Z. Wahrscheinlichkeitstheorie verw. Geb. 26, 207 – 217 (1973)

On Entropy and Information Gain in Random Fields

Hans Föllmer

0. Introduction

By a *random field* we mean a stochastic process $(X_t)_{t\in T}$ which assumes values in some finite set S and whose parameter set T is the d-dimensional lattice. Alternatively we may say that a random field is a probability measure μ on S^T. A random field is called a *Markov field* if the conditional distribution of X_t, given the states $(X_s)_{s\neq t}$, depends only on the nearest neighbor states $(X_s)_{\|s-t\|=1}$. For $d=1$ stationary Markov fields reduce to Markov chains (cf. [11]).

For the classical case $d=1$ Spitzer has shown that, among stationary random fields, a Markov field is characterized by the fact that it minimizes a suitable *free energy* (cf. [12]). His method does not seem to carry over to higher dimensions. The purpose of this paper is to present a different approach which does two things: (a) it works for arbitrary dimension $d\geqq 1$ and (b) it shows that the variational characterization of the Markov property is intimately related to the theorem of McMillan in the sense that both can be based on essentially the same argument.

The key is formula (2.5) for the entropy of a stationary random field. Its proof yields the d-dimensional version of McMillan's theorem, first published by Thouvenot in [13], as an aside. In Section 3 we introduce the information gain of a random field μ with respect to a random field ν. Here (2.5) implies that the above conditional distributions are the same for μ and ν as soon as the information gain is zero. Section 4 explains why this amounts to a variational principle.

Markov fields are known to be examples of *Gibbs fields* in the sense of Dobrušin [3] and Lanford and Ruelle [8] where the conditional distributions above are given in terms of a potential function (cf. [1, 11, 2]). Thus Spitzer's characterization of the Markov property is contained in the variational principle for Gibbs fields due to Lanford and Ruelle [8]. Our translation of the information theoretic result of Section 3 into the *thermodynamics* of Section 4 is made such that it covers Lanford and Ruelle's result. As a by-product we obtain, in the case of Gibbs fields, the d-dimensional version of Breiman's theorem that there is almost sure convergence behind the theorem of McMillan.

1. Definitions

Let S be a finite set. We set

$$T=\{t=(t_1,\ldots,t_d)\,|\,t_i \text{ integer}\}$$

and denote by Ω the set of all maps $\omega\colon T\to S$. For $V\subseteq T$ let $\mathscr{F}_V$ be the σ-field generated by the projections $\omega\to\omega(t)$ $(t\in V)$; $\mathscr{F}_\emptyset$ is meant to be the trivial σ-field $\{\emptyset,\Omega\}$.

(1.1) *Definition.* A *random field* is a probability measure μ on $(\Omega, \mathscr{F}_T)$.

Let μ be a random field and let us use the notation

$$\mu_t(a|V)(.)=\mu(\omega(t)=a|\mathscr{F}_V)(.)$$

for conditional probabilities.

(1.2) *Definition.* The conditional probabilities

$$\mu_t(.|T-\{t\}) \qquad (t\in T)$$

are called the *local characteristics* of the random field μ. If the local characteristics are positive and depend only on nearest neighbors, i.e. if

$$\mu_t(.|T-\{t\})=\mu_t(.|N(t))>0 \qquad (t\in T) \tag{1.3}$$

where $N(t)=\{s\in T|\|s-t\|=1\}$, then μ is called a *Markov field.* Replacing $N(t)$ by $N_r(t)=\{s\in T|0<\|s-t\|\leqq r\}$ we obtain the notion of a *Markov field of order r.*

For each lattice point $t\in T$ we denote by θ_t the corresponding *shift*, i.e. the map on Ω defined through

$$[\theta_t(\omega)](s)=\omega(t+s) \qquad (s\in T).$$

The random field μ is called *stationary* if it is invariant under all shifts: $\mu\theta_t=\mu$ $(t\in T)$. In this case the local characteristics of μ are stationary in the sense that $\mu_0(a|T-\{0\})\circ\theta_t^{-1}$ is a version of $\mu_t(a|T-\{t\})$ for any $t\in T$.

(1.4) *Remark.* It is well known that a random field is in general not determined by its local characteristics (*possibility of phase transition*), and that stationarity of the local characteristics does not imply that the random field is stationary (*possibility of symmetry breakdown*). Cf. [11].

