Consistent Estimation of Stationary Processes and Stationary Random Fields

Stuart Geman and Athanasios Kehagias
Division of Applied Mathematics
Brown University
Providence, RI 02912 USA

Hans Künsch
Seminar fur Statistik
ETH Zentrum
CH-8092 Zurich
SWITZERLAND

November, 1993

[1]Supported by Army Research Office contract DAAL03-92-G-0115 to the Center for Intelligent Control Systems, National Science Foundation grant DMS-8813699, and Office of Naval Research contract N00014-91-J-1021.

1 Introduction

The purpose of this report is to provide the proofs of the consistency results in the paper Künsch, Geman and Kehagias ([4]). We repeat here the definitions and statement of the Theorems so that this report is self-contained. For motivation, background and examples we refer however to [3] and [4].

2 The one-dimensional case

We imagine observing a stationary process Z with state space $E = \{0, 1, ... M-1\}$, for some (finite) $M > 1$. Let μ_o be the (unknown) distribution, or law, of Z. Following the notation of §2 ([4]), the process Z is to be approximated by a hidden process $Y = \bar{f}(X)$, where X is nearest-neighbor with state space $E' = \{0, 1, ... N\}$. Henceforth, the hiding function f (and consequently $\bar{f}$ as well) is fixed: $f(x) = x \bmod M$. In the one-dimensional problem ($S = \mathcal{Z}$), and X is first-order Markov. Let

$$\begin{aligned} \mathcal{M}_N &= \{m = \{m_{ij}\}_{i,j=0}^{N} : \ m \textit{ trans. prob. matrix}, \\ &\quad \textit{and } m_{ij} \geq e^{-N} \ \forall\, 0 \leq i, j \leq N\}. \end{aligned}$$

N will serve as a "regularization" or "smoothing" parameter, and will eventually be tied to the number n of observations, $Z_0 = z_0, \ Z_1 = z_1, ... Z_n = z_n$, through an increasing function. For any $m \in \mathcal{M}_N$, denote by μ_m the distribution of the hidden Markov process $Y = \{Y_t\}, \ Y_t = f(X_t)$, where $\{X_t\}$ is the unique stationary Markov process with transition matrix m. The results of §2 ([4]) suggest that μ_o can be approximated by a distribution μ_m, for suitable m and large enough N. Having observed $Z_0 = z_0,\ Z_1 = z_1, ... Z_n = z_n$, we denote by $ML_{N,n}$ the set of maximum likelihood matrices from within $\mathcal{M}_N$:

$$ML_{N,n} = ML_{N,n}(z) = \{m \in \mathcal{M}_N : \mu_m(z_0, z_1, \dots z_n) = \sup_{q \in \mathcal{M}_N} \mu_q(z_0, z_1, \dots z_n)\}.$$

(In general, $ML_{N,n}$ has more than one element. In any case, it is never empty: $\mathcal{M}_N$ is compact and μ_q is continuous in q.) Under an additional condition on Z, there exists a sequence N_n such that the set of HMM's associated with $ML_{N_n,n}$ is consistent for μ_o:

Theorem 2.1 *Let $\{Z_t\}_{t=-\infty}^{\infty}$ be a stationary ergodic process with finite state space, $Z_t \in \{0, 1, \dots M-1\}$, $M < \infty$, and distribution function μ_o. If $\exists\ \delta > 0\ \ni\ \mu_o(z_0 | z_1, \dots z_{-t}) \geq \delta\ \forall\ t,\ (z_0, \dots z_{-t}) \in \{0, 1, \dots M-1\}^{t+1}$, then for all $N_n \uparrow \infty$ sufficiently slowly*

$$\sup_{m \in ML_{N_n,n}} \int \log \frac{\mu_o(z_0 | z_{-1}, z_{-2}, \dots)}{\mu_m(z_0 | z_{-1}, z_{-2}, \dots)} d\mu_o(z) \rightarrow 0\ a.s.\ (\mu_o)$$

Remarks.

1. More precisely, there exists a sequence $N_n \uparrow \infty$ such that the assertion holds for all sequences $N'_n \uparrow \infty$ satisfying $N'_n \leq N_n\ \forall n$.

2. Unfortunately, $N_n = N_n(\mu_o)$; roughly speaking, $\{Z_t\}$ can yield information arbitrarily slowly.

3. There is nothing special about the regularization $m_{ij} \geq e^{-N}$. If instead, $m_{ij} \geq g(N)$, where $g(N) \downarrow 0$, then there will be a relationship between

$g(N)$ and N_n such that the *faster* $g(N) \downarrow 0$ the *slower* $N_n \uparrow \infty$, in order to insure consistency.

2.1 Two basic lemmas

The proof is based upon two lemmas. The first is a kind of uniform law of large numbers for the probabilities μ_m, $m \in \mathcal{M}_n$, reminiscent of the Shannon-McMillan-Breiman Theorem (cf. Billingsley [1]):

Lemma 2.1.1

$$\lim_{n\to\infty} \sup_{m\in\mathcal{M}_{N_n}} \left|\frac{1}{n}\log\mu_m(z_0, z_1, \ldots z_n) - \int \log\mu_m(z_0|z_{-1}, z_{-2}, \ldots)d\mu_o(z)\right| = 0 \quad a.s.\ (\mu_o)$$

for all $N_n \uparrow \infty$ *sufficiently slowly.*

The second lemma insures that there is *some* sequence $m_N \in \mathcal{M}_N$ such that μ_{m_N} approaches μ_o:

Lemma 2.1.2 *There exists a sequence of matrices* $m_N \in \mathcal{M}_N$ *such that*

$$\lim_{N\to\infty} \int \log\mu_{m_N}(z_0|z_{-1}, z_{-2}, \ldots)d\mu_o(z) = \int \log\mu_o(z_0|z_{-1}, z_{-2}, \ldots)d\mu_o(z).$$

(The proof of lemma 2.1.2 will be by construction.)

Now assume that the lemmas are true. By Jensen's inequality,

$$\int \log\ \mu_m(z_0|z_{-1}, \ldots)d\mu_o(z) \le \int \log\ \mu_o(z_0|z_{-1}, \ldots)d\mu_o(z)$$

for all N and $m \in M_N$, so it is enough to show that

$$\liminf_{n\to\infty} \inf_{m\in ML_{N_n,n}} \int \log\ \mu_m(z_0|z_{-1}, \ldots)d\mu_o(z) \ge \int\ \mu_o(z_0|z_{-1}, \ldots)d\mu_o(z)\ a.s.$$

By application of the lemmas:

$$\liminf_{n\to\infty} \inf_{m\in ML_{N_n,n}} \int \log\ \mu_m(z_0|z_{-1},\ldots)d\mu_o(z)$$

$$\begin{aligned}
&= \liminf_{n\to\infty} \inf_{m\in ML_{N_n,n}} \{(\int \log\ \mu_m(z_0|z_{-1},\ldots)d\mu_o(z) - \frac{1}{n}\log\ \mu_m(z_0,z_1,\ldots z_n)) \\
&\quad + \frac{1}{n}\log\ \mu_m(z_0,z_1,\ldots z_n)\} \\
&\geq \liminf_{n\to\infty} \inf_{m\in ML_{N_n,n}} \{\frac{1}{n}\log\ \mu_m(z_0,z_1,\ldots z_n) - |\frac{1}{n}\log\ \mu_m(z_0,z_1,\ldots z_n) \\
&\quad - \int \log\ \mu_m(z_0|z_{-1},\ldots)d\mu_o(z)|\} \\
&= \liminf_{n\to\infty} \inf_{m\in ML_{N_n,n}} \frac{1}{n}\log\ \mu_m(z_0,z_1,\ldots z_n) \quad (a.s.,\ by\ lemma\ 2.1.1) \\
&\geq \liminf_{n\to\infty} \frac{1}{n}\log\ \mu_{m_{N_n}}(z_0,z_1,\ldots z_n) \\
&= \liminf_{n\to\infty} \int \log\ \mu_{m_{N_n}}(z_0|z_{-1},\ldots)d\mu_o(z) \quad (again,\ a.s.,\ by\ lemma\ 2.1.1) \\
&= \int \log\ \mu_o(z_0|z_{-1},\ldots)d\mu_o(z) \quad (by\ lemma\ 2.1.2)
\end{aligned}$$

2.2 Proof of Lemma 2.1.1

For any $m \in \cup_{N=1}^{\infty}\mathcal{M}_N$, let μ_m be the distribution of the hidden Markov model associated with the process $Y_t = f(X_t)$, where X_t is the stationary Markov process with transition probability matrix m. For any such m, and any $y = \{y_t\}_{t=-\infty}^{\infty}$, let

$$g_n(y,m) = \frac{1}{n}\log\ \mu_m(y_0,y_1,\ldots y_n) - \int \log\ \mu_m(z_0|z_{-1},z_{-2},\ldots)d\mu_o(z).$$

With this notation, Lemma 2.1.1 can be written

$$\lim_{n\to\infty} \sup_{m\in\mathcal{M}_{N_n}} |g_n(y,m)| = 0\ a.s. \qquad (1)$$

for all $N_n \uparrow \infty$ sufficiently slowly.

