The Annals of Statistics
1997, Vol. 25, No. 1, 293–304

CONSISTENT ESTIMATION OF JOINT DISTRIBUTIONS FOR SUFFICIENTLY MIXING RANDOM FIELDS

BY JEFFREY E. STEIF[1]

Chalmers University of Technology and the Mittag Leffler Institute

The joint distribution of a d-dimensional random field restricted to a box of size k can be estimated by looking at a realization in a box of size $n \gg k$ and computing the empirical distribution. This is done by sliding a box of size k around in the box of size n and computing frequencies. We show that when $k = k(n)$ grows as a function of n, then the total variation distance between this empirical distribution and the true distribution goes to 0 a.s. as $n \to \infty$ provided $k(n)^d \le (\log n^d)/(H + \varepsilon)$ (where H is the entropy of the random field) and providing the random field satisfies a condition called quite weak Bernoulli with exponential rate. This class of processes, studied previously, includes the plus state for the Ising model at a variety of parameter values and certain measures of maximal entropy for certain subshifts of finite type. Marton and Shields have proved such results in one dimension and this paper is an attempt to extend their results to some extent to higher dimensions.

1. Introduction. Let $\{X_m\}_{m \in \mathbf{Z}^d}$ be an ergodic stationary random field taking values in a finite set A. The marginal joint distribution

$$\{X_m\}_{m \in \{(i_1, i_2, \dots, i_d):\, 1 \le i_j \le k\}}$$

can be estimated by a realization

$$\{X_m\}_{m \in \{(i_1, i_2, \dots, i_d):1 \le i_j \le n\}}$$

$(n \gg k)$ by counting frequencies. This means that we look at all translates of $\{(i_1, i_2, \dots, i_d)\colon 1 \le i_j \le k\}$ inside of $\{(i_1, i_2, \dots, i_d)\colon 1 \le i_j \le n\}$ and use these translates together with the realization

$$\{X_m\}_{m \in \{(i_1, i_2, \dots, i_d):\, 1 \le i_j \le n\}}$$

to give us an empirical distribution for

$$\{X_m\}_{m \in \{(i_1, i_2, \dots, i_d):\, 1 \le i_j \le k\}}.$$

(More precise statements will be given later.) For k fixed, the multidimensional ergodic theorem (see [13]) guarantees that with probability 1, as $n \to \infty$, these empirical distributions will converge to the true distribution. The question that interests us is when k also goes to ∞ (but much more

Received July 1995; revised December 1995.
[1] Research supported by grants from the Swedish Natural Science Research Council and from the Royal Swedish Academy of Sciences.
AMS 1991 *subject classifications*. Primary 62M40; secondary 60G60.
Key words and phrases. Estimation, empirical distributions, mixing conditions, random fields.

slowly than n), how slowly must k grow relative to n so that the above estimation scheme works. We briefly recall the results of [11] (see also [12]), where only one dimension was considered. In these papers, it was shown that for any ergodic process in one dimension if $k(n) \geq (\log n)/(H - \varepsilon)$ (where H is the entropy of the process), then such an estimation scheme cannot work. It was also mentioned that given any sequence $k(n) \to \infty$, there is an ergodic process for which this estimation scheme, using $\{k(n)\}$, does not work. The main results of their paper were to show that if $k(n) \leq (\log n)/(H + \varepsilon)$ and the process is nice enough (i.i.d., irreducible Markov, ψ-mixing, weak Bernoulli), then the above estimation scheme works. In this paper, we want to extend, to some degree, these results to higher dimensions.

In the remainder of this introduction, we would like to explain some of the problems that one encounters when going from one to higher dimensions. While there is no reason to give the definitions of ψ-mixing and weak Bernoulli for one dimension here, we mention that these definitions formally extend to higher dimensions. Using these definitions, the results in [11] mentioned above extend as well. However, it was pointed out by Bradley in [2] (see [5] and [6] for further discussion concerning this point) that any process which satisfies this formal generalization of weak Bernoulli or ψ-mixing in dimensions greater than one is in fact finitely dependent, which means that there is some fixed number l such that any two sets of random variables associated to two index sets which are more than l apart are completely independent. Since most interesting random fields (e.g., Ising models in statistical mechanics) will not satisfy this latter condition, these extensions of the definition of weak Bernoulli and ψ-mixing to dimensions greater than 1 are then obviously too strong. (We mention here in passing that a more useful definition of weak Bernoulli involving the notion of a coupling surface is proposed in [6].) We therefore would like to extend the results in [11] to some class which contains interesting and natural random fields. In [5], a certain mixing property, called quite weak Bernoulli with exponential rate, was introduced. It turns out that one can extend the Marton and Shields result to this class (in fact, a slightly larger class) and, on the other hand, this class is general enough that it contains a number of interesting random fields. In the next section, after the definitions are given, we give some examples of random fields which satisfy this property and mention other facts about this property.

While in one dimension, as mentioned above, Marton and Shields can prove their result for any irreducible Markov chain, it seems that the more complicated structure of Markov random fields ($d \geq 2$) does not allow one to carry out (at least as far as the author sees) such an argument for general Markov random fields. Proving such a result for general Markov random fields seems to be an interesting problem and this paper is a small contribution to this problem.