2. Specific Entropy

Let μ be a random field. For any $V\subseteq T$ denote by μ_V the restriction of μ to $\mathscr{F}_V$. If V is finite then $\mathscr{F}_V$ is finite and we can define the *entropy* of μ_V as

$$H(\mu_V)=-E_\mu[\log\mu(\omega_V)] \tag{2.1}$$

where E_μ denotes expectation with respect to μ and ω_V is the restriction of ω to V (which can be identified with an atom of $\mathscr{F}_V$). It is well known that $\frac{1}{|V|}H(\mu_V)$ converges to some limit in the interval $[0, H(\mu_{\{0\}})]$, the *specific entropy* $h(\mu)$ of the random field μ, if μ is stationary and if V expands to infinity in a suitable way ($|V|$ is the cardinality of V). The aim of this section is to find a convenient formula for that limit. Its existence as well as the d-dimensional version of McMillan's theorem will be by-products of the argument below.

For $k=1, 2, \ldots, \infty$ let T_k be the set of all lattice points $t=(t_1, \ldots, t_d)\in T$ with $t_i=n_i k$ for some integer n_i and define

$$T_k^*=\{s\in T|s<0 \text{ or } s\notin T_k\} \tag{2.2}$$

where $<$ denotes the lexicographical ordering on T.

(2.3) *Examples.* 1) $T_1 = T$, and T_1^* is the *past* $\{s \in T \mid s < 0\}$.

2) T_2 is the set of *even* lattice points, and T_2^* is the union of the *past* and the *uneven* part of the *future* $\{s \in T \mid s > 0\}$.

3) T_∞ is meant to be $\{0\}$ so that $T_\infty^* = T - \{0\} = \bigcup_{k=1}^{\infty} T_k^*$.

For $a = (a_1, \ldots, a_d) \in T$ we set

$$V(a) = \{t \in T \mid 0 \leqq t_i < a_i\}$$

and

$$V_k(a) = V(a) \cap T_k, \qquad V_k^*(a) = V(a) \cap T_k^* = V(a) - V_k(a).$$

We write $a \to \infty$ if $\min_{1 \leqq i \leqq d} a_i \to \infty$.

(2.4) **Theorem.** *Let μ be stationary. Then the limits*

$$h(\mu) = \lim_{a \to \infty} \frac{1}{|V(a)|} H(\mu_{V(a)}), \qquad h_k^*(\mu) = \lim_{a \to \infty} \frac{1}{|V_k^*(a)|} H(\mu_{V_k^*(a)})$$

exist and satisfy

$$h(\mu) = (1 - k^{-d})\, h_k^*(\mu) + k^{-d}\, E_\mu[H(\mu_0(.\mid T_k^*))] \tag{2.5}$$

where $H(\mu_0(.\mid T_k^*))(\omega)$ denotes the entropy of the probability measure $\mu_0(.\mid T_k^*)(\omega)$ on S.

(2.6) *Remarks.* 1) We may write

$$\begin{aligned} E_\mu[H(\mu_0(.\mid T_k^*))] &= -E_\mu\Big[\sum_{a \in S} \mu_0(a \mid T_k^*)(\omega) \log \mu_0(a \mid T_k^*)(\omega)\Big] \\ &= -E_\mu[\log \mu_0(\omega(0) \mid T_k^*)(\omega)]. \end{aligned}$$

2) For $k = 1$ we obtain

$$h(\mu) = E_\mu[H(\mu_0(.\mid T_1^*))]. \tag{2.7}$$

Part 2) of the proof shows that there is a L^1-convergence behind the existence of specific entropy:

$$-\frac{1}{|V(a)|} \log \mu(\omega_{V(a)}) \to E_\mu[H(\mu_0(.\mid T_1^*)) \mid \mathscr{J}] \tag{2.8}$$

in L_μ^1, where $E_\mu[.\mid \mathscr{J}]$ denotes conditional expectation with respect to the σ-field $\mathscr{J}$ of stationary sets (all $A \in \mathscr{F}_T$ with $\theta_t^{-1}(A) = A$ $(t \in T)$). This is the d-dimensional version of McMillan's theorem (cf. [13]). For a d-dimensional version of Breiman's theorem on almost sure convergence see (4.28) below.