Given N, and given m, $m' \in \mathcal{M}_N$, define

$$\|m - m'\| = \sup_{0 \le i,j \le N} |m_{ij} - m'_{ij}|.$$

The proof of Lemma 2.1.1 is based on:

Lemma 2.2.1 *(i) For any N and every $\epsilon > 0$, there exists $\delta = \delta(N, \epsilon) > 0$, such that*

$$\sup_{0 \le k \le \infty} \sup_{y} \sup_{\substack{m,m' \in \mathcal{M}_N \\ \|m-m'\| < \delta}} |\mu_m(y_0|y_1, \dots y_{-k}) - \mu_{m'}(y_0|y_{-1}, \dots y_{-k})| < \epsilon.$$

(ii) For every N,

$$\inf_{0 \le k \le \infty} \inf_{y} \inf_{m \in \mathcal{M}_N} \mu_m(y_0|y_{-1}, \dots y_{-k}) > 0.$$

(Where we interpret $\mu_m(y_0|y_{-1}, \dots y_{-k})$ as $\mu_m(y_0)$ when $k = 0$ and as $\mu_m(y_0|y_{-1}, y_{-2}, \dots)$ when $k = \infty$.)

Let us postpone the proof of Lemma 2.2.1 for the time being. To prove Lemma 2.1.1, we make the following two observations:

01. For every N, $m \in \mathcal{M}_N$, $g_n(y, m) \to 0$ a.s. as $n \to \infty$.

02. For every N, $\epsilon > 0$, $\exists \delta = \delta(N, \epsilon) > 0 \ni$

$$\sup_{n} \sup_{y} \sup_{\substack{m,m' \in \mathcal{M}_N \\ \|m-m'\| < \delta}} |g_n(y, m) - g_n(y, m')| < \epsilon$$

The first observation can be established by following, essentially line for line, the proof of the Shannon-McMillan-Breiman Theorem, as presented for example in Billingsley [1], pp. 129–132. The result **01** is actually easier, by virtue of the uniform positivity asserted in Lemma 2.2.1 (ii); we will forego the details. As for **02**, this is a direct consequence of Lemma 2.2.1, as can be seen by rewriting

$$\frac{1}{n}\log\ \mu_m(y_0, y_1, \dots y_n)$$

as

$$\frac{1}{n}\log\ \mu_m(y_0) + \frac{1}{n}\sum_{k=1}^{n}\log\ \mu_m(y_k|y_{k-1}, \dots y_0)$$

and using the stationarity of the process associated with μ_m.

Lemma 2.1.1 (i.e. equation 1) is now proven by first establishing the more modest result:

$$\lim_{n\to\infty}\sup_{m\in\mathcal{M}_N}|g_n(y,m)| = 0\ a.s. \tag{2}$$

for every *fixed* $N = 1, 2, \dots$. To establish (2), fix N and $\epsilon > 0$ and choose $\delta = \delta(N, \epsilon)$ as in **02**. For any $m \in \mathcal{M}_N$ let

$$B(m,\delta) = \{m' \in \mathcal{M}_N :\ \|m - m'\| < \delta\}.$$

$\mathcal{M}_N$ is clearly compact, so we can choose $m_1, m_2, \dots m_r \in \mathcal{M}_N$ such that

$$\mathcal{M}_N = \bigcup_{i=1}^{r} B(m_r, \delta)$$

($r = r(N,\delta)$). By **01**, $\sup\limits_{1\le i\le r}|g_n(y,m_i)| \to 0$ a.s.

Hence

$$\overline{\lim_{n\to\infty}}\sup_{m\in\mathcal{M}_N}|g_n(y,m)| = \overline{\lim_{n\to\infty}}\sup_{1\le i\le r}\sup_{m\in B(m_i,\delta)}|g_n(y,m)|$$

$$\le \overline{\lim_{n\to\infty}}\sup_{1\le i\le r}\sup_{m\in B(m_i,\delta)}|g_n(y,m) - g_n(y,m_i)| + \overline{\lim_{n\to\infty}}\sup_{1\le i\le r}|g_n(y,m_i)|$$

$$\le \overline{\lim_{n\to\infty}}\sup_{\substack{m,m'\in\mathcal{M}_N \\ \|m-m'\|<\delta}}|g_n(y,m) - g_n(y,m')| < \epsilon$$

Since ϵ is arbitrary, (2) is established.

To get from (2) to (1) we use a Borel-Cantelli argument: Let

$$f_n^N(y) = \sup_{m \in \mathcal{M}_N} |g_n(y,m)|$$

Then $f_n^N(y) \to 0$ a.s. for every $N = 1, 2, \ldots$. We will construct $N_n \uparrow \infty$ such that $f_n^{N'_n} \to 0$ a.s. for any sequence $N'_n \uparrow \infty$ with $N'_n \leq N_n$ for all n, thereby showing that $f_n^{N_n} \to 0$ a.s. "for all $N_n \uparrow \infty$ sufficiently slowly." First, choose a sequence n_N, strictly increasing in N, such that

$$\mu_o\left(\sup_{n > n_N} f_n^N > \frac{1}{N} \right) < \frac{1}{N^2} \tag{3}$$

for each $N = 1, 2, \ldots$. For $n \leq n_2$, set $N_n = 1$. For $n_k < n \leq n_{k+1}$, $k \geq 2$, set $N_n = k$. Then $N_n \uparrow \infty$. Let $N'_n \uparrow \infty$ be any sequence such that $N'_n \leq N_n$ for all n. Fix $\epsilon > 0$. For each k, $N'_n = k$ for at most finitely many n. Hence:

$$\begin{aligned}
&\mu_o(f_n^{N'_n} > \epsilon \text{ infinitely often n} \in \{1, 2, \ldots\}) \\
&= \mu_o\left(\sup_{n \ni N'_n = k} f_n^k > \epsilon \text{ infinitely often k} \in \{\mathrm{N}'_\mathrm{i}\}_{\mathrm{i}=1}^{\infty} \right) \\
&\leq \mu_o(\sup_{n \ni N'_n \geq k} f_n^k > \epsilon \text{ infinitely often k} \in \{1, 2, \ldots\}) \\
&\leq \mu_o(\sup_{n \ni N_n \geq k} f_n^k > \epsilon \text{ infinitelyoften k} \in \{1, 2, \ldots\}) \\
&= \mu_o(\sup_{n > n_k} f_n^k > \epsilon \text{ infinitely often k} \in \{1, 2, \ldots\}).
\end{aligned}$$

By (3), and the Borel-Cantelli lemma,

$$\mu_o(\sup_{n > n_k} f_n^k > \frac{1}{k} \text{ infinitely often k} \in \{1, 2, \ldots\}) = 0,$$

from which it follows that

$$\mu_o(\sup_{n>n_k} f_n^k > \epsilon \text{ infinitely often k} \in \{1,2,\ldots\}) = 0$$

for all $\epsilon > 0$, and, hence, that $f_n^{N_n} \rightarrow 0$ a.s. for all $N_n \uparrow \infty$ sufficiently slowly.

To complete the proof of Lemma 2.1.1, we now need to prove Lemma 2.2.1.

2.3 Proof of Lemma 2.1.3

Both parts are based on the observation that, given m, and given $y_{-1}, y_{-2}, \ldots y_{-k}$ (k finite or infinite), $\{x_t\}$ is a (inhomogeneous) first order Markov process, on the (inhomogeneous) state space defined by $f(x_t) = y_t \; \forall t \in \{-1, -2, \ldots -k\}$. (This "compatibility condition," i.e. $f(x_t) = y_t$, will often be implicitly assumed in the following discussion.)

Part (ii) is immediate, since

$$\mu_m(y_0|y_{-1}, \ldots y_{-k}) = \mu_m(x_0 \in f^{-1}(y_0)|y_{-1}, \ldots y_{-k})$$

$$= \sum_{x_{-1}} \mu_m(x_0 \in f^{-1}(y_0)|x_{-1})\mu_m(x_{-1}|y_{-1}, \ldots y_{-k})$$

$$\geq e^{-N} \sum_{x_{-1}} \mu_m(x_{-1}|y_{-1}, \ldots y_{-k}) \geq e^{-N}.$$

As for part (i), the first step is to establish that this conditional Markov process possesses a minimum transition probability that depends on N, but is independent of m, k, and y.