2. Definitions and statements of results. Let $\mathbf{Z}$ denote the integers and let $\mathbf{N}$ denote the positive integers. Let $\{X_m\}_{m \in \mathbf{Z}^d}$ be a stationary ergodic process taking values in a finite set A. Stationarity means that for all sets

$S \subseteq \mathbf{Z}^d$ and all $u \in \mathbf{Z}^d$, the joint distributions of $\{X_m\}_{m \in S}$ and $\{X_m\}_{m \in S+u}$ are the same. Ergodicity means that any event which is invariant under all translations of the lattice has probability 0 or 1. We will usually identify such a process with its distribution which is a probability measure μ on $A^{\mathbf{Z}^d}$ which is invariant under translations of $\mathbf{Z}^d$.

For $u, v \in \mathbf{Z}^d$, we write $u \le v$ if $u_i \le v_i$ for $i = 1, \dots, d$. For $m, n \in \mathbf{Z}^d$, with $m \le n$, let

$$B_m^n = \{x \in \mathbf{Z}^d : m \le x \le n\}.$$

B_m^∞ is used to represent the obvious thing. Given $a \in A^{B_m^n}$ (which we think of as a finite configuration defined on B_m^n taking values in A), we let

$$[a] = \left\{\eta \in A^{\mathbf{Z}^d} : \eta(x) = a(x) \text{ for all } x \in B_m^n\right\}$$

denote the corresponding cylinder set. If $\eta \in A^S$, $m \le n$ and $B_m^n \subseteq S$, we let η_m^n denote the restriction of η to B_m^n.

If k is an integer, we will also use k to represent the vector $(k, k, \dots, k) \in \mathbf{Z}^d$. It will always be clear from context which interpretation is intended. If $k \le n$, $\eta \in A^{B_1^n}$ and $a \in A^{B_1^k}$, let

$$f_k^n(a|\eta) = \left|\left\{i \in B_1^{n-k+1} : \eta_i^{i+k-1} = a\right\}\right|$$

and let

$$\hat{\mu}_k^n(a|\eta) = \frac{f_k^n(a|\eta)}{(n-k+1)^d}.$$

Clearly $\hat{\mu}_k^n(\cdot|\eta)$ is a probability measure on $A^{B_1^k}$, which is the empirical distribution using the configuration $\eta \in A^{B_1^n}$. In words, we look at η in the box B_1^n and look to see how often the different configurations in the box B_1^k occur in η and use these frequencies to define an empirical distribution which hopefully will estimate the true distribution of μ on the box B_1^k. We will denote by μ_k the restriction of μ to the box B_1^k.

The entropy of a stationary process is defined as follows. Letting $H_n(\mu) = -\sum_{a \in A^{B_1^n}} \mu([a]) \log \mu([a])$, we then define the entropy of the process μ, $H(\mu)$, to be $\lim_{n \to \infty} H_n(\mu)/n^d$, which always exists by a subadditivity argument (see [8]). We will write H instead of $H(\mu)$ if the measure μ is clear from context. It is well known that H is at most $\log(|A|)$ with equality holding only in the i.i.d. uniform case. See [16] for a discussion of entropy for one-dimensional processes.

If p and q are probability measures on a measurable set $(\Omega, \mathscr{B})$, the total variation distance between them, which we denote by $\|p - q\|$, is defined by

$$\sup_{E \in \mathscr{B}} |p(E) - q(E)|.$$

DEFINITION 2.1. The nondecreasing sequence $\{(k(n)\}$ $(\le n)$ is admissible for the ergodic process μ if

$$\lim_{n \to \infty} \|\hat{\mu}_{k(n)}^n(\cdot|\eta) - \mu_{k(n)}\| = 0 \quad \mu \text{ a.s.}$$

In words, the sequence $\{k(n)\}$ is admissible for μ if eventually (for large n) one estimates the distribution of the process restricted to the box $B_1^{k(n)}$, $\mu_{k(n)}$, well using the empirical distribution obtained by looking at a realization on the (much larger) box B_1^n. We feel that from an applied point of view, one perhaps would only be interested in that the above holds in probability rather than a.s., but we formulate this stronger result anyway.

We first mention the following easy result whose proof will not be given. We mention that this result follows fairly easily from the multidimensional ergodic theorem.

THEOREM 2.2. *For any ergodic process* μ, *there exists a sequence* $\{k(n)\}$ *which is admissible for* μ.

What one is really after is to find a sequence $\{k(n)\}$ which is universally admissible, that is, admissible for all ergodic processes or perhaps for a large class of ergodic processes. The following negative result is mentioned in [11].

THEOREM 2.3. *Given any sequence* $\{k(n)\}$, *there exists an ergodic process* μ *taking only the values* 0 *and* 1 *for which* $\{k(n)\}$ *is not admissible.*

The next negative result, which is of more interest to us, is identical to a result in the one-dimensional case described in [11]. Although the proof is also identical, we nonetheless feel it worthwhile to give.

THEOREM 2.4. *For any ergodic process* μ, *if* $\varepsilon > 0$ *and if* $k(n)^d \geq (\log n^d)/(H(\mu) - \varepsilon)$, *then* $\{k(n)\}$ *is not admissible for* μ.