Proof. 1) For $V = V(a)$ and $\omega \in \Omega$ we write

$$\mu(\omega_V) = \mu(\omega_{V_k}) \prod_{t \in V_k} [\mu_0(\omega(0) \mid V_t)(\omega)] \circ \theta_t^{-1}$$

where $V_t = (V_k^* \cup \{s \in V_k | s < t\}) - t \subseteq T_k^*$ and where we identify ω_W with an atom of $\mathscr{F}_W$ ($\omega_\emptyset$ is identified with Ω). Thus

$$\frac{1}{|V|} \log \mu(\omega_V) = \frac{1}{|V|} \log \mu(\omega_{V_k^*}) + \frac{1}{|V|} \sum_{t \in V_k} [\log \mu_0(\omega(0)|V_t)(\omega)] \circ \theta_t^{-1}. \tag{2.9}$$

2) Part 5) below implies that for any $\varepsilon > 0$ there is an integer $M > 0$ such that

$$\|\log \mu_0(\omega(0)|V_t)(.) - \log \mu_0(\omega(0)|T_k^*)\|_\mu < \varepsilon \tag{2.10}$$

as soon as $V_t \supseteq T_k^* \cap N_M(0)$ ($\|\cdot\|_\mu$ denotes the L_μ^1-norm). Thus

$$\left\| \frac{1}{|V_k|} \sum_{t \in V_k} [\log \mu_0(\omega(0)|V_t) - \log \mu_0(\omega(0)|T_k^*)] \circ \theta_t^{-1} \right\|_\mu \leqq \varepsilon + \frac{|V_k'|}{|V_k|} 2c(\mu) \tag{2.11}$$

where $|V_k'|$ is the number of those $t \in V_k$ which have distance not greater than M from $T - V_k$, and where

$$c(\mu) = \sup_{W_k \subseteq T_k^*} \|\log \mu_0(\omega(0)|W)(.)\|_\mu < \infty \tag{2.12}$$

again by 5) below. Note that $|V_k'(a)|\,|V_k(a)|^{-1} \to 0$ for $a \to \infty$.

3) Consider the case $k=1$ where $V_k(a) = V(a)$ and let us discuss L^1-convergence of the left side in (2.9) as $a \to \infty$. The first term on the right vanishes, and so part 2) shows that the question is reduced to the L^1-convergence of

$$\frac{1}{|V(a)|} \sum_{t \in V(a)} [\log \mu_0(\omega(0)|T_1^*)(\omega)] \circ \theta_t^{-1}.$$

But here we can apply the d-dimensional ergodic theorem ([5], VIII, 6.9) and this yields (2.8). Taking expectations we obtain (2.7) and in particular the existence of $h(\mu)$.

4) Let us return to the general case and let us discuss convergence of the expectation of the right side in (2.9) as $a \to \infty$. Since we have convergence on the left side by part 3), it is enough to consider the second term on the right. But part 2) shows that, in the limit, we may replace its expectation by

$$\frac{1}{|V|} E_\mu\Big[\sum_{t \in V_k} \log \mu_0(\omega(0)|T_k^*) \circ \theta_t^{-1}\Big] = \frac{|V_k|}{|V|} E_\mu[\log \mu_0(\omega(0)|T_k^*)],$$

and this converges to $-k^{-d} E_\mu[H(\mu_0(.|T_k^*))]$. Thus we have (2.5) and in particular the existence of $h_k^*(\mu)$.

5) Here we recall, in a form which is suitable for the application above, a familiar step in the classical proof of McMillan's theorem. Let $\mathscr{W}$ be a collection of finite subsets $\mathscr{W} \subseteq T - \{0\}$ filtering to the right with respect to inclusion. Then

$$\log \mu_0(\omega(0)|W)(\omega) \to \log \mu_0(\omega(0)| \bigcup_{W \in \mathscr{W}} W)$$

in L^p_μ for each $p \geqq 1$. To prove this, note that the functions

$$g_W(\omega) = -\log \mu_0(\omega(0)|W)(\omega) \qquad (W \in \mathscr{W})$$

form a submartingale $\geqq 0$ with respect to the σ-fields $\mathscr{G}_W = \mathscr{F}_{W \cup \{0\}}$. Moreover, (g_W^p) is uniformly integrable for each $p \geqq 1$:

$$E_\mu[g_W^p; n \leqq g_W^p < n+1] \leqq |S|(n+1)e^{-n^{1/p}}$$

(same proof as in [7], p. 311) which is the tail of a convergent series. Thus (g_W) converges in L^p to some limit g (same argument as in [9], V.20). In order to identify the limit note that $\mu_0(\omega(0)|W)(\omega) \to \mu_0(\omega(0)|\bigcup W)$ in L^1_μ ([9], V.20) and use almost sure convergence along increasing sequences (W_n) plus continuity of $\log(.)$ to conclude $g(\omega) = -\log \mu_0(\omega(0)|\bigcup W)(\omega)$ μ-a.s.