Note first that if $t \geq -1$ then $\mu_m(x_{t+1} = b|x_t = a, y_{-1}, y_{-2}, \ldots y_{-k}) = \mu_m(x_{t+1} = b|x_t = a) = m_{ab} \geq e^{-N}$. If, on the other hand, $t < -1$, then

$$\mu_m(x_{t+1} = b|x_t = a, y_{-1}, y_{-2}, \ldots y_{-k})$$

$$= \frac{\sum\limits_{\substack{x_{t+2}, x_{t+3}, \dots x_{-1} \\ \text{compatible}}} \mu_m(x_t = a | y_{-1}, \dots y_{-k}) m_{ab} m_{bx_{t+2}} m_{x_{t+2}} m_{x_{t+3}} \dots m_{x_{-2}x_{-1}}}{\sum\limits_{\substack{x_{t+1}, x_{t+2}, \dots x_{-1} \\ \text{compatible}}} \mu_m(x_t = a | y_{-1}, \dots y_{-k}) m_{ax_{t+1}} m_{x_{t+1}x_{t+2}} m_{x_{t+2}x_{t+3}} \dots m_{x_{-2}x_{-1}}}$$

where "compatible" means compatible with the conditioning as explained earlier. Let

$$g_{x_{t+2}} = \sum_{\substack{x_{t+3}, \dots x_{-1} \\ \text{compatible}}} m_{x_{t+2}x_{t+3}} \dots m_{x_{-2}x_{-1}}$$

if $t < -2$, and $g_{x_{t+2}} = 1$ if $t = 2$.

$$\mu_m(x_{t+1} = b | x_t = a, y_{-1}, y_{-2}, \dots y_{-k})$$

$$= \frac{\mu_m(x_t = a | y_{-1}, \dots y_{-k}) m_{ab} \sum\limits_{\substack{x_{t+2} \\ \text{compatible}}} m_{bx_{t+2}} g_{x_{t+2}}}{\mu_m(x_t = a | y_{-1}, \dots y_{-k}) \sum\limits_{\substack{x_{t+1}, x_{t+2} \\ \text{compatible}}} m_{ax_{t+1}} m_{x_{t+1}x_{t+2}} g_{x_{t+2}}}$$

$$\geq \frac{e^{-2N} \sum\limits_{\substack{x_{t+2} \\ \text{compatible}}} g_{x_{t+2}}}{(N+1) \sum\limits_{\substack{x_{t+2} \\ \text{compatible}}} g_{x_{t+2}}} = \frac{e^{-2N}}{N+1}$$

Let $\gamma_N = e^{-2N}/(N+1)$, which is, then, a lower bound on transition probabilities of the inhomogeneous process $\{x_t\}$, conditioned on arbitrary $y_{-1}, y_{-2}, \dots y_{-k}$.

For (i), we need to show

$$\overline{\lim_{\delta \to 0}} \sup_{0 \leq k \leq \infty} \sup_{y} \sup_{\substack{m, m' \in \mathcal{M}_N \\ \|m - m'\| < \delta}} |\mu_m(y_0 | y_{-1}, \dots y_{-k}) - \mu_{m'}(y_0 | y_{-1}, \dots y_{-k})| = 0$$

The left-hand side can be bounded as follows:

$$\overline{\lim_{\delta \to 0}} \sup_{0 \leq k \leq \infty} \sup_{y} \sup_{\substack{m, m' \in \mathcal{M}_N \\ \|m - m'\| < \delta}} |\mu_m(y_0 | y_{-1}, \dots y_{-k}) - \mu_{m'}(y_0 | y_{-1}, \dots y_{-k})|$$

$$= \overline{\lim_{t\to-\infty}} \lim_{\delta\to 0} \sup_{0\le k\le\infty} \sup_{y} \sup_{\substack{m,m'\in\mathcal{M}_N \\ \|m-m'\|<\delta}} \sup_{\substack{x_t \\ \text{compatible}}}$$

$$|(\mu_m(y_0|y_{-1},\ldots y_{-k}) \quad - \mu_m(y_0|x_t,y_{-1},\ldots y_{-k}))$$

$$-(\mu_{m'}(y_0|y_{-1},\ldots y_{-k}) - \mu_m(y_0|x_t,y_{-1},\ldots y_{-k}))$$

$$+(\mu_m(y_0|x_t,y_{-1},\ldots y_{-k}) - \mu_{m'}(y_0|x_t,y_{-1},\ldots y_{-k}))|$$

$$\le 2\overline{\lim_{t\to-\infty}} \sup_{0\le k\le\infty} \sup_{y} \sup_{m\in\mathcal{M}_N} \sup_{\substack{x_t \\ \text{compatible}}} |\mu_m(y_0|y_{-1},\ldots y_{-k}) - \mu_m(y_0|x_t,y_{-1},\ldots y_{-k})|$$

$$+ \overline{\lim_{t\to-\infty}} \lim_{\delta\to 0} \sup_{0\le k\le\infty} \sup_{y} \sup_{\substack{m,m'\in\mathcal{M}_N \\ \|m-m'\|<\delta}} \sup_{\substack{x_t \\ \text{compatible}}} \tag{4}$$

$$|\mu_m(y_0|x_t,y_{-1},\ldots y_{-k}) - \mu_{m'}(y_0|x_t,y_{-1},\ldots y_{-k})|$$

We will show that each of the two terms in (4) is zero. To address the first term, rewrite $\mu_m(y_0|y_{-1},\ldots y_{-k})$ as follows:

$$\mu_m(y_0|y_{-1},\ldots y_{-k}) = \sum_{\substack{x'_t \\ \text{compatible}}} \mu_m(y_0|x'_t,y_{-1},\ldots y_{-k}) \cdot \mu_m(x'_t|y_{-1},\ldots y_{-k})$$

Then

$$|\mu_m(y_0|y_{-1},\ldots y_{-k}) - \mu_m(y_0|x_t,y_{-1},\ldots y_{-k})|$$

$$\le \sup_{\substack{x_t,x'_t \\ \text{compatible}}} |\mu_m(y_0|x_t,y_{-1},\ldots y_{-k}) - \mu_m(y_0|x'_t,y_{-1},\ldots y_{-k})|$$

and it is enough to show that

$$\sup_{\substack{x_t,x'_t \\ \text{compatible}}} |\mu_m(y_0|x_t,y_{-1},\ldots y_{-k}) - \mu_m(y_0|x'_t,y_{-1},\ldots y_{-k})| \to 0$$

as $t \to -\infty$, uniformly in k, y, and m, which we now do via a simple coupling argument.

Fix k, y, m, t, x_t, and x'_t, and consider two stochastic processes, one beginning at x_t at time t and the other beginning at x'_t at time t, each of which follows the inhomogeneous transition probabilities of the x process conditioned on $y_{-1}, \dots y_{-k}$. We couple these processes by defining a common source of randomness, namely a sequence of independent uniform random variables on $[0,1], U_1, U_2, U_3, \dots$. U_1 governs the transition from x_t to x_{t+1} *and* from x'_t to x'_{t+1}. U_2 governs the transition from x_{t+1} to x_{t+2} *and* from x'_{t+1} to x'_{t+2}, and so on. To be specific, let $\varphi_1, \varphi_2, \dots \varphi_r$ be the allowed states of x_{t+s} (i.e. $\{\varphi_1, \dots \varphi_r\} = f^{-1}(y_{t+s})$ if $-k \le t+s \le -1$, and $\{\varphi_1, \dots \varphi_r\} = \{0, 1, \dots N\}$ otherwise). These same states are the allowed states for x'_{t+s}. Given x_{t+s-1} and x'_{t+s-1}, let

$$\psi_i = \mu_m(x_{t+s} = \varphi_i | x_{t+s-1}, y_{-1}, \dots y_{-k})$$

and

$$\psi'_i = \mu_m(x'_{t+s} = \varphi_i | x'_{t+s-1}, y_{-1}, \dots y_{-k}).$$

The x process goes from x_{t+s-1} to $x_{t+s} = \varphi_i$ if

$$\sum_{j=1}^{i-1} \psi_j \le U_s < \sum_{j=1}^{i} \psi_i.$$

The same rule applies to the transition from x'_{t+s-1} to $x'_{t+s} = \varphi_i$ for the x' process, except that $\{\psi'_j\}$ is used in place of $\{\psi_j\}$.

Observe that $x_{t+s} = x'_{t+s} \Longrightarrow x_{t+u} = x'_{t+u}\ \forall u > s$, i.e. the processes couple. Observe also that the probability of the processes coupling at time $t+s$, given that $x_{t+s-1} \ne x'_{t+s-1}$, is at least $\min(\psi_1, \psi'_1) \ge \gamma_N$.

Finally, letting P denote measure under this coupling,

$$|\mu_m(y_0 | x_t, y_{-1}, \dots y_{-k}) - \mu_m(y_0 | x'_t, y_{-1}, \dots y_{-k})|$$

$$= |P(x_0 \in f^{-1}(y_0)) - P(x'_0 \in f^{-1}(y_0))|$$

$$\leq P(x_0 \neq x'_0) \leq (1-\gamma_N)^{-t}$$

independent of k, y, m, x_t and x'_t. Hence

$$\sup_{\substack{x_t, x'_t \\ \text{compatible}}} |\mu_m(y_0|x_t, y_{-1}, \dots y_{-k}) - \mu_m(y_0|x'_t, y_{-1}, \dots y_{-k})| \to 0$$

uniformly in k, y, and m as $t \to -\infty$, as required.