PROOF. First note that for every $\eta \in B_1^n$, the probability measure $\hat{\mu}_{k(n)}^n(\cdot|\eta)$ clearly has at most n^d point masses since there are at most this many translates of $B_1^{k(n)}$ in B_1^n. Next, since $n^d \leq \exp((H - \varepsilon)k(n)^d)$, the multidimensional Shannon–McMillan–Breiman theorem (see [13]) implies that for large n, the $\mu_{k(n)}$ probability of any subset of $A^{B_1^{k(n)}}$ with at most $\exp((H - \varepsilon)k(n)^d)$ elements must be arbitrarily close to 0. Since, for all η, $\mu_{k(n)}^n(\cdot|\eta)$ concentrates all its mass on such a set, this implies that the total variation distance between $\mu_{k(n)}^n(\cdot|\eta)$ and $\mu_{k(n)}$ must be close to 1 for all η. $\square$

The point is now to show that if $k(n)^d \leq (\log n^d)/(H + \varepsilon)$ and μ is "nice" in some sense, then $\{k(n)\}$ is admissible for μ. As described earlier in the Introduction, although this was proved in [11] for one-dimensional Markov chains, the structure of higher dimensional Markov random fields is much more complicated and their methods do not go through. It would be reasonable to hope that for any Markov random field μ, any sequence $\{k(n)\}$ satisfying the necessary entropy constraint $k(n)^d \leq (\log n^d)/(H + \varepsilon)$ given in Theorem 2.4 would be admissible for μ, but we cannot prove this. We will, however, show that for certain random fields, to be defined below, which have reasonably strong mixing properties, any sequence $\{k(n)\}$ satisfying the necessary entropy constraint is admissible.

All types of mixing conditions give a type of asymptotic independence and there are many such precise formulations of this. The following formulation has some resemblance to the notion of weak Bernoulli (equivalently, absolute regularity) mentioned (but not defined) earlier in one dimension. For $S \subseteq \mathbf{Z}^d$ and μ a probability measure on $A^{\mathbf{Z}^d}$, we let $\mu|_S$ be the restriction (or projection) of μ to the points in S. We also let Λ_n be $[-n, n]^d \cap \mathbf{Z}^d$. Finally, in the definitions below, $n(1-\alpha)$ will mean $\lfloor n(1-\alpha) \rfloor$, the greatest integer less than or equal to $n(1-\alpha)$.

Definition 2.5. A translation invariant ergodic measure μ on $A^{\mathbf{Z}^d}$ is called *quite weak Bernoulli* (QWB) if for all $\alpha > 0$,

$$\lim_{n \to \infty} \| \mu|_{(\mathbf{Z}^d \setminus \Lambda_n) \cup \Lambda_{n(1-\alpha)}} - \mu|_{\mathbf{Z}^d \setminus \Lambda_n} \times \mu|_{\Lambda_{n(1-\alpha)}} \| = 0.$$

Two stronger versions of the above definition follow:

Definition 2.6. A translation invariant ergodic measure μ on $A^{\mathbf{Z}^d}$ is called *quite weak Bernoulli with summable rate* (QWBS) if for all $\alpha > 0$,

$$\sum_n \| \mu|_{(\mathbf{Z}^d \setminus \Lambda_n) \cup \Lambda_{n(1-\alpha)}} - \mu|_{\mathbf{Z}^d \setminus \Lambda_n} \times \mu|_{\Lambda_{n(1-\alpha)}} \| < \infty.$$

Definition 2.7. A translation invariant ergodic measure μ on $A^{\mathbf{Z}^d}$ is called *quite weak Bernoulli with exponential rate* (QWBE) if for all $\alpha > 0$, there exist constants $\sigma_\alpha > 0$, $c_\alpha > 1$ so that

$$\| \mu|_{(\mathbf{Z}^d \setminus \Lambda_n) \cup \Lambda_{n(1-\alpha)}} - \mu|_{\mathbf{Z}^d \setminus \Lambda_n} \times \mu|_{\Lambda_{n(1-\alpha)}} \| \le c_\alpha \exp(-\sigma_\alpha n)$$

for all n.

Note that QWBE implies QWBS, which in turn implies QWB. Before stating our main theorem, we discuss these properties in some detail. If the process μ satisfies the QWB property, then it is well known that the process is isomorphic (in the sense of ergodic theory) to an independent process (see [15]). This means that there is an invertible measure preserving transformation from μ to some product measure (representing an i.i.d. process) which commutes with all translations. Such processes are called Bernoulli shifts and play a fundamental role in pure ergodic theory.

There are a number of random fields which satisfy the QWBE property. One of these is the plus state for the Ising model at a variety of parameter values. Some of the parameter values where this property has been proved is any d, zero external field and low temperature (see [9]) and any d, zero external field and high temperature (see [9] or combine the methods in [1] and [5]). See [10] for other parameter values. Standard methods together with a percolation result in [7] show that for $d = 2$ and zero external field, temperatures less than the critical temperature belong to this class. Finally, it is proved in [14] that for $d = 2$, we are in this class if there is a zero external field and the temperature is greater than the critical temperature or if there is a nonzero external field and the temperature is arbitrary. The last two results show that in two dimensions, we always have the QWBE property except if we are at the critical point where the external field is zero and the temperature is the critical temperature.

