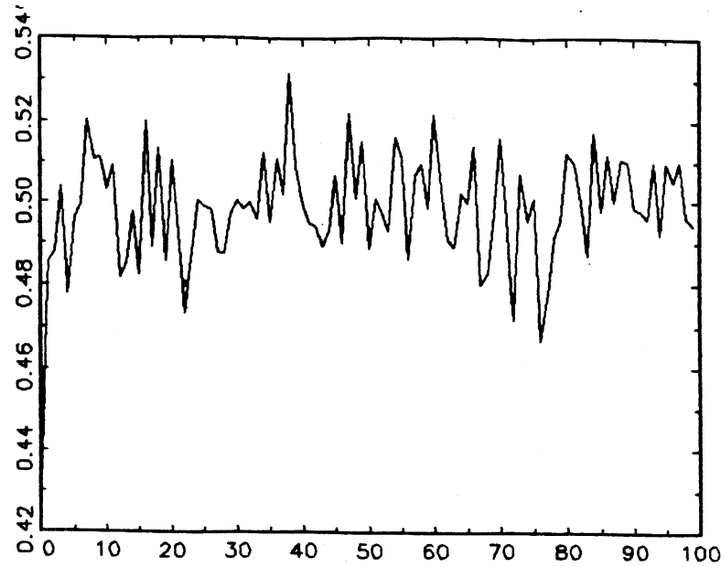

Figure 2

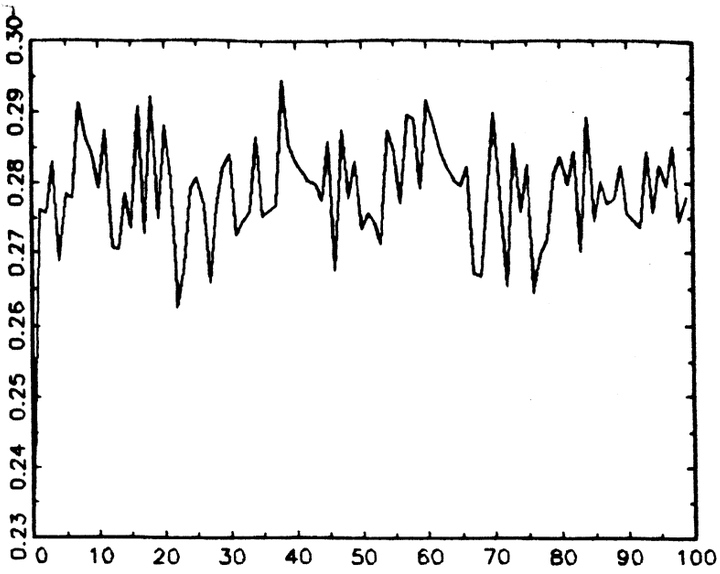

Figure 3

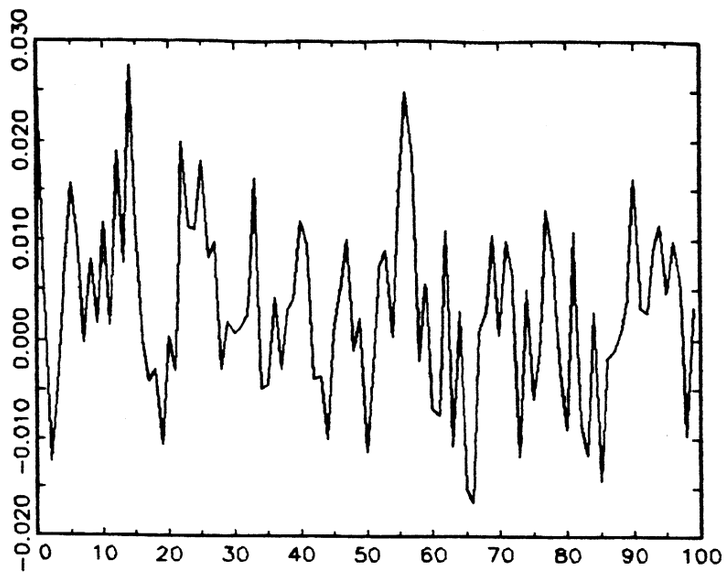

Figure 4

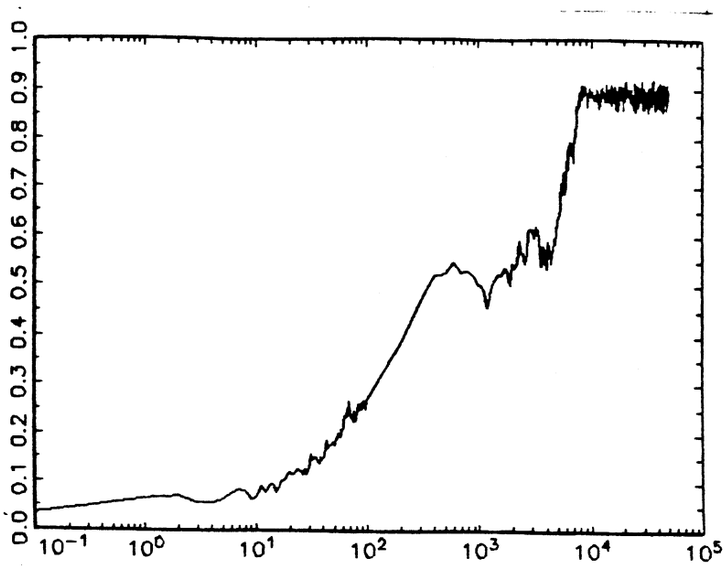