It remains to show that the second term in (4) is also zero. We will show that, in fact, for each fixed t:

$$\overline{\lim_{\delta \to 0}} \sup_{0 \leq k \leq \infty} \sup_{y} \sup_{\substack{m, m' \in \mathcal{M}_N \\ \|m - m'\| < \delta}} \sup_{\substack{x_t \\ \text{compatible}}} |\mu_m(y_0|x_t, y_{-1}, \dots y_{-k})$$

$$-\mu_{m'}(y_0|x_t, y_{-1}, \dots y_{-k})| = 0.$$

We can write down each of the conditional probabilities directly. Let $\tau = \max(t+1, -k)$. Then

$$\mu_m(y_0|x_t, y_{-1}, \dots y_{-k}) = \mu_m(y_0|x_t, y_{-1}, \dots y_\tau)$$

$$= \frac{\mu_m(y_0, y_{-1}, \dots y_\tau | x_t)}{\mu_m(y_{-1}, \dots y_\tau | x_t)}$$

$$= \frac{\displaystyle\sum_{\substack{x_{t+1}, \dots x_{-1} \\ \text{compatible} \\ x_0 \in f^{-1}(y_0)}} m_{x_t x_{t+1}} \cdots m_{x_{-2} x_{-1}} m_{x_{-1} x_0}}{\displaystyle\sum_{\substack{x_{t+1}, \dots x_{-1} \\ \text{compatible}}} m_{x_t x_{t+1}} \cdots m_{x_{-2} x_{-1}}},$$

and the same expression applies to $\mu_{m'}(y_0|x_t, y_{-1}, \dots y_{-k})$, except that m is replaced by m'. Since $m \in \mathcal{M}_N \Longrightarrow m_{ab} \geq e^{-N}$,

$$\sup_{\substack{m, m' \in \mathcal{M}_N \\ \|m - m'\| < \delta}} \sup_{\substack{x_t \\ \text{compatible}}} \mu_m(y_0|x_t, y_{-1}, \dots y_{-k}) - \mu_{m'}(y_0|x_t, y_{-1}, \dots y_{-k})|$$

is uniformly (in k and y) small when δ is small. This completes the proof of Lemma 2.2.1, and hence also the proof of Lemma 2.1.1.

2.4 Proof of Lemma 2.1.2

For convenience, we take the special case $M = 2$ (so that $y_t \in \{0,1\}$). The general case, $2 \leq M < \infty$, follows from the same argument.

The proof is by construction. Let $L_N = [\log_2(N+1)]$ (where $[x]$ = greatest integer less than or equal to x), and for any $\xi \in \{0,1\}^{L_N}$ ($\xi = (\xi_1, \xi_2, \ldots \xi_{L_N})$) define

$$a(\xi) = \xi_1 + 2\xi_2 + 2^2\xi_3 + \cdots + 2^{L_N-1}\xi_{L_N}.$$

Notice that $0 \leq a(\xi) \leq 2^{L_N} - 1 \leq N \ \ \forall \xi \in \{0,1\}^{L_N}$. Let $g_N = 2^{L_N} - 1$, so $a : \{0,1\}^{L_N} \to \{0, 1, \ldots g_N\}$ is one-to-one. Observe that $f(a(\xi)) = \xi_1$.

Now define the probability transition matrix m_N:

$$(m_N)_{ij} = \begin{cases} \mu_o(y_0 = a^{-1}(j)_1 | (y_{-1}, \ldots y_{-L_N}) = a^{-1}(i))(1 - (N-1)e^{-N}) \\ \qquad \text{if } 0 \leq \mathrm{i} \leq \mathrm{g_N},\ 0 \leq \mathrm{j} \leq \mathrm{g_N}, \text{ and} \\ \qquad a^{-1}(j)_k = a^{-1}(i)_{k-1}\ 2 \leq k \leq L_N \\ e^{-N} \\ \qquad \text{if } 0 \leq \mathrm{i} \leq \mathrm{g_N} \text{ andeither } \mathrm{j} > \mathrm{g_N} \text{ or} \\ \qquad \exists\, 2 \leq k \leq L_N \ni\ a^{-1}(j)_k \neq a^{-1}(i)_{k-1} \\ \frac{1-(N-g_N)e^{-N}}{g_N+1} \\ \qquad \text{if } \mathrm{g_N} < \mathrm{i} \leq \mathrm{N},\ 0 \leq \mathrm{j} \leq \mathrm{g_N} \\ e^{-N} \\ \qquad \text{if } \mathrm{g_N} < \mathrm{i} \leq \mathrm{N},\ \mathrm{g_N} < \mathrm{j} \leq \mathrm{N} \end{cases}$$

Recall the constraint on $m \in \mathcal{M}_N: \ m_{ij} \geq e^{-N} \ \forall i,j$. Since $g_N \leq N$,

$$\frac{1-(N-g_N)e^{-N}}{g_N+1} \geq e^{-N}$$

for all N sufficiently large. Thus $m_N \in \mathcal{M}_N$ for all N sufficiently large. We will show that

$$\int \log \mu_{M_N}(y_0|y_{-1}\ldots)d\mu_o(y) - \int \log \mu_o(y_0|y_{-1},\ldots)d\mu_o(y) \to 0.$$

$$\int \log \mu_{M_N}(y_0|y_{-1}\ldots)d\mu_o(y) - \int \log \mu_o(y_0|y_{-1},\ldots)d\mu_o(y)$$

$$= \Big\{ \int \log \mu_{M_N}(y_0|y_{-1},\ldots)d\mu_o(y) - \int \log \mu_{M_N}(x_0 = a(y_0, y_{-1}, \ldots y_{-L_N+1})|$$
$$x_{-1} = a(y_{-1}, y_{-2}, \ldots y_{-L_N}))d\mu_o(y) \Big\}$$
$$+\Big\{ \int \log \mu_{M_N}(x_0 = a(y_0, y_{-1}, \ldots y_{-L_N+1})|x_{-1} = a(y_{-1}, y_{-2}, \ldots y_{-L_N}))d\mu_o(y)$$
$$- \int \log \mu_o(y_0|y_{-1}, y_{-2}, \ldots y_{-L_N})d\mu_o(y) \Big\}$$
$$+\Big\{ \int \log \mu_o(y_0|y_{-1}, y_{-2}, \ldots y_{-L_N})d\mu_o(y) - \int \log \mu_o(y_0|y_{-1}, y_{-2}, \ldots)d\mu_o(y) \Big\}$$

There are three terms. The third goes to zero by dominated convergence (recall that $\forall\, y_0, y_{-1}, \ldots \ \mu_o(y_0|y_{-1},\ldots) \geq \delta > 0$, which implies $\mu_o(y_0|y_{-1},\ldots y_{-L_N}) \geq \delta$ as well, and observe that $\mu_o(y_0|y_{-1},\ldots y_{-L_N}) \to \mu_o(y_0|y_{-1}, y_{-2}, \ldots)$ a.s. $d\mu_o$). As for the second term, when N is sufficiently large,

$$\frac{\mu_{M_N}(x_0 = a(y_0, y_{-1}, \ldots y_{-L_N+1})|x_{-1} = a(y_{-1}, y_{-2}, \ldots y_{-L_N}))}{\mu_o(y_0|y_{-1}, y_{-2}, \ldots y_{-L_N}))}$$
$$= 1-(N-1)e^{-N} \to 1$$

(see definition of m_N).

It remains to show that

$$\int \log \frac{\mu_{M_N}(y_0|y_{-1},\ldots)}{\mu_{M_N}(x_0 = a(y_0,\ldots y_{-L_N+1})|x_{-1} = a(y_{-1},\ldots y_{-L_N}))} d\mu_o(y) \to 0.$$

$$\int \log \frac{\mu_{M_N}(y_0|y_{-1},\ldots)}{\mu_{M_N}(x_0 = a(y_0,\ldots y_{-L_N+1})|x_{-1} = a(y_{-1},\ldots y_{-L_N}))} d\mu_o(y)$$

$$= \int \log \frac{\mu_{M_N}(y_0|y_{-1},\ldots)}{\mu_{M_N}(x_0 = a(y_0,\ldots y_{-L_N+1})|x_{-1} = a(y_{-1},\ldots y_{-L_N}), y_{-1}, y_{-2},\ldots)} d\mu_o(y)$$

$$=^{1} \int \log \lim_{k\to\infty}$$

$$\frac{\mu_{M_N}(y_0|y_{-1},\ldots y_{-k})}{\mu_{M_N}(x_0 = a(y_0,\ldots y_{-L_N+1})|x_{-1} = a(y_{-1},\ldots y_{-L_N}), y_{-1}, y_{-2},\ldots y_{-k})} d\mu_o(y)$$

$$= \int \log \lim_{k\to\infty}$$

$$\frac{\mu_{M_N}(y_0, y_{-1},\ldots y_{-k})\mu_{M_N}(x_{-1} = a(y_{-1},\ldots y_{-L_N}), y_{-1},\ldots y_{-k})}{\mu_{M_N}(x_0 = a(y_0,\ldots y_{-L_N+1}), x_{-1} = a(y_{-1},\ldots y_{-L_N}), y_{-1},\ldots y_{-k})\mu_{M_N}(y_{-1},\ldots y_{-k})} d\mu_o(y)$$

$$= \int \log \lim_{k\to\infty}$$

$$\frac{\mu_{M_N}(y_0, y_{-1},\ldots y_{-k})\mu_{M_N}(x_{-1} = a(y_{-1},\ldots y_{-L_N}), y_{-1},\ldots y_{-k})}{\mu_{M_N}(x_0 = a(y_0,\ldots y_{-L_N+1}), x_{-1} = a(y_{-1},\ldots y_{-L_N}), y_0,\ldots y_{-k})\mu_{M_N}(y_{-1},\ldots y_{-k})} d\mu_o(y)$$

$$= \int \log \lim_{k\to\infty} \frac{\mu_{M_N}(x_{-1} = a(y_{-1},\ldots y_{-L_N})|y_{-1},\ldots y_{-k})}{\mu_{M_N}(x_0 = a(y_0,\ldots y_{-L_N+1}), x_{-1} = a(y_{-1},\ldots y_{-L_N})|y_0, y_{-1},\ldots y_{-k})} d\mu_o(y)$$

$$= \int \log \frac{\mu_{M_N}(x_{-1} = a(y_{-1},\ldots y_{-L_N})|y_{-1}, y_{-2},\ldots)}{\mu_{M_N}(x_0 = a(y_0,\ldots y_{-L_N+1}), x_{-1} = a(y_{-1},\ldots y_{-L_N})|y_0, y_{-1},\ldots)} d\mu_o(y).$$

---

[1]Conditioned on $y_{-1}, y_{-2}, \ldots$, x is first order Markov with positive (and inhomogeneous) transition probabilities bounded uniformly below (see above proof of Lemma 2.2.1). From this, it follows that x is strongly mixing, and the conditional distributions converge uniformly in y as $k \to \infty$.