Another collection of measures for which the QWBE property has been established are certain measures of maximal entropy for subshifts of finite type (see [5]). Subshifts of finite type (see [3] for a discussion) are certain dynamical systems which arise in ergodic theory and which have connections to statistical mechanics. Such objects have natural measures associated to them, so-called measures of maximal entropy. In [5], it is proved that for a certain family of subshifts of finite type, all the ergodic measures of maximal entropy satisfy the QWBE property. Within this class, one can find, for any d and k, a subshift of finite type in d dimensions with exactly k ergodic (and hence QWBE by the above) measures of maximal entropy (see [4]).

We also mention that the QWBE property implies a central limit theorem (see [5]) and that the QWB property is one (of a few) possible extensions to higher dimensions of the definition of absolute regularity introduced by Kolmogorov (which is equivalent to the definition of weak Bernoulli).

We now state our main theorem.

THEOREM 2.8. *If μ is an ergodic process which is QWBS, $\varepsilon > 0$, and $\{k(n)\}$ is a nondecreasing sequence approaching ∞ and satisfying $k(n)^d \le (\log n^d)/(H(\mu) + \varepsilon)$ for all n, then $\{k(n)\}$ is admissible for μ.*

It seems that in order to find an admissible sequence, one needs to know the entropy of the process in order to insure that $k(n)^d \le (\log n^d)/(H + \varepsilon)$ for all n. However, since the entropy is always at most $\log|A|$, by taking $k(n)^d \le (\log n^d)/(\log|A| + \varepsilon)$ for all n, the earlier inequality will be assured. In most cases where the above theorem can be applied, one probably has the stronger QWBE property. However (after the suggestion of a referee), we formulate our result in the more general setting above so that one sees precisely what is needed. The final section of this paper is devoted to the proof of Theorem 2.8.

3. Proof of main result. We assume that μ is an ergodic stationary process which is QWBS and that $\{k(n)\}$ is a nondecreasing sequence approaching ∞ and satisfying $k(n)^d \le (\log n^d)/(H + \varepsilon)$ for all n. Most of the argument of one of the results in [11] (see also [12]) can be extended. One of the differences is that, at some point in the argument, we will need to partition the lattice into boxes where the space between the boxes grows with the size of the box (as opposed to in one dimension, where they do not). Another difference is that there is a martingale argument in the one-dimensional case which one cannot carry out, but the summable rate in our mixing assumption will circumvent this difficulty.

Let $k \in \mathbf{N}$, $\alpha > 0$ and $r \in B_1^{k+\alpha k}$. (In the rest of the paper, αk will always mean $\lfloor \alpha k \rfloor$, the greatest integer less than or equal to αk.) For $j \in \mathbf{Z}^d$, denote

$B_{r+(j-1)(k+\alpha k)}^{r+(j-1)(k+\alpha k)+k-1}$ by $V(j,r,k,\alpha)$. If $\eta \in A^{\mathbf{Z}^d}$, we let $\eta_j(r,k,\alpha)$ be the restriction of η to $V(j,r,k,\alpha)$, and let $[\eta_j(r,k,\alpha)]$ be the corresponding cylinder set. We note that for different j's, the boxes $V(j,r,k,\alpha)$ are separated by distance αk. This is different from the one-dimensional case in that this separation grows with k.

We now need to place a linear order $\prec^d$ on $\mathbf{Z}^d$. (This is the usual ordering used in ergodic theory for higher dimensional processes.) This ordering will be defined by induction on d and can be thought of as a backwards dictionary ordering. For $d = 1$, we use the natural linear order $<$. For $d \geq 2$, we define $\prec^d$ by induction via

$$(x_1,\dots,x_d) \prec^d (y_1,\dots,y_d)$$

if and only if

$$x_d < y_d \quad \text{or} \quad x_d = y_d \text{ and } (x_1,\dots,x_{d-1}) \prec^{d-1} (y_1,\dots,y_{d-1}).$$

If $U \subseteq \mathbf{Z}^d$, let $\overline{U} = U \cup \{x \in \mathbf{Z}^d\colon\ x \prec^d u \text{ for some } u \in U\}$. Next, let $\mathscr{F}(j,r,k,\alpha)$ be the sub-σ-algebra $\sigma\big(X_i\colon i \in \overline{\bigcup_{j' \prec^d j} V(j',r,k,\alpha)}\big)$. If $\gamma > 0$, we say that $j \in B_1^\infty$ is a (γ,r,k,α) splitting index for $\eta \in A^{\mathbf{Z}^d}$ if

$$\mu\big(\big[\eta_j(r,k,\alpha)\big] | \mathscr{F}(j,r,k,\alpha)\big) \leq (1+\gamma)\mu\big(\big[\eta_j(r,k,\alpha)\big]\big) \quad \mu \text{ a.s.}$$

We let $B_j(\gamma,r,k,\alpha) = \{\eta \in A^{\mathbf{Z}^d}\colon j \text{ is a } (\gamma,r,k,\alpha) \text{ splitting index for } \eta\}$. Note that for any $\tilde{j}$ with $j \prec^d \tilde{j}$, $B_j(\gamma,r,k,\alpha)$ is measurable with respect to $\mathscr{F}(\tilde{j},r,k,\alpha)$.