3. Specific Information Gain

Throughout this section μ and ν will be two stationary random fields. For any finite $V \subseteq T$ we may define the *information gain* of μ_V with respect to ν_V as

$$H(\mu_V|\nu_V) = E_\mu\left[\log \frac{\mu(\omega_V)}{\nu(\omega_V)}\right], \tag{3.1}$$

a number in $[0, \infty]$ (cf. [10]).

(3.2) **Proposition.** *Suppose that ν is Markov of order r. Then the specific information gains*

$$h(\mu|\nu) = \lim_{a \to \infty} \frac{1}{|V(a)|} H(\mu_{V(a)}|\nu_{V(a)})$$

and

$$h_k^*(\mu|\nu) = \lim_{a \to \infty} \frac{1}{|V_k^*(a)|} H(\mu_{V_k^*(a)}|\nu_{V_k^*(a)})$$

exist and satisfy

$$h(\mu|\nu) = (1-k^{-d})\, h_k^*(\mu|\nu) + k^{-d} E_\mu[H(\mu_0(.|T_k^*))|\nu_0(.|T_k^*)] \qquad (k > r) \tag{3.3}$$

where the integrand on the right denotes the information gain of $\mu_0(.|T_k^*)(\omega)$ with respect to $\nu_0(.|T_k^*)(\omega) = \nu_0(.|N_r(0))(\omega)$.

Proof. Write

$$\frac{1}{|V|} H(\mu_V|\nu_V) = E_\mu\left[\frac{1}{|V|}\log \mu(\omega_V) - \frac{1}{|V|}\log \nu(\omega_V)\right]$$

and split up the two integrands in the manner of (2.9).

Let $k > r$. Since ν is Markov of order r we have $\nu_0(.|W) = \nu_0(.|N_r(0)) = \nu_0(.|T_k^*)$ as soon as $W \supseteq N_r(0)$. Thus we may certainly say that for any $\varepsilon > 0$ there is an integer $M > 0$ such that

$$\|\log \nu_0(\omega(0)|V_t) - \log \nu_0(\omega(0)|T_k^*(\|_\mu < \varepsilon \qquad \text{as soon as } V_t \supseteq T_k^* \cap N_M(0). \tag{3.4}$$

Moreover we have

$$(3.5) \qquad c(\mu, \nu) = \sup_{W \subset T_k^*} \|\log \nu_0(\omega(0)|W)\|_\mu < \infty$$

since $\nu_0(.|T-\{0\})(.)>0$, which depends only on the finitely many atoms of $\mathscr{F}_{N_r(0)}$, is bounded away from 0 by some $\alpha>0$ so that

$$\nu_0(.|W)(.) = E_\nu[\nu_0(.|T-\{0\})|\mathscr{F}_W](.)$$

may be assumed to be $\geqq \alpha$ for any $W \subseteq T-\{0\}$. Combining (3.4) and (3.5) with (2.10) and (2.12), we can proceed as in the proof of (2.4) and approximate the terms

$$\frac{1}{|V|} E_\mu \left[\sum_{t \in V_k} \log \frac{\mu_0(.|V_t)}{\nu_0(.|V_t)} \circ \theta_t^{-1} \right]$$

by

$$\frac{|V_k|}{|V|} E_\mu \left[\log \frac{\mu_0(.|T_k^*)}{\nu_0(.|T_k^*)} \right].$$

Granting the existence of $h(\mu|\nu)$ which will come out independently in (4.25) below, and thereby the existence of $h_k^*(\mu|\nu)$, we obtain (3.3).