Figure 5

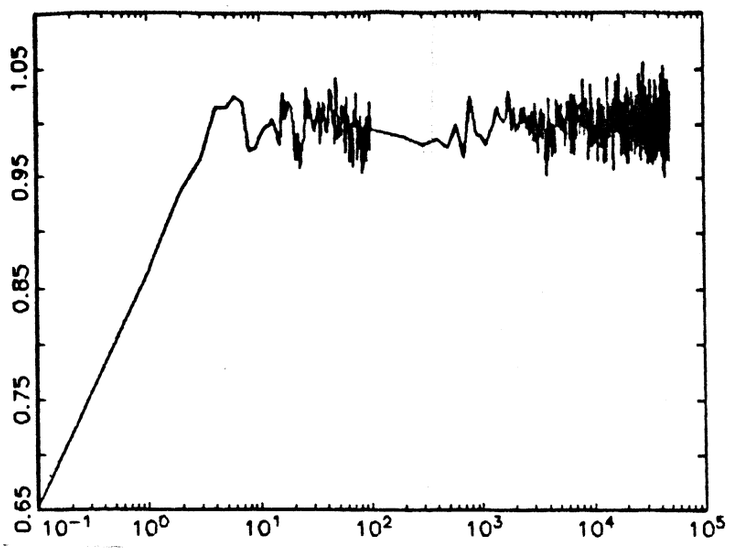

Figure 6

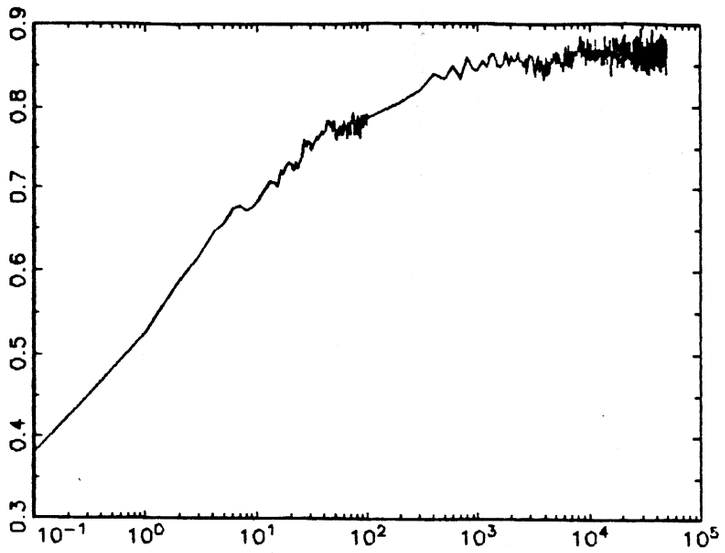

Figure 7

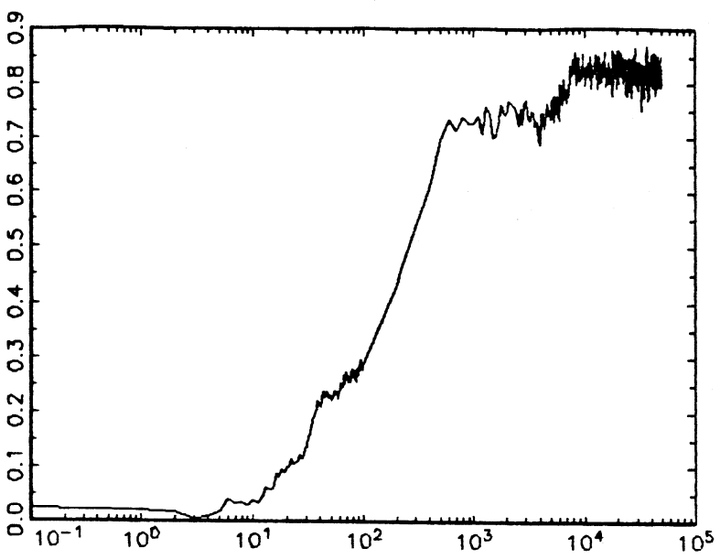

Figure 8