Our first lemma is completely analogous to Lemma 6A in [12].

LEMMA 3.1. *Fix* $\gamma > 0$, $k \in \mathbf{N}$, $\alpha > 0$ *and* $r \in B_1^{k+\alpha k}$ *and let* J *be a finite set contained in* B_1^∞. *Then given* $\eta_j(r,k,\alpha) \in A^{V(j,r,k,\alpha)}$ *for* $j \in J$, *we have*

$$\mu\bigg(\bigcap_{j \in J}\big(\big[\eta_j(r,k,\alpha)\big] \cap B_j(\gamma,r,k,\alpha)\big)\bigg) \leq (1+\gamma)^{|J|}\prod_{j \in J}\mu\big(\big[\eta_j(r,k,\alpha)\big]\big).$$

PROOF. This is proved as in the one-dimensional case, Lemma 6A in [12], the only difference being that to start off the induction, we let j be the maximum element in J with respect to the linear ordering $\prec^d$. □

Given n, k and α, let $t = t(n,k,\alpha)$ be the largest integer t satisfying $t(k+\alpha k) + k - 1 \leq n$.

LEMMA 3.2. *If* $0 < \gamma < 1/2$ *and* $\alpha > 0$, *then there exist integers* $k(\gamma,\alpha)$ *and* $t(\gamma,\alpha)$ *and a sequence of measurable sets* $\{G_n(\gamma,\alpha)\}$ *such that*:

(i) $\mu(G_n(\gamma,\alpha)$ *eventually*$) = 1$.

(ii) *If* $k \geq k(\gamma,\alpha)$, $t(n,k,\alpha) \geq t(\gamma,\alpha)$ *and* $\eta \in G_n(\gamma,\alpha)$, *then there are at least* $(1-\gamma)(k+\alpha k)^d$ *values of* $r \in B_1^{k+\alpha k}$ *for which there are at least* $(1-\gamma)t(n,k,\alpha)^d$ *indices* $j \in B_1^{t(n,k,\alpha)}$ *that are a* (γ,r,k,α) *splitting index for* η.

Note: $G_n(\gamma,\alpha)$ eventually refers as usual to the event $\bigcup_{m=1}^\infty \bigcap_{n=m}^\infty G_n(\gamma,\alpha)$.

PROOF. It is easy to show that

$$E\left|1-\frac{\mu(\eta_1^n)}{\mu(\eta_1^n|\mathscr{F}(1,1,n,\alpha))}\right| \le 2\|\mu|_{(\mathbf{Z}^d\setminus\Lambda_n)\cup\Lambda_{n(1-\alpha)}} - \mu|_{\mathbf{Z}^d\setminus\Lambda_n}\times\mu|_{\Lambda_{n(1-\alpha)}}\|.$$

Denoting $\|\mu|_{(\mathbf{Z}^d\setminus\Lambda_n)\cup\Lambda_{n(1-\alpha)}} - \mu|_{\mathbf{Z}^d\setminus\Lambda_n}\times\mu|_{\Lambda_{n(1-\alpha)}}\|$ by $q_{\alpha,n}$, Markov's inequality gives

$$\mu\left(\left|1-\frac{\mu(\eta_1^n)}{\mu(\eta_1^n|\mathscr{F}(1,1,n,\alpha))}\right| \ge \frac{\gamma^2}{2}\right) \le \frac{2q_{\alpha,n}}{\gamma^2/2} = \frac{4q_{\alpha,n}}{\gamma^2}.$$

Using the fact that μ is QWBS, choose $k(\gamma,\alpha)$ so that

$$\frac{1}{\gamma^2}\sum_{l=k(\gamma,\alpha)}^{\infty} 4q_{\alpha,l} < \frac{\gamma^2}{2}.$$

This gives

$$\mu\left(\left|1-\frac{\mu(\eta_1^l)}{\mu(\eta_1^l|\mathscr{F}(1,1,l,\alpha))}\right| < \frac{\gamma^2}{2}\ \forall l \ge k(\gamma,\alpha)\right) > 1-\frac{\gamma^2}{2}.$$

Let C denote the event

$$\left\{\left|1-\frac{\mu(\eta_1^l)}{\mu(\eta_1^l|\mathscr{F}(1,1,l,\alpha))}\right| < \frac{\gamma^2}{2}\ \forall l \ge k(\gamma,\alpha)\right\}.$$

The multidimensional ergodic theorem applied to the indicator function I_C, tells us that if we let $G_n(\gamma,\alpha)$ be

$$\left\{\eta\colon \frac{1}{n^d}\sum_{i\in B_1^n} I_C(T^{i-1}\eta) > 1-\frac{\gamma^2}{2}\right\},$$

then (i) holds. [Here T^i is the transformation on $A^{\mathbf{Z}^d}$ which shifts configurations by the vector i, that is, $T^i(\eta)(x) = \eta(x-i)$.]