We will show that both $\mu_{M_N}(x_{-1} = a(y_{-1}, \ldots y_{-L_N})|y_{-1}, y_{-2}, \ldots)$ and $\mu_{M_N}(x_0 = a(y_0, \ldots y_{-L_N+1}), x_{-1} = a(y_{-1}, \ldots y_{-L_N})|y_0, y_{-1}, \ldots)$ converge uniformly (in y) to 1, which will then complete the proof of the lemma.

The argument is much the same for both terms; we give details for the second only.

Given $y_0, y_{-1}, \ldots$, call the sequence $x_{-L_N}, x_{-L_N+1}, \ldots x_0$ "canonical" if $x_t \in \{0, 1, \ldots g_N\}$ and $a^{-1}(x_t)_1 = y_t$ (consistent with $f(x_t) = y_t$) $\forall t = 0, -1, \ldots - L_N$, and if, for every $t = 0, -1, \ldots - L_N + 1$,

$$a^{-1}(x_t)_k = a^{-1}(x_{t-1})_{k-1} \; \forall 2 \le k \le L_N. \tag{5}$$

Given x_{-L_N}, and given $y_0, y_{-1}, \ldots$, there is only one canonical sequence, since (5) determines all but the first component of $a^{-1}(x_t)$ in terms of x_{t-1}, and since $a^{-1}(x_t)_1 = y_t$. Also, given $y_0, y_{-1}, \ldots$, every canonical sequence ends with $x_{-1} = a(y_{-1}, \ldots y_{-L_N})$ and $x_0 = a(y_0, \ldots y_{-L_N+1})$. Hence

$$\mu_{M_N}(x_0 = a(y_0, \ldots y_{-L_N+1}), x_{-1} = a(y_{-1}, \ldots y_{-L_N})|y_0, y_{-1}, \ldots)$$

$$\ge \mu_{M_N}(x_{-L_N}, \ldots x_0 \text{ canonical}|\mathrm{y}_0, \mathrm{y}_{-1}, \ldots)$$
$$= \sum_{\xi=0}^{N} \mu_{M_N}(x_{-L_N}, \ldots x_0 \text{ canonical}|\mathrm{x}_{-\mathrm{L_N}-1} = \xi, \mathrm{y}_0, \mathrm{y}_{-1}, \ldots)$$
$$\cdot \mu_{M_N}(x_{-L_N-1} = \xi|y_0, y-1, \ldots).$$

Given x_{-L_N-1}, let us call the "weight" of a sequence $x_{-L_N}, \ldots x_0$ its *a priori* probability under μ_{M_N}:

$$\text{weight } (\mathrm{x}_{-\mathrm{L_N}}, \ldots \mathrm{x}_0|\mathrm{x}_{-\mathrm{L_N}-1}) \doteq \prod_{\mathrm{k}=-\mathrm{L_N}}^{0} (\mathrm{m_N})_{\mathrm{x}_{\mathrm{k}-1}, \mathrm{x}_\mathrm{k}}.$$

With this notation:

$$\mu_{M_N}(x_0 = a(y_0, \dots y_{-L_N+1}), x_{-1} = a(y_{-1}, \dots y_{-L_N}) | y_0, y_{-1}, \dots)$$

$$\geq \sum_{\xi=0}^{N} \frac{\sum\limits_{\substack{x_{-L_N}, \dots x_0 \\ \text{canonical}}} \mathrm{weight}(\mathrm{x_{-L_N}}, \dots \mathrm{x_0} | \xi)}{\sum\limits_{\substack{\mathrm{x_{-L_N}}, \dots \mathrm{x_0} \\ \mathrm{f(x_t) = y_t, t=0,-1,\dots -L_N}}} \mathrm{weight}(\mathrm{x_{-L_N}}, \dots \mathrm{x_0} | \xi)} \mu_{M_N}(x_{-L_N-1} = \xi | y_0, y_{-1}, \dots).$$

Therefore, it is enough to show that

$$\frac{\sum\limits_{\substack{x_{-L_N}, \dots x_0 \\ \text{canonical}}} \mathrm{weight}(\mathrm{x_{-L_N}}, \dots \mathrm{x_0} | \xi)}{\sum\limits_{\substack{\mathrm{x_{-L_N}}, \dots \mathrm{x_0} \\ \mathrm{f(x_t) = y_t, t=0,-1,\dots -L_N}}} \mathrm{weight}(\mathrm{x_{-L_N}}, \dots \mathrm{x_0} | \xi)} \longrightarrow 1$$

uniformly in $\xi \in \{0, \dots N\}$. Let

$$A_N = \sum_{\substack{x_{-L_N}, \dots x_0 \\ \text{canonical}}} \mathrm{weight}(\mathrm{x_{-L_N}}, \dots \mathrm{x_0} | \xi)$$

and

$$B_N = \sum_{\substack{x_{-L_N}, \dots x_0 \\ \text{notcanonical} \\ f(x_t) = y_t, t=0,1,\dots -L_N}} \mathrm{weight}(\mathrm{x_{-L_N}}, \dots \mathrm{x_0} | \xi).$$

Then, in terms of A_N and B_N, it remains to show that $A_N/(A_N + B_N) \to 1$, equivalently $B_N/A_N \to 0$, uniformly in $\xi \in \{0, \dots N\}$.

Fix $\xi \in \{0, \dots g_N\}$, and define $\hat{x}_{-L_N}, \dots \hat{x}_0$ to be the unique canonical sequence satisfying $a^{-1}(x_{-L_N})_k = a^{-1}(\xi)_{k-1}$, $2 \leq k \leq L_N$. Then

$$A_N \geq \mathrm{weight}(\hat{\mathrm{x}}_{-\mathrm{L_N}}, \dots \hat{\mathrm{x}}_0 | \xi) \geq \delta^{\mathrm{L_N}+1}(1 - (\mathrm{N}-1)\mathrm{e}^{-\mathrm{N}})^{\mathrm{L_N}+1}.$$

On the other hand, if $\xi \in \{g_N + 1, \dots N\}$, pick $\hat{x}_{-L_N} \in \{0, \dots g_N\}$ arbitrary and let $\hat{x}_{-L_N}, \dots \hat{x}_0$ be the resulting canonical sequence. Then

$$A_N \geq \mathrm{weight}(\hat{\mathrm{x}}_{-\mathrm{L_N}}, \dots \hat{\mathrm{x}}_0 | \xi) \geq \frac{1 - (\mathrm{N} - \mathrm{g_N})\mathrm{e}^{-\mathrm{N}}}{\mathrm{g_N} + 1} \delta^{\mathrm{L_N}}(1 - (\mathrm{N}-1)\mathrm{e}^{-\mathrm{N}})^{\mathrm{L_N}}. \quad (6)$$

When N is large, the bound in (6) is smaller, and therefore serves as a lower bound for A_N for all $\xi \in \{0, \ldots N\}$.

As for B_N, the expression for the weight of each non-canonical sequence contains at least one term of size e^{-N}, and therefore is no bigger than e^{-N}. Since there are no more than $(N+1)^{L_N+1}$ terms in the sum that constitutes B_N:

$$B_N \leq (N+1)^{L_N+1} e^{-N}.$$

It is now an easy matter to verify that B_N / A_N, which is bounded above by

$$\frac{(N+1)^{L_N+1} e^{-N}}{\frac{1-(N-g_N)e^{-N}}{g_N+1} \delta^N (1-(N-1)e^{-N})^{L_N}},$$

converges to zero, which completes the proof.

3 The general case

3.1 Gibbs measures

Let S be the d-dimensional lattice $\mathcal{Z}^d$ and E be a finite set, the state space. The configuration space is then $\Omega = E^S$. A shift-invariant, summable potential is a collection of functions $\Phi = \{\Phi_V\}_{V \subset S, finite}$, such that

1. $\Phi_V : E^V \longrightarrow \mathcal{R}$,

2. $\Phi_{V+t} = \Phi_V \quad \forall t \in S$,

3. $\|\Phi\| = \sum_{V \ni 0} \sup_{x \in \Omega} |\Phi_V(x_V)| < \infty$.

A Gibbs measure with potential Φ is any probability measure ν on Ω such that for any finite $V \subset S$ and $x \in \Omega$

$$\nu[X_V = x_V | X_{V^c} = x_{V^c}] = \pi_V^\Phi[x_V | x_{V^c}] = Z_V^\Phi(x_{V^c}) \exp(-H_V^\Phi(x))$$

where

$$H_V^\Phi(x) = \sum_{W \cap V \neq \phi} \Phi_W(x_W)$$

is the energy of configuration x in V and

$$Z_V^\Phi(x_{V^c}) = \sum_{x_V} \exp(-H_V^\Phi(x))$$

is a normalizing constant, the so-called partition function. The set of all stationary Gibbs measures with potential Φ will be denoted by $\mathcal{G}_s(\Phi)$.