(3.2) was motivated by the following

(3.6) *Remark.* From (3.3) we get

$$(3.7) \qquad h(\mu|\nu) \geqq E_\mu[H(\mu_0(.|T_k^*))|\nu_0(.|N_r(0))]$$

for $k>r$. Now suppose that we have $h(\mu|\nu)=0$. Then (3.7) yields

$$H(\mu_0(.|T_k^*))|\nu_0(.|N_r(0))(.) = 0 \qquad \mu\text{-a.s.}$$

and thus

$$\mu_0(.|T_k^*)(.) = \nu_0(.|N_r(0))(.) \qquad \mu\text{-a.s.}$$

(cf. [10]). Since $\bigcup_k T_k^* = T_\infty^* = T-\{0\}$, the martingale convergence theorem implies

$$\mu_0(.|T-\{0\}) = \lim_k \mu_0(.|T_k^*) = \nu_0(.|N_r(0)) \qquad \mu\text{-a.s.}$$

Thus μ and ν have the same local characteristics as soon as the specific information gain $h(\mu|\nu)$ vanishes. In particular μ is also Markov of order r.

We will show in the next section that this remark amounts to a variational characterization of the Markov property. But let us first extend it to the general case of this section where μ and ν just are stationary random fields.

(3.8) **Theorem.** *If $h(\mu|\nu)=0$ then μ and ν have the same local characteristics.*

Proof. 1) We may assume without loss of generality that μ is absolutely continuous with respect to ν with a bounded density. For if necessary take $\nu' = \frac{1}{2}(\mu+\nu)$. It is easy to check that $h(\mu|\nu)=0$ implies $h(\mu|\nu')=0$ and that μ and ν have the same local characteristics as soon as μ and ν' do.

2) Part 2) of the proof of (2.4), applied to the random field ν, shows that (3.4) and (3.5) hold with respect to the L_ν^1-norm. Now we use our assumption in 1) to replace the L_ν^1-norm by the L_μ^1-norm. But (3.4) and (3.5) is all we needed in order

to get (3.3). Thus $h(\mu|\nu)=0$ implies, as in the previous remark, that

$$\mu_0(.|T_k^*)=\nu_0(.|T_k^*) \qquad \mu\text{-a.s.}$$

for all $k\geqq 1$. Applying the martingale convergence theorem on both sides and using absolute continuity of μ with respect to ν we obtain

$$\mu_0(.|T-\{0\})=\nu_0(.|T-\{0\}) \qquad \mu\text{-a.s.}$$

4. Specific Energy

Let us denote by $\mathscr{A}$ the class of non-void finite subsets of T.

(4.1) *Definition.* A *potential* U is a collection of maps

$$U(A,.)\colon\ S^A \to R \qquad (A\in\mathscr{A}).$$

If $U(A,.)=0$ as soon as the diameter of A is greater than some fixed integer r then U is said to be of *finite range*. If this is the case with $r=1$ then U is called a *nearest neighbor potential*. We say that U is *stationary* if

(4.2) $$U(A,\omega_A)=U(A+t,\omega_{A+t})\circ\theta_t \qquad (t\in T,\, A\in\mathscr{A},\, \omega\in\Omega).$$

Let us fix for the rest of the paper a stationary potential U such that

(4.3) $$\|U\|=\sum_{0\in A\in\mathscr{A}}\|U(A,.)\|<\infty$$

where $\|U(A,.)\|$ is the supremum of $|U(A,.)|$ on S^A. If, for example, U is of finite range then (4.3) is clearly satisfied.

For $V\in\mathscr{A}$, $\xi\in S^V$, $\varphi\in S^{T-V}$ we define the *energy* of ξ on V given the *environment* φ as

(4.4) $$E_{V,\varphi}(\xi)=\sum_{\substack{A\in\mathscr{A}\\ A\cap V\neq\emptyset}} U(A,\tilde{\omega}_A)$$

where $\tilde{\omega}$ coincides with ξ on V and with φ on $T-V$. Note that the sum on the right is absolutely convergent due to (4.3). The probability measure $\pi_{V,\varphi}$ on S^V defined through

(4.5) $$\pi_{V,\varphi}(\xi)=Z_{V,\varphi}^{-1}\, e^{-E_{V,\varphi}(\xi)}$$

is called the *Gibbs distribution* on V given φ. The normalizing factor

(4.6) $$Z_{V,\varphi}=\sum_{\xi\in S^V} e^{-E_{V,\varphi}(\xi)}$$

is often called the (grand canonical) *partition function* on V given φ.