To proceed with (ii), we first define $t(\gamma,\alpha) = \lceil 2/\gamma^2\rceil$. Now assume that $k \ge k(\gamma,\alpha)$, $t(n,k,\alpha) \ge t(\gamma,\alpha)$ and $\eta \in G_n(\gamma,\alpha)$. Write t for $t(n,k,\alpha)$. Since $t \ge \lceil 2/\gamma^2\rceil$, we then have

$$\frac{1}{t^d(k+\alpha k)^d}\sum_{i\in B_1^{t(k+\alpha k)}} I_C(T^{i-1}(\eta)) > 1-\gamma^2$$

(the idea being that since t is large, this average cannot differ much from the earlier one), which implies

$$\frac{1}{(k+\alpha k)^d}\sum_{r\in B_1^{k+\alpha k}}\frac{1}{t^d}\sum_{j\in B_1^t} I_C(T^{r+(j-1)(k+\alpha k)-1}(\eta)) > 1-\gamma^2.$$

This implies that there exists $R \subseteq B_1^{(k+\alpha k)}$ such that $|R| \ge (1-\gamma)(k+\alpha k)^d$ and for all $r \in R$,

$$\frac{1}{t^d}\sum_{j\in B_1^t} I_C(T^{r+(j-1)(k+\alpha k)-1}(\eta)) > 1-\gamma.$$

Next, if $r \in R$, there exists $J \subseteq B_1^t$ such that $|J| \geq (1-\gamma)t^d$ and for all $j \in J$, $T^{r+(j-1)(k+\alpha k)-1}(\eta) \in C$. The proof will be completed if we can show that each such j is a (γ, r, k, α) splitting index for η. Since $k \geq k(\gamma, \alpha)$, the above implies that for such j,

$$\left|1 - \frac{\mu\left(\eta_{r+(j-1)(k+\alpha k)}^{r+(j-1)(k+\alpha k)+k-1}\right)}{\mu\left(\eta_{r+(j-1)(k+\alpha k)}^{r+(j-1)(k+\alpha k)+k-1} | \mathscr{F}(j, r, k, \alpha)\right)}\right| < \frac{\gamma^2}{2}.$$

Multiplying by the denominator $\mu(\eta_{r+(j-1)(k+\alpha k)}^{r+(j-1)(k+\alpha k)+k-1} | \mathscr{F}(j, r, k, \alpha))$, simple algebra and using the fact that $1/(1-\gamma^2/2) \leq 1+\gamma$ show that

$$\mu\left(\left[\eta_j(r, k, \alpha)\right] | \mathscr{F}(j, r, k, \alpha)\right) \leq (1+\gamma)\mu\left(\left[\eta_j(r, k, \alpha)\right]\right)$$

showing that such a j is a (γ, r, k, α) splitting index for η as desired. □

We will now replace our empirical distribution by nonoverlapping averages in the following way. For $\eta \in A^{B_1^n}$, $a \in A^{B_1^k}$, $r \in B_1^{k+\alpha k}$ and $J \subseteq B_1^{t(n,k,\alpha)}$, let

$$f_k^{r,\alpha,n,J}(a|\eta) = \left|\left\{j \in J : \eta_{r+(j-1)(k+\alpha k)}^{r+(j-1)(k+\alpha k)+k-1} = a\right\}\right|$$

and let

$$\hat{\mu}_k^{r,\alpha,n,J}(a|\eta) = \frac{f_k^{r,\alpha,n,J}(a|\eta)}{|J|}.$$

In words, we partition B_r^n into blocks of size k separated by distance αk and consider the empirical distribution over the boxes $\{V(j, r, k, \alpha)\}_{j \in J}$ of size k which are nonoverlapping, separated from each other by distance $\geq \alpha k$ and all contained in B_1^n. The following lemma is clear and the proof is left to the reader.

LEMMA 3.3. *For all* $\delta > 0$, *there exists* $\gamma \in (0, 1/2)$ *and* $\alpha > 0$ *so that if* $k/n < \gamma$ *and* $\|\hat{\mu}_k^n(\cdot|\eta) - \mu_k\| \geq \delta$, *then it follows that for at least* $2\gamma(k+\alpha k)^d$ *indices* $r \in B_1^{k+\alpha k}$, *we have that for all* $J \subseteq B_1^{t(n,k,\alpha)}$ *with* $|J| \geq (1-\gamma)t(n,k,\alpha)^d$, $\|\hat{\mu}_k^{r,\alpha,n,J}(\cdot|\eta) - \mu_k\| \geq \delta/2$.

To prove that $\{k(n)\}$ is admissible for μ, it suffices to show that $\forall \delta > 0$,

$$\mu\left(\eta : \|\hat{\mu}_{k(n)}^n(\cdot|\eta) - \mu_{k(n)}\| \geq \delta \text{ i.o.}\right) = 0, \tag{3.1}$$

where i.o. stands for infinitely often.

Now given $\delta > 0$, Lemma 3.3 gives us a γ and α. We also assume that γ is so small that

$$\sum_n 2k(n)\exp\left(-c\gamma(\log\gamma)\frac{(2n)^d}{k(n)^d}\right)(1+\gamma)^{(2n)^d/(k(n)^d)} \times (n^d)^{(n^d)^{(H+\varepsilon/2)/(H+\varepsilon)}} 2^{-n^d/(2^{d+1}k(n)^d)C(\delta^2/100)} < \infty,$$

where c and C are two universal constants to be given later. Given our γ and α, we then apply Lemma 3.2 to obtain integers $k(\gamma, \alpha)$ and $t(\gamma, \alpha)$ and a sequence of measurable sets $\{G_n(\gamma, \alpha)\}$ such that conditions (i) and (ii) of Lemma 3.2 hold.