A nearest neighbor potential is a potential Φ with $\Phi_V \equiv 0$ except if $V = \{t\}$ or $V = \{t, s\}$ with $\|t - s\| = 1$. For such a potential we denote $\Phi_{\{t\}}$ by Ψ_0 and $\Phi_{\{t,t+e_i\}}$ by $\Psi_i (1 \leq i \leq d)$ where $e_i \in \mathcal{Z}^d$ has i-th component 1 and all other components 0.

We briefly introduce some thermodynamic quantities we will use in the sequel: For any shift-invariant, summable potential Φ the pressure $p(\Phi)$ is defined as

$$p(\Phi) = \lim_{V \uparrow S} |V|^{-1} \log Z_V^\Phi(x_{V^c}).$$

For two stationary probabilities μ and ν on Ω we define the specific entropy

$$s(\mu) = - \lim_{V \uparrow S} |V|^{-1} E_\mu[\log \mu[X_V = x_V]]$$

and the specific relative entropy

$$h(\mu, \nu) = \liminf_{V \uparrow S} |V|^{-1} E_\mu[\log(\mu[X_V = x_V] / \nu[X_V = x_V])].$$

The first two limits do indeed exist and p is independent of the boundary condition x provided the limit $V \uparrow S$ is taken along any sequence (V_n) with $|\partial V_n|/|V_n| \to 0$ (see e.g. Georgii [2], Chapter 15). Obviously we can write

$$h(\mu, \nu) = -s(\mu) - d(\mu, \nu)$$

where

$$d(\mu, \nu) = \limsup_{V \uparrow S} |V|^{-1} E_\mu[\log \nu[X_V = x_V]].$$

Finally we will need that for $\nu \in \mathcal{G}_s(\Phi)$

$$h(\mu, \nu) = -s(\mu) + \sum_{V \ni 0} E_\mu[\Phi_V(x_V)]/|V| + p(\Phi)$$

or respectively

$$d(\mu, \nu) = -\sum_{V \ni 0} E_\mu[\Phi_V(x_V)]/|V| - p(\Phi)$$

(Georgii [2], Theorem 15.30). This follows easily upon replacing $\nu[X_V = x_V]$ by $\pi_V^\Phi[x_V | x_{V^c}]$.

3.2 Main result

We suppose that we observe one realization of a stationary random field Z with state space $E = \{0, 1, \ldots, M-1\}$ for some $M > 1$ on a window $V \subset S$ finite. Let μ_0 be the unknown law of Z. We want to approximate μ_0 by the law of a hidden field $Y = \overline{f}(X)$ where X is a nearest neighbor Gibbs measure with state space $E' = \{0, 1, \ldots, N\}$. Here $\overline{f} : E'^s \to E^s$ is connected to a local hiding function $f : E' \to E$ by $\overline{f}(x)_t = f(x_t)(t \in S)$ and f is fixed once and for all

$$f(k) = k \bmod M.$$

The potential of the unobserved X is chosen from the following set

$$\begin{aligned}\mathcal{P}_N = \quad & \{ \text{ nearest neighbor potentials } \Psi = (\Psi_0, \ldots, \Psi_d) \\ & \text{on } \{0, 1, \ldots, N\} \text{ with } |\Psi_0(k)| \leq N, \\ & |\Psi_i(j,k)| \leq N,\ 1 \leq i \leq d,\ 0 \leq j, k \leq N\}.\end{aligned}$$

To any $\Psi \in \mathcal{P}_N$ we denote the set of hidden Gibbs laws on E^S by

$$\mathcal{H}(\Psi) = \{\mu = \nu \circ \overline{f}^{-1}; \nu \in \mathcal{G}_s(\Psi)\}.$$

For fixed N we select the potential $\Psi \in \mathcal{P}_N$ by maximizing the approximate log likelihood

$$L_V(\Psi, z_V, x_{\partial V}) = \log(\sum_{\substack{x_V \\ f(x_t)=z_t, t \in V}} \pi_V^{\Psi}[x_V | x_{\partial V}])$$

with arbitrary but fixed $x_{\partial V}$. For a discussion of the reasons for this choice see section 3.1 of [4].

Finally define $M_{N,V}$ to be the set of maximum likelihood potentials within $\mathcal{P}_N$:

$$M_{N,V} = M_{N,V}(z_V, x_{\partial V}) = \{\Psi \in \mathcal{P}_N : L_V(\Psi, z_V, x_{\partial V}) = \sup_{\Psi \in \mathcal{P}_N} L_V(\Psi, z_V, x_{\partial V})\}.$$

Then our main result in this Chapter shows consistency for a broad class of true distributions μ_0 for a suitable choice of the sieve parameter N (cf. [4]).

Theorem 3.1 *Let μ_0 be an ergodic Gibbs measure on $\{0, 1, \ldots, M-1\}^S$ with respect to a shift-invariant summable potential Φ and let V_n be a sequence of finite subsets of $S, \cup_n V_n = S, |\partial V_n|/|V_n| \to 0$. Then for all sequences $N_n \uparrow \infty$ sufficiently slowly*

$$\sup_x \sup_{\Psi \in ML_{N_n, V_n}} \sup_{\mu \in \mathcal{H}(\Psi)} h(\mu_0, \mu) \to 0 \ a.s.\ (\mu_0).$$

3.3 Two basic lemmas

As in the one-dimensional case, the proof splits up into a uniform law of large numbers for log likelihood L and an approximation result for μ_0 by some sequence $\mu_N \in \mathcal{H}(\Psi_N), \Psi_N \in \mathcal{P}_N$. Define

$$g_V(\Psi, z_V) = \sup_{\mu \in \mathcal{H}(\Psi)} \sup_{x_{\partial V}} \big|\, |V|^{-1} L_V(\Psi, z_V, x_{\partial V}) - d(\mu_0, \mu)\big|.$$

Lemma 3.3.1 *For all $N_n \uparrow \infty$ sufficiently slowly* $\lim_n \sup_{\Psi \in \mathcal{P}_{N_n}} g_V(\Psi, z_V) = 0$ *a.s.* (μ_0).

Lemma 3.3.2 *There exists a sequence $\Psi_N \in \mathcal{P}_N$ such that*

$$\lim_N \sup_{\mu_N \in \mathcal{H}(\Psi_N)} h(\mu_0, \mu_N) = 0.$$

These two lemmas together imply Theorem 3.1. To see this, take any $\Psi \in ML_{N_n, V_n}, \mu \in \mathcal{H}(\Psi), \Psi_{N_n}$ as in Lemma 3.3.2. and $\mu_{N_n} \in \mathcal{H}(\Psi_{N_n})$. Then

$$\begin{aligned}
0 &\leq h(\mu_0, \mu) = -s(\mu_0) - d(\mu_0, \mu) \\
&\leq -s(\mu_0) - |V_n|^{-1} L_{V_n}(\Psi, z_{V_n}, x_{\partial V_n}) + \sup\nolimits_{\Psi \in \mathcal{P}_{N_n}} g_{V_n}(\Psi, z_{V_n}) \\
&\leq -s(\mu_0) - |V_n|^{-1} L_{V_n}(\Psi_{N_n}, z_{V_n}, x_{\partial V_n}) + \sup\nolimits_{\Psi \in \mathcal{P}_{N_n}} g_{V_n}(\Psi, z_{V_n}) \\
&\leq -s(\mu_0) - d(\mu_0, \mu_{N_n}) + 2 \sup\nolimits_{\Psi \in \mathcal{P}_{N_n}} g_{V_n}(\Psi, z_{V_n}) \\
&= h(\mu_0, \mu_{N_n}) + 2 \sup\nolimits_{\Psi \in \mathcal{P}_{N_n}} g_{V_n}(\Psi, z_{V_n}).
\end{aligned}$$

The two terms on the right go to zero by the two lemmas above.

3.4 Proof of the first basic lemma

The basic estimates to prove Lemma 3.3.1 are collected in

Lemma 3.4.1 *For any* $V \subset S$ *finite,* $W \subseteq V, z \in \{0, \dots, M-1\}^S, x, x' \in \{0, \dots, N\}^S, \Psi, \Psi' \in \mathcal{P}_N, \mu \in \mathcal{H}(\Psi)$ *the following holds:*

i) $|L_V(\Psi, z_V, x_{\partial V}) - L_V(\Psi, z_V, x'_{\partial V})| \leq 4dN|\partial V|$,

ii) $|L_V(\Psi, z_V, x_{\partial V}) - \log \mu[Y_V = z_V]| \leq 4dN|\partial V|$,

iii) $|L_V(\Psi, z_V, x_{\partial V}) - L_V(\Psi', z_V, x_{\partial V})| \leq 2(2d+1)\|\Psi - \Psi'\| \; |V|$,

iv) $|L_V(\Psi, z_V, x_{\partial V}) - L_W(\Psi, z_W, x_{\partial W}) - L_{V \setminus W}(\Psi, z_{V \setminus W}, x_{\partial(V \setminus W)})|$
$\leq 4dN|\partial W|$.