(4.7) *Definition.* A random field μ is called a *Gibbs field*, and we write $\mu\in G(U)$, if

(4.8) $$\mu(\omega_V=\xi|\mathscr{F}_{T-V})(\omega)=\pi_{V,\omega_{T-V}}(\xi) \qquad \mu\text{-a.s.}$$

whenever $V\in\mathscr{A}$ and $\xi\in S^V$. $G_0(U)$ will denote the set of stationary Gibbs fields.

(4.9) *Remarks.* 1) In general we have $|G(U)| \geqq |G_0(U)| \geqq 1$. In other words: (4.3) guarantees the existence but not the uniqueness of Gibbs fields resp. stationary Gibbs fields corresponding to U (cf. [6]).

2) It is enough to require (4.8) for the one-point sets $V=\{t\}$ with $t\in T$ (cf. [4]). Thus, in order to check if μ is a Gibbs field, one has only to look at the local characteristics of μ.

3) Any Markov field is a Gibbs field with respect to some nearest neighbor potential (cf. [1] and [11]). More generally: any Markov field of some finite order corresponds to some potential of finite range (cf. [2]).

(4.10) **Theorem.** *Let μ be a stationary random field. For each $a\in T$ choose a $\varphi(a)\in S^{T-V(a)}$. Then*

$$e(\omega)=\lim_{a\to\infty}\frac{1}{|V(a)|}E_{V(a),\varphi(a)}(\omega_{V(a)}) \tag{4.11}$$

exists, both μ-a.s. and in $L^p_\mu\,(p\geqq 1)$, and satisfies

$$e(.)=E_\mu\left[\sum_{0\in A\in\mathscr{A}}\frac{U(A,\omega_A)}{|A|}\,\middle|\,\mathscr{J}\right](.)\qquad \mu\text{-a.s.} \tag{4.12}$$

In particular $e(.)$ does not depend on the choice of the $\varphi(a)$'s.

(4.13) *Definition.* Let us call $e(\omega)$ the *specific energy* of the configuration ω (under μ) and the expectation

$$e(\mu)=E_\mu[e(.)] \tag{4.14}$$

the specific energy of μ. The *specific free energy* of μ is defined as

$$f(\mu)=e(\mu)-h(\mu). \tag{4.15}$$

Proof. 1) Define

$$g(\omega)=\sum_{0\in A\in\mathscr{A}}\frac{U(A,\omega_A)}{|A|}\qquad(\omega\in\Omega).$$

The d-dimensional ergodic theorem says that

$$\frac{1}{|V(a)|}\sum_{t\in V(a)}g\circ\theta_t$$

converges to the right side in (4.12), both μ-a.s. and in L^p_μ ([5], VIII, 6.9). It is therefore enough to show

$$\Big|E_{V,\varphi}(\omega_V)-\sum_{t\in V}g\circ\theta_t(\omega)\Big|\leqq 2\Delta(V) \tag{4.16}$$

where $\Delta(V)$ does not depend on ω and φ and satisfies

$$\Delta(V(a))\,|V(a)|^{-1}\to 0\quad\text{as } a\to\infty. \tag{4.17}$$

2) Write

$$E_{V,\varphi}(\omega_V)=\sum_{V\supseteq A\in\mathscr{A}}U(A,\omega_A)+\sum_{\substack{A\in\mathscr{A}\\ A\cap V\neq\emptyset\\ A\cap(T-V)\neq\emptyset}}U(A,\tilde{\omega}_A) \tag{4.18}$$

where $\tilde{\omega}$ coincides with ω on V and with φ on $T-V$, and

$$\text{(4.19)} \qquad \sum_{t\in V} g\circ\theta_t(\omega) = \sum_{t\in V}\sum_{t\in A\subseteq V} \frac{U(A,\omega_A)}{|A|} + \sum_{t\in V}\sum_{\substack{t\in A\in\mathscr{A}\\ A\cap(T-V)\neq\emptyset}} \frac{U(A,\omega_A)}{|A|}.$$

Now note that the first term on the right is the same in (4.18) and (4.19), and that the second term on the right both in (4.18) and (4.19) is dominated in absolute value by

$$\Delta(V) = \sum_{t\in V}\sum_{\substack{t\in A\in\mathscr{A}\\ A\cap(T-V)\neq\emptyset}} \|U(A,.)\|.$$