Since $\mu(G_n(\gamma, \alpha)$ eventually$) = 1$ by (i), in order to prove (3.1), it suffices by Borel–Cantelli to show that

$$\sum_n \mu\Big(\big\{\eta\colon \|\hat{\mu}^n_{k(n)}(\cdot|\eta) - \mu_{k(n)}\| \geq \delta\big\} \cap G_n(\gamma, \alpha)\Big) < \infty. \tag{3.2}$$

Since $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$, Lemmas 3.2 and 3.3 imply that for large n, if

$$\eta \in \big\{\eta\colon \|\hat{\mu}^n_{k(n)}(\cdot|\eta) - \mu_{k(n)}\| \geq \delta\big\} \cap G_n(\gamma, \alpha),$$

then there exists $r \in B_1^{k(n)+\alpha k(n)}$ such that $\|\hat{\mu}^{r,\alpha,n,J}_{k(n)}(\cdot|\eta) - \mu_{k(n)}\| \geq \delta/2$ for all $J \subseteq B_1^{t(n,k(n),\alpha)}$ with $|J| \geq (1-\gamma)t(n, k(n), \alpha)^d$ and such that there are at least $(1-\gamma)t(n, k(n), \alpha)^d$ indices $j \in B_1^{t(n,k(n),\alpha)}$ which are a $(\gamma, r, k(n), \alpha)$ splitting index for η. Let $\mathscr{J}$ be $\{J \subseteq B_1^{t(n,k(n),\alpha)}\colon |J| \geq (1-\gamma)t(n, k(n), \alpha)^d\}$. This implies that for large n,

$$\begin{aligned}&\big\{\eta\colon \|\hat{\mu}^n_{k(n)}(\cdot|\eta) - \mu_{k(n)}| \geq \delta\big\} \cap G_n(\gamma, \alpha)\\ &\quad\subseteq \bigcup_{r=1}^{k(n)+\alpha k(n)} \bigcup_{J\in\mathscr{J}} \left\{\eta\colon \|\hat{\mu}^{r,\alpha,n,J}_{k(n)}(\cdot|\eta) - \mu_{k(n)}\| \geq \frac{\delta}{2}\right\} \cap D_n(r, J),\end{aligned}$$

where

$$D_n(r, J) = \{\eta\colon j \text{ is a } (\gamma, r, k(n), \alpha) \text{ splitting index for } \eta\ \forall j \in J\}.$$

Fixing $r \in B_1^{k(n)+\alpha k(n)}$ and $J \in \mathscr{J}$, note that $D_n(r, J) = \bigcap_{j\in J} B_j(\gamma, r, k(n), \alpha)$ and therefore Lemma 3.1 gives us that if $\eta_j(r, k(n), \alpha) \in A^{V(j,r,k(n),\alpha)}$ for $j \in J$, then

$$\begin{aligned}&\mu\Big(\bigcap_{j\in J}\big([\eta_j(r, k(n), \alpha)]\big) \cap D_n(r, J)\Big)\\ &\qquad\leq (1+\gamma)^{|J|} \prod_{j\in J} \mu\big([\eta_j(r, k(n), \alpha)]\big).\end{aligned} \tag{3.3}$$

We will now use a large deviation result for i.i.d. random variables. Let $Y_1, Y_2, \dots$ be i.i.d. random variables taking values in $A^{B_1^k}$ and having distribution μ_k. Let $\hat{\mu}^n_k(\cdot|Y_1, Y_2, \dots)$ be the empirical distribution on $A^{B_1^k}$ given the Y_i's. As before, this means that for $\eta \in A^{B_1^k}$, $\hat{\mu}^n_k(\eta|Y_1, Y_2, \dots) = (1/n)\sum_{i=1}^n I_{\{Y_i = \eta\}}$. Letting

$$W_k = \left\{\eta \in A^{B_1^k}\colon \exp\left(-\left(H + \frac{\varepsilon}{2}\right)k^d\right) \leq \mu_k(\eta) \leq \exp\left(-\left(H - \frac{\varepsilon}{2}\right)k^d\right)\right\},$$

Lemma 2 in [11] implies that

$$\left(\|\hat{\mu}^n_k(\cdot|Y_1, Y_2, \dots) - \mu_k\| \geq \frac{\delta}{2}\right) \leq (n+1)^{|W_k|} 2^{-nC(\delta^2/100)}, \tag{3.4}$$

where C is a universal constant provided that $\mu_k(W_k) \geq 1 - \delta/10$, which in turn holds for large k by the multidimensional Shannon–McMillan–Breiman theorem.