Proof

For i) we use that for any $x_V, x_{\partial V}, x'_{\partial V}$

$$|H_V^\Psi(x) - H_V^\Psi(x_V x'_{\partial V})| \leq 2dN|\partial V| \tag{7}$$

because each $s \in \partial V$ is connected with at most $2d$ neighbors. Taking exponentials and summing over x_V this implies

$$\exp(-2dN|\partial V|)Z_V^\Psi(x'_{\partial V}) \leq Z_V^\Psi(x_{\partial V}) \leq \exp(2dN|\partial V|)Z_V^\Psi(x'_{\partial V}). \tag{8}$$

Hence

$$\exp(-4dN|\partial V|)\pi_V^\Psi(x_V|x'_{\partial V})$$
$$\leq \pi_V^\Psi(x_V|x_{\partial V}) \leq \exp(4dN|\partial V|)\pi_V^\Psi(x_V|x'_{\partial V}) \tag{9}$$

Summing this over all x_V with $f(x_t) = z_t (t \in V)$ shows i).

For ii) we observe that for $\nu \in \mathcal{G}_s(\Psi)$

$$\nu[X_V = x_V] = \int \pi_V^\Psi(x_V|x'_{\partial V})d\nu(x').$$

Thus (3.3.4) implies that also

$$\exp(-4dN|\partial V|)\pi_V^\Psi(x_V|x_{\partial V}) \leq \nu[X_V = x_V] \leq \exp(4dN|\partial V|)\pi_V^\Psi(x_V|x_{\partial V}).$$

From this ii) follows immediately.

The proofs of iii) and iv) are similar to the proof of i). First we estimate the difference between energies and then we deduce inequalities for partition functions and conditional probabilities. Details are left to the reader.

□

Lemma 3.4.2 *For any fixed N we have*

i) $d(\mu_0, \mu) = \lim_n |V_n|^{-1} \int L_{V_n}(\Psi, z_{V_n}, x_{\partial V_n})\mu_0(dz)$ *for any* $x, \Psi \in \mathcal{P}_N, \mu \in \mathcal{H}(\Psi)$.

ii) $\lim_n g_{V_n}(\Psi, z_{V_n}) = 0$ *a.s.* (μ_0) *for any* $\Psi \in \mathcal{P}_N$.

iii) $\lim_n \sup_{\Psi \in \mathcal{P}_N} g_{V_n}(\Psi, z_{V_n}) = 0$ *a.s.* (μ_0).

Proof

Let Λ_m be the cube $\{1, 2, \ldots m\}^d, \Lambda_{m,t} = \Lambda_m + tm (t \in S), I_n = \{t \in S; \Lambda_{m,t} \subseteq V_n\}$.

From Lemma 3.4.1 iii) and iv)

$$|L_{V_n}(\Psi, z_{V_n}, x_{\partial V_n}) - \sum_{t \in I_n} L_{\Lambda_{m,t}}(\Psi, z_{\Lambda_{m,t}}, x_{\partial \Lambda_{m,t}})|$$

$$\leq 8d^2 N m^{d-1}|I_n| + 2(2d+1)(d+1)N(|V_n| - |I_n|m^d).$$

Consider first the particular boundary condition $x_t \equiv 0 (t \in S)$. Then by the stationarity of μ_0

$$\begin{aligned} &| \, |V_n|^{-1} \int L_{V_n}(\Psi, z_{V_n}, x_{\partial V_n}) \mu_0(dz) \\ &- |I_n| m^d |V_n|^{-1} m^{-d} \int L_{\Lambda_m}(\Psi, z_{\Lambda_m}, x_{\partial\Lambda_m}) \mu_0(dz)| \\ &\leq \text{const.} m^{-1} + \text{const.}(1 - |I_n| m^d |V_n|^{-1}). \end{aligned}$$

But $|\partial V_n|/|V_n| \to 0$ implies that for any fixed m $|I_n|m^d|V_n|^{-1} \to 0$. Hence by choosing first m and then n one sees easily that with $x_t \equiv 0$ the limit

$$\lim_n |V_n|^{-1} \int L_{V_n}(\Psi, z_{V_n}, x_{\partial V_n}) \mu_0(dz)$$

exists. Part i) follows now easily using Lemma 3.4.1 i) and ii).

In order to prove ii), we again choose first $x_t \equiv 0$. Then by the ergodic theorem

$$|I_n|^{-1} \sum_{t \in I_n} L_{\Lambda_{m,t}}(\Psi, z_{\Lambda_{m,t}}, x_{\partial\Lambda_{m,t}}) \to \int L_{\Lambda_m}(\Psi, z_{\Lambda_m}, x_{\partial\Lambda_m}) \mu_0(dz)$$

a.s. (μ_0) for any fixed m. Together with the previous estimate and part i) we see by choosing first m and then n that a.s. (μ_0)

$$\lim_n |V_n|^{-1} L_{V_n}(\Psi, z_{V_n}, x_{\partial V_n}) = d(\mu_0, \mu).$$

To handle an arbitrary boundary condition, we use Lemma 3.4.1, i).

Finally iii) is an immediate consequence of part ii), Lemma 3.4.1 iii) and the compactness of $\mathcal{P}_N$ with respect to $\|\Psi\|$. □

Finally to obtain Lemma 3.3.1 from Lemma 3.4.2 iii) we use the same Borel-Cantelli argument as in the one-dimensional case (see section 2).

3.5 Proof of the second basic lemma

To make the notation easier, we assume $M = 2$. Denote by $C_\ell = \{-\ell, -\ell+1, \ldots, \ell\}^d$ a cube of side length $2\ell + 1$. First we assume that $N + 1 = 2^{|C_\ell|}$ for some $\ell \in \mathcal{N}$. Then there is a bijection between $x_t \in \{0, 1, \ldots, N\}$ and $(\xi_{t,r})_{r \in C_\ell} \in \{0,1\}^{C_\ell}$ such that $x_t \bmod 2 = \xi_{t,0}$. We are going to use the same construction as in the approximation theorem 2.2.1 of [4]. So we define the approximating potential Ψ_N by

$$\Psi_{N,0}(x_t) = \sum_{V \subset C_\ell} \Phi_V(\xi_{t,V})/|C_\ell|$$

$$\Psi_{N,j}(x_t, x_{t+e_j}) = \beta_N \sum_{r \in C_\ell \cap C_\ell + e_j} 1_{[\xi_{t,r} \neq \xi_{t+e_j, r-e_j}]}$$

where $\beta_N = N/(2\ell(2\ell+1)^{d-1}) = O(N/\log N)$ is chosen such that $\Psi_N \in \mathcal{P}_N$ for N large enough. We are going to compute $h(\mu, \mu_N), \mu_N \in \mathcal{H}(\Psi_N)$, by considering the particular sequence $V_n = \{1, \ldots, n\}^d$. It turns out to be most convenient to work with periodic boundary conditions, i.e. we define the energy

$$H_{V_n}^{\Psi_N}(x_{V_n}|\text{per}) = \sum_{t \in V_n} (\Psi_{N,0}(x_t) + \sum_{j=1}^{d} \Psi_{N,j}(x_t, x_{t+e_j}))$$

where all additions $t + e_j$ are modulo n. Moreover we define

$$Z_{V_n}^{\Psi_N}(\text{per}) = \sum_{x_{V_n}} \exp(-H_{V_n}^{\Psi_N}(x_{V_n}|\text{per})),$$

$$\pi_{V_n}^{\Psi_N}(x_{V_n}|\text{per}) = Z_{V_n}^{\Psi_N}(\text{per})^{-1} \exp(-H_{V_n}^{\Psi_N}(x_{V_n}|\text{per})),$$

$$L_{V_n}(\Psi_N, z_{V_n}, \text{per}) = \log(\sum_{\substack{x_{V_n} \\ f(x_t) = z_t}} \pi_{V_n}^{\Psi_N}(x_{V_n}|\text{per})).$$

As in section 3.4 we can show that for any $\mu_N \in \mathcal{H}(\Psi_N)$

$$h(\mu_0, \mu_N) = -s(\mu_0) - \lim_n n^{-d} \int L_{V_n}(\Psi_N, z_{V_n}, \text{per})\mu_0(dz).$$

In order to analyze the integral on the righthand side, we decompose

$$L_{V_n}(\Psi_N, z_{V_n}, \text{per}) = \log \tilde{Z}_{V_n}^{\Psi_N}(z_{V_n}, \text{per}) - \log Z_{V_n}^{\Psi_N}(\text{per})$$

where

$$\tilde{Z}_{V_n}^{\Psi_N}(z_{V_n}, \text{per}) = \sum_{\substack{x_{V_n} \\ f(x_t)=z_t}} \exp(-H_{V_n}^{\Psi_N}(x_{V_n}|\text{per})).$$

Note that $\tilde{Z}_{V_n}^{\Psi_N}$ is the partition function for the potential

$$\Phi_{\{t\}}(x_t) = \Psi_{N,0}(x_t) + \infty 1_{[f(x_t)\neq z_t]}$$

$$\Phi_{\{t,t+e_j\}}(x_t, x_t + e_j) = \Psi_{N,j}(x_t, x_{t+e_j}).$$

We are going to show that only those configurations x_{V_n} contribute asymptotically to the partition function for which the compatibility constraints

$$\xi_{t,r} = \xi_{s,r+t-s}(t \in V_n, s \in V_n, r \in C_\ell, r+t-s \in C_\ell)$$

are satisfied. For $\tilde{Z}_{V_n}^{\Psi_N}$ this leaves only one configuration, namely $\xi_{t,r} = z_{t+r}(t \in V_n, r \in C_\ell$, additions modulo $n)$. For this configuration we have

$$\begin{aligned} H_{V_n}^{\Psi_N}(x_{V_n}|\text{per}) &= \textstyle\sum_{t\in V_n} \Psi_{N,0}(x_t) = \sum_{t\in V_n} \sum_{W\subset C_\ell} \Phi_W(z_{t+W})/|C_\ell| \\ &= \textstyle\sum_{W\cap V_n \neq \phi} \Phi_W(z_W)\alpha(W,\ell), \end{aligned}$$

where $\alpha(W,\ell) = |\{t \in S; W+t \subset C_\ell\}|/|C_\ell|$ and the configuration z_{V_n} is extended periodically.