3) We have still to check that the terms $\Delta(V)$ defined in 2) satisfy (4.17). For $N\geqq 1$ let us denote by $\mathscr{A}^N$ the class of those sets $A\in\mathscr{A}$ which have diameter $\leqq N$. For $\varepsilon>0$ we can choose N such that

$$\sum_{t\in A\in\mathscr{A}-\mathscr{A}^N} \|U(A,.)\| < \varepsilon \qquad (t\in T),$$

due to (4.3) and the stationarity of U. Then we have

$$\begin{aligned}\Delta(V) &\leqq \varepsilon\,|V| + \sum_{t\in V}\sum_{\substack{t\in A\in\mathscr{A}^N\\ A\cap(T-V)\neq\emptyset}} \|U(A,.)\| \\ &= \varepsilon\,|V| + |V^N|\,\|U\|\end{aligned}$$

where $|V^N|$ is the number of points $t\in V$ which have distance not greater than N from $T-V$. But for any N we have $|V^N(a)|\,|V(a)|^{-1}\to 0$ as $a\to\infty$, and this shows (4.17).

Now let us take up again the notion of information gain discussed in Section 3.

(4.20) **Lemma.** *If μ is a random field and ν a Gibbs field then*

$$\lim_{a\to\infty} \frac{1}{|V(a)|}\left[H(\mu_{V(a)}|\nu_{V(a)}) - H(\mu_{V(a)}|\pi_{V(a),\varphi(a)})\right] = 0$$

for any choice of the $\varphi(a)$'s.

Proof. (4.16) implies $|E_{V,\varphi}(\xi) - E_{V,\psi}(\xi)| \leqq \Delta(V)$ for any ξ, φ, ψ, and this yields

$$\text{(4.21)} \qquad e^{-2\Delta(V)} \leqq \frac{\pi_{V,\psi}(\xi)}{\pi_{V,\varphi}(\xi)} \leqq e^{2\Delta(V)}$$

via (4.5) and (4.6). Now consider

$$H(\mu_V|\pi_{V,\varphi}) - H(\mu_V|\nu_V) = E_\mu\left[\log\frac{\nu(\omega_V)}{\pi_{V,\varphi}(\omega_V)}\right].$$

Since

$$\frac{\nu(\omega_V)}{\pi_{V,\varphi}(\omega_V)} = \int \frac{\pi_{V,\omega_{T-V}}(\omega_V)}{\pi_{V,\varphi}(\omega_V)}\, d\nu(\omega_{T-V})$$

is a mixture of terms such as in (4.21), we obtain

$$\left|\log \frac{\nu(\omega_V)}{\pi_{V,\varphi}(\omega_V)}\right| \leqq 2\Delta(V), \tag{4.22}$$

and by (4.17) the lemma follows.

For any random field μ we can write

$$\frac{1}{|V|} H(\mu_V|\pi_{V,\varphi}) = \frac{1}{|V|} \log Z_{V,\varphi} + \frac{1}{|V|} E_\mu[E_{V,\varphi}(\,.)] - \frac{1}{|V|} H(\mu_V). \tag{4.23}$$

Now assume that μ is stationary. Then the last two terms on the right converge for $V=V(a)$ and $a\to\infty$ to $e(\mu)-h(\mu)$ (cf. (4.10), (2.4)). This yields the theorem of Lee and Yang, i.e. the existence of

$$p = \lim_{a\to\infty} \frac{1}{|V(a)|} \log Z_{V(a),\varphi(a)} \tag{4.24}$$

and its independence of the $\varphi(a)$'s (use (4.20) with $\mu=\nu$ so that the left side in (4.23) converges to 0). Thus all the terms in (4.23) converge. Now take $\nu \in G(U)$. Combining (4.20) and (4.23) we obtain the existence of the specific information gain $h(\mu|\nu)$ and the relations

$$\begin{aligned} h(\mu|\nu) &= \lim_{a\to\infty} \frac{1}{|V(a)|} H(\mu_{V(a)}|\pi_{V(a),\varphi(a)}) \\ &= p + e(\mu) - h(\mu). \end{aligned} \tag{4.25}$$

We may also write

$$0 \leqq h(\mu|\nu) = f(\mu) - f(\nu) \tag{4.26}$$

(note that $f(\nu) = -p$ as soon as $\nu \in G_0(U)$ due to (4.25) with $\mu=\nu$, *define* $f(\nu) = -p$ if $\nu \in G(U) - G_0(U)$, and recall that information gains are always non-negative).