We use this to bound the probability of

$$\left\{\eta\colon \|\hat{\mu}_{k(n)}^{r,\alpha,n,J}(\cdot|\eta) - \mu_{k(n)}\| \geq \frac{\delta}{2}\right\} \cap D_n(r, J).$$

Equations (3.3) and (3.4) imply that this probability is bounded by

$$(1+\gamma)^{|J|}(|J|+1)^{|W_{k(n)}|}2^{-|J|C(\delta^2/100)}.$$

The multidimensional Shannon–McMillan–Breiman theorem implies that $|W_{k(n)}| \leq \exp((H+\varepsilon/2)k(n)^d)$ for large n. Also, letting t denote $t(n,k,\alpha)$ we have $t^d/2 \leq |J| \leq t^d$ and $n/2k(n) \leq t \leq 2n/k(n)$ for large n. Since $k(n)^d \leq (\log n^d)/(H+\varepsilon)$, this gives us that the above is bounded by

$$\begin{aligned}(1+\gamma)^{(2n)^d/(k(n)^d)}&(n^d)^{\exp((H+\varepsilon/2)\log(n^d)/H+\varepsilon)}2^{-(n^d/(2^{d+1}k(n)^d))C(\delta^2/100)}\\ &\leq (1+\gamma)^{(2n)^d/(k(n)^d)}(n^d)^{(n^d)^{(H+\varepsilon/2)/(H+\varepsilon)}}2^{-(n^d/(2^{d+1}k(n)^d))C(\delta^2/100)}.\end{aligned}$$

Using the fact that the number of subsets of $\{1,\dots,t^d\}$ with at least $(1-\gamma)t^d$ elements is a most $\exp(-c\gamma(\log\gamma)t^d)$ for some universal constant c, this gives

$$\begin{aligned}\mu\Big(\Big\{\eta\colon &\|\hat{\mu}_k^n(n)(\cdot|\eta) - \mu_{k(n)}\| \geq \delta\Big\} \cap G_n(\gamma,\alpha)\Big)\\ &\leq 2k(n)\exp\left(-c\gamma(\log\gamma)\frac{(2n)^d}{k(n)^d}\right)(1+\gamma)^{(2n)^d/(k(n)^d)}\\ &\quad\times(n^d)^{(n^d)^{(H+\varepsilon/2)/(H+\varepsilon)}}2^{-(n^d/(2^{d+1}k(n)^d))C(\delta^2/100)}.\end{aligned}$$

We note that this is summable in n by the way γ was chosen. This completes the proof. □

Acknowledgment. The author would like to thank a referee for asking whether it was necessary to assume the QWBE property in order to prove the the main theorem (this was the formulation in the first version of this paper). The author would also like to thank Roberto Schonmann for providing some of the Ising model references in this paper.

REFERENCES

[1] VAN DEN BERG, J. and MAES, C. (1994). Disagreement percolation in the study of Markov fields. *Ann. Probab.* **22** 749–763.

[2] BRADLEY, R. (1989). A caution on mixing conditions for random fields. *Statist. Probab. Lett.* **8** 489–491.

[3] BURTON, R. and STEIF, J. E. (1994). Nonuniqueness of measures of maximal entropy for subshifts of finite type. *Ergodic Theory Dynamic Systems* **14** 213–235.

[4] BURTON, R. and STEIF, J. E. (1995). New results on measures of maximal entropy. *Israel J. Math.* **89** 275–300.

[5] Burton, R. and Steif, J. E. (1995). Quite weak Bernoulli with exponential rate and percolation for random fields. *Stochastic Process. Appl.* **58** 35–55.
[6] Burton, R. and Steif, J. E. (1996). Coupling surfaces and weak Bernoulli in one and higher dimensions. Unpublished manuscript.
[7] Chayes, J., Chayes, L. and Schonmann, R. H. (1987). Exponential decay of connectivities in the two-dimensional Ising model. *J. Statist. Phys.* **49** 433–445.
[8] Georgii, H. (1988). *Gibbs Measures and Phase Transitions*. de Gruyter, New York.
[9] di Liberto, F., Gallavotti, G. and Russo, L. (1973). Markov processes, Bernoulli schemes, and Ising model. *Comm. Math. Phys.* **33** 259–282.
[10] Martinelli, F. and Olivieri, E. (1994). Approach to equilibrium of Glauber dynamics in the one phase region I. The attractive case. *Comm. Math. Phys.* **161** 447–486.
[11] Marton, K. and Shields, P. (1994). Entropy and the consistent estimation of joint distributions. *Ann. Probab.* **22** 960–977.
[12] Marton, K. and Shields, P. (1996). Correction to "Entropy and the consistent estimation of joint distributions." *Ann. Probab.* **24** 541–545.
[13] Ornstein, D. S. and Weiss, B. (1983). The Shannon–McMillan–Breiman theorem for a class of amenable groups. *Israel J. Math.* **44** 53–60.
[14] Schonmann, R. H. and Shlosman, S. B. (1995). Complete analyticity for 2D Ising completed. *Comm. Math. Phys.* **170** 453–482.
[15] Steif, J. E. (1991). Space–time Bernoullicity of the lower and upper stationary processes for attractive spin systems. *Ann. Probab.* **19** 609–635.
[16] Walters, P. (1975). *An Introduction to Ergodic Theory*. Springer, New York.

Department of Mathematics
Chalmers University of Technology
S-412 96 Gothenburg
Sweden
E-mail: steif@math.chalmers.se