Defining the potential Φ_N by

$$\Phi_{N,V}(z_V) = \Phi_V(z_V)\alpha(V,\ell),$$

we have

$$H_{V_n}^{\Psi_N}(x_{V_n}|\text{per}) = H_{V_n}^{\Phi_N}(z_{V_n}|\text{per}).$$

Next we bound the contribution of other configurations to $\tilde{Z}{}_{V_n}^{\Psi_N}$

Lemma 3.5.1 *For all N and n*

$$-H_{V_n}^{\Phi_N}(z_{V_n}|per)$$

$$\leq \log \tilde{Z}{}_{V_n}^{\Psi_N}(z_{V_n}|per) \leq -H_{V_n}^{\Phi_N}(v_n|per)+|V_n|(|C_\ell|-1)\exp(-\beta_N/\log N+2\|\Phi\|).$$

Proof

Let $x_{V_n}^0$ be the configuration with $\xi_{t,r}^0 = z_{t+r}$ and x_{V_n} any other configuration with $f(x_t) = \xi_{t,0} = z_t = \xi_{t,0}^0 (t \in V_n)$. Consider the set

$$I = \{(t,r) \in V_n \times C_\ell; \xi_{t,r} \neq \xi_{t,r}^0\}$$

and assume $|I| = k$. We claim that

$$\sum_{t\in V_n}\sum_{j=1}^{d}\Psi_{N,j}(x_t, x_{t+e_j}) \geq \frac{k}{|C_\ell|}\beta_N = k\beta_N/\log N. \tag{10}$$

To show this consider the graph with vertices $V_n \times C_\ell$ and edges between (t,r) and (s,u) iff $\|t-s\| = 1$ and $t+r = s+u$. In this graph two vertices (t,r) and (s,u) are linked through a chain of edges iff $t+r = s+u$. Hence the graph decomposes into $|V_n|$ connected components of size $|C_\ell|$. So I contains vertices in at least $[k/(|C_\ell|-1)]+1$ different such components ($[a]$ denotes the integer part of a real number a). On the other hand, each component which contains a vertex from I contains at least one edge $(t,r),(s,u)$ such that $\xi_{t,r} \neq \xi_{s,u}$. This is easily seen by a counter argument. If no such edge exists, then $\xi_{t,r}$ must be constant on this component. But $\xi^0_{t,r}$ is constant on each component and $\xi_{t,0} = \xi^0_{t,0}$ for each $t \in V$ which is a contradiction. Together this proves (10).

In addition we obviously have

$$|\sum_{t\in V_n} \Psi_{N,0}(x_t^0) - \sum_{t\in V_n} \Psi_{N,0}(x_t)| \le 2 \sum_{\substack{t\in V_n \\ (t,C_\ell)\cap I\neq\emptyset}} \sum_{W\subseteq C_\ell} \sup_{z_W} |\Phi_W(z_W)|/|C_\ell| \le 2k\|\Phi\|. \tag{11}$$

Taking (10) and (11) together we obtain

$$\exp(-H_{V_n}^{\Psi_N}(x_{V_n}|\text{per})) \le \exp(-H_{V_n}^{\Phi_N}(z_{V_n}|\text{per}) - k(\beta_N/\log N - 2\|\Phi\|))$$

Summing over all configurations x_{V_n} with $f(x_t) = z_t (t \in V_n)$ thus gives with $m = |V_n|(|C_\ell| - 1)$

$$\begin{aligned} &\overset{\sim\Psi_N}{Z_{V_n}}(z_{V_n}|\text{per}) \\ &\le \exp(-H_{V_n}^{\Phi_N}(z_{V_n}|\text{per})) \sum_{k=0}^{m} \binom{m}{k} \exp(-k(\beta_N/\log N - 2\|\Phi\|)) \\ &= \exp(-H_{V_n}^{\Phi_N}(z_{V_n}|\text{per}))(1+\exp(-\beta_N/\log N + 2\|\Phi\|))^m. \end{aligned}$$

Since $\log(1+a) \le a$, this proves the second inequality. The first inequality is obvious because all terms contributing to $\overset{\sim\Psi_N}{Z_{V_n}}$ are positive.

□

Next we consider $Z_{V_n}^{\Psi_N}(\text{per})$. Because

$$Z_{V_n}^{\Psi_N}(\text{per}) = \sum_{z_{V_n}} \tilde{Z}_{V_n}^{\Psi_N}(z_{V_n}|\text{per}),$$

the following is an immediate consequence of Lemma 3.5.1.

Lemma 3.5.2 *For all N and n:*

$$\log Z_{V_n}^{\Phi_N}(per) \leq \log Z_{V_n}^{\Psi_N}(per)$$

$$\leq \log Z_{V_n}^{\Phi_N}(per) + |V_n|(|C_\ell| - 1)\exp(-\beta_N/\log N + 2\|\Phi\|).$$

From these two lemmas we obtain after dividing by $|V_n|$ and letting $n \to \infty$ that for any $N, \mu_N \in \mathcal{H}(\Psi_N)$

$$|d(\mu_0, \mu_N) + \sum_{W \ni 0} E_{\mu_0}[\Phi_W(z_W)]|W|^{-1}\alpha(W, \ell) + p(\Phi_N)|$$
$$\leq 2(|C_\ell| - 1)\exp(-\beta_N/\log N + 2\|\Phi\|).$$

Finally we let N tend to infinity. Since $|C_\ell| = O(\log N)$ and $\beta_N = O(N/\log N)$ we obtain (c.f. section 3.1)

$$\lim_N d(\mu_0, \mu_N) = -\sum_{W \ni 0} E_{\mu_0}[\Phi_W(z_W)]|W|^{-1} - p(\Phi) = d(\mu_0, \mu_0) = -s(\mu_0).$$

This is the claim of Lemma 3.3.2 for the subsequence $N = 2^{|C_\ell|} - 1, \ell \in \mathcal{N}$. It remains to define μ_N for general N. For this we construct to any $\Psi_N \in \mathcal{P}_N$ a $\Psi_{N+1} \in \mathcal{P}_{N+1}$ such that

$$h(\mu_0, \mu_N) = h(\mu_0, \mu_{N+1}) \quad (\mu_j \in \mathcal{H}(\Psi_j), j = N, N+1). \tag{12}$$

Let $\varphi : \{0, \ldots, N+1\} \to \{0, \ldots, N\}$ be defined as $\varphi(i) = i(i \leq N), \varphi(N+1) = \varphi(N-1)$ and set

$$\Psi_{N+1,0}(x_t) = \begin{cases} \Psi_{N,0}(x_t) \text{ if } x_t \neq N-1, N+1 \\ \Psi_{N,0}(N-1) + \log(2) \text{ otherwise,} \end{cases}$$

$$\Psi_{N+1,j}(x_t, x_s) = \Psi_{N,j}(\varphi(x_t), \varphi(x_s)).$$

This means that the two states $N-1$ and $N+1$ are interchangeable. In particular

$$H_{V_n}^{\Psi_{N+1}}(x_{V_n}|\text{per}) = H_{V_n}^{\psi_N}((\varphi(x_t))_{t \in V_n}|\text{per}) + k \log 2$$

where $k = \{t \in V_n : x_t = N-1 \text{ or } x_t = N+1\}$. Because also $f(\varphi(x_t)) = f(x_t)$, this shows that

$$\tilde{Z}{}_{V_n}^{\Psi_{N+1}}(z_{V_n}, \text{per}) = \tilde{Z}{}_{V_n}^{\Psi_N}(z_{V_n}|\text{per}).$$

and

$$Z_{V_n}^{\Psi_{N+1}}(\text{per}) = Z_{V_n}^{\Psi_n}(\text{per}).$$

The previous arguments show that this is sufficient for 12, so the second Lemma is proved.

References

[1] P. Billingsley. *Ergodic Theory and Information.* John Wiley & Sons, New York, 1964.

[2] H.-O. Georgii. *Gibbs Measures and Phase Transitions.* Walter de Gruyter, New York, 1988.

[3] A. Kehagias. *Approximation of Stochastic Processes by Hidden Markov Models.* PhD thesis, Division of Applied Mathematics, Brown University, 1992.

[4] H. Künsch, S. Geman, and A. Kehagias. *Hidden Markov Random Fields.* Technical Report, Division of Applied Mathematics, Brown University, Providence, RI 02912, 1993.