We are now in a position to translate our result (3.8) on specific information gain into Lanford and Ruelle's characterization of Gibbs states by a variational principle, that is, the equivalence (i) $\Leftrightarrow$ (iv) below.

(4.27) **Theorem.** *Let μ be a stationary random field and $\nu \in G_0(U)$. Then the following statements are equivalent:*

(i) $\mu \in G(U)$.

(ii) $\lim_{a\to\infty} \frac{1}{|V(a)|} H(\mu_{V(a)}|\pi_{V(a),\varphi(a)}) = 0$ *for any choice of the $\varphi(a)$'s.*

(iii) $h(\mu|\nu) = 0$.

(iv) *The specific free energy $f(.)$ assumes in μ its minimum.*

Proof. (iii) $\Rightarrow$ (i), which is the crucial step, is contained in (3.8). (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iii) $\Leftrightarrow$ (iv) is a summary of this section.

Let us conclude with the d-dimensional version, in the case of Gibbs fields, of Breiman's theorem that there is almost sure convergence behind the existence of specific entropy:

(4.28) *Remark.* Let ν be a stationary Gibbs field. Since

$$-\frac{1}{|V(a)|}\log \pi_{V(a),\varphi(a)} = \frac{1}{|V(a)|}\log Z_{V(a),\varphi(a)} + \frac{1}{|V(a)|} E_{V(a),\varphi(a)}(\omega_V)$$

converges ν-a.s. to $p+e(\omega)$ as $a\to\infty$ (cf. (4.24), (4.10)), and since

$$\lim_{a\to\infty} \frac{1}{|V(a)|} |\log \pi_{V(a),\varphi(a)}(\omega_{V(a)}) - \log \nu(\omega_{V(a)})| = 0$$

for any ω by (4.22), we obtain the ν-almost sure existence of

$$h(\omega) = \lim_{a\to\infty} \left[-\frac{1}{|V(a)|}\log \nu(\omega_{V(a)})\right]$$

and the relation $h(.)=p+e(.)$ ν-a.s.

References

1. Averintsev, M. B.: A way of describing random fields with discrete parameter (in Russian). Problemy Peredači Informacii, **6**, 100–109 (1970).
2. Averintsev, M. B.: Description of Markov random fields by conditional Gibbs distributions (in Russian). Teor. Ver. i ee Prim. **17**, 2, 21–35 (1972).
3. Dobrušin, R. L.: Gibbsian random fields for lattice systems with pair interactions. Functional Anal. Appl. **2**, 281–301 (1968).
4. Dobrušin, R. L.: Description of a random field by means of conditional probabilities and conditions of its regularity. Theor. Probability Appl. **13**, 197–224 (1968).
5. Dunford, N. and Schwartz, J. T.: Linear Operators, I. New York: Interscience 1958.
6. Georgii, H.-O.: Phasenübergang 1. Art bei Gittergasmodellen. Lecture Notes in Physics **16**. Heidelberg-Berlin-New York: Springer 1972.
7. Jacobs, K.: Lecture Notes on Ergodic Theory. Aarhus Universitet: Aarhus 1963.
8. Lanford III, O. E., Ruelle, D.: Observables at infinity and states with short range correlations in statistical mechanics. Comm. Math. Phys. **13**, 194–215 (1969).
9. Meyer, P. A.: Probability and Potentials. Waltham-Toronto-London: Blaisdell 1966.
10. Rényi, A.: Probability Theory. North-Holland: Amsterdam 1970.
11. Spitzer, F.: Random fields and interacting particle systems. Notes on lectures given at the 1971 MAA Summer Seminar, Williams College, Williamstown, Mass. Mathematical Association of America 1971.
12. Spitzer, F.: A variational characterization of finite Markov chains. Ann. Math. Statist. **93**, 1, 1303–1307 (1971).
13. Thouvenot, J.-P.: Convergence en moyenne de l'information pour l'action de Z^2. Z. Wahrscheinlichkeitstheorie verw. Gebiete **24**, 2, 135–137 (1972).

Hans Föllmer
Mathematisches Seminar der
Johann Wolfgang Goethe-Universität
Institut für Angewandte Mathematik
D-6000 Frankfurt am Main 1
Robert-Mayer-Str. 6–10
Federal Republic of Germany

(Received January 15, 1973)