Image Analysis, Random Fields and Dynamic Monte Carlo Methods

A Mathematical Introduction

With 59 Figures
This chapter will be theoretical and—besides the examples—possibly a bit dry. No doubt, some basic ideas can be imparted without this material. But a deeper understanding, in particular of topics like texture, parameter estimation or parallel algorithms, requires some abstract background and therefore one has to learn random fields. In this chapter, we present some basic notions and elementary results.

### 3.1 Markov Random Fields

Discrete images were represented by elements of finite product spaces and special probability distributions on the set of such images were discussed. An appropriate abstract setting will now be introduced.

Let $S$ be a finite index set—the set of sites; for every site $s \in S$ let $X_s$ be a finite space of states $x_s$. The product $X = \prod_{s \in S} X_s$ is the space of (finite) configurations $x = (x_s)_{s \in S}$. We consider probability measures or distributions $\Pi$ on $X$, i.e., vectors $\Pi = (\Pi(x))_{x \in X}$ such that $\Pi(x) \geq 0$ and $\sum_{x \in X} \Pi(x) = 1$. Subsets $E \subset X$ are called events; the probability of an event $E$ is given by $\Pi(E) = \sum_{x \in E} \Pi(x)$. A strictly positive probability measure $\Pi$ on $X$, i.e., $\Pi(x) > 0$ for every $x \in X$, is called a stochastic or random field.

For $A \subset S$ let $X_A = \prod_{s \in A} X_s$ denote the space of configurations $x_A = (x_s)_{s \in A}$ on $A$; the map

$$X_A : X \rightarrow X_A, \quad x = (x_s)_{s \in S} \mapsto (x_s)_{s \in A}$$

is the projection of $X$ onto $X_A$. We shall use the short-hand notation $X_s$ for $X\{s\}$ and $\{X_A = x_A\}$ for $\{x \in X : X_A(x) = x_A\}$. Commonly one writes $\{X_A = x_A, X_B = x_B\}$ for intersections $\{X_A = x_A\} \cap \{X_B = x_B\}$. For a random field $\Pi$ the random vector $X = (X_s)_{s \in S}$ on the probability space $(X, \Pi)$ is also frequently called a random field.

For events $E$ and $F$ the conditional probability of $F$ given $E$ is defined by $\Pi(F|E) = \Pi(F \cap E)/\Pi(E)$. Conditional probabilities of the form

$$\Pi(X_A = x_A | X_{S \setminus A} = x_{S \setminus A}), \ A \subset S, \ x_A \in X_A, \ x_{S \setminus A} \in X_{S \setminus A}$$
are called local characteristics. They are always defined since random fields are assumed to be strictly positive. They express the probability that the configuration is $x_A$ on $A$ and $x_{S\setminus A}$ on the rest of the world. Later on, we shall use the short-hand notation $\Pi(x_A \, | \, x_{S\setminus A})$.

We compute now local characteristics for a simple random field.

**Example 3.1.1.** Let $X_s = \{-1, 1\}$ for all $s \in S$. Then

$$\Pi(x) = \frac{1}{Z} \exp \left( \sum_{\langle s, t \rangle} x_s x_t \right),$$

where $Z$ is the normalization constant. The index set $S$ is a finite square lattice and $\langle s, t \rangle$ means that $t$ is the site next to $s$ on the right or left or the next upper or lower site (or more generally, $S$ is a finite undirected graph with bonds $\langle s, t \rangle$). Then

$$\Pi(X_t = x_t \, | \, X_r = x_r, \, r \neq t) = \frac{\Pi(X_s = x_s \text{ for all } s)}{\Pi(X_s = x_s \text{ for all } s \neq t)}$$

$$= \frac{\exp \left( \sum_{\langle s, t \rangle} x_s x_t \right) \exp \left( \sum_{\langle r, s \rangle, \, r \neq t, s \neq t} x_r x_s \right)}{\sum_{x_t \in X_t} \exp \left( \sum_{\langle s, t \rangle} x_s x_t \right) \exp \left( \sum_{\langle r, s \rangle, \, r \neq t, s \neq t} x_r x_s \right)}$$

$$= \frac{\exp \left( x_t \sum_{\langle s, t \rangle} x_s \right)}{\sum_{x_t \in X_t} \exp \left( x_t \sum_{\langle s, t \rangle} x_s \right)}.$$

Hence the conditional probabilities have a particularly simple form; for example,

$$\Pi(X_t = -1 \, | \, X_r = x_r, \, r \neq t) = \frac{1}{1 + \exp \left( 2 \sum_{\langle t, r \rangle} x_r \right)}.$$

This shows: The probability for the state $x_t$ in $t$ given the configuration on the rest of $S$ depends on the states on the (four) neighbours of $t$ only. It is not affected by a change of colours in sites which are no neighbours of $t$.

The local characteristics of the other distributions in the last chapter also depend only on a small number of neighbouring sites. If so, conditional distributions can be computed in reasonable time whereas the computing time would not be feasible for dependence say on all states if the underlying space is large. Later on, we shall develop algorithms for the approximate computation of MAP estimates. They will depend on the iterative computation.
of local characteristics and there local dependence will be crucial. We shall
discuss local dependence in more detail now.

Those sites which possibly influence the local characteristic at a site \( s \) will
be called the neighbours of \( s \). The relation \( s \) and \( t \) are neighbours’ should
fulfill some axioms.

**Definition 3.1.1.** A collection \( \partial = \{ \partial(s) : s \in S \} \) of subsets of \( S \) is called
a *neighbourhood system*, if (i) \( s \notin \partial(s) \) and (ii) \( s \in \partial(t) \) if and only if
\( t \in \partial(s) \). The sites \( s \in \partial(t) \) are called *neighbours* of \( t \). A subset \( C \) of \( S \) is
called a *clique* if two different elements of \( C \) are always neighbours. The set
of cliques will be denoted by \( C \). We shall frequently write \( (s, t) \) if \( s \) and \( t \) are
neighbours of each other.

**Remark 3.1.1.** The neighbourhood relation induces an undirected graph with
vertices \( s \in S \) and a bond between \( s \) and \( t \) if and only if \( s \) and \( t \) are neigh-
bours. Conversely, an undirected graph induces a neighbourhood system. The
‘complete’ sets in the graph correspond to the cliques.

**Example 3.1.2.** (a) A degenerate neighbourhood system is given by \( \partial(s) = \emptyset \)
for all \( s \in S \). There are no nonempty cliques and the sites act independently
of each other.

(b) The other extreme is \( \partial(s) = S \setminus \{ s \} \) for all \( s \in S \). All subsets of \( S \) are
cliques and all sites influence each other.

(c) Some of the neighbourhood systems used in the last chapter are of the
following type: The index set is a finite lattice

\[
S = \{(i, j) \in \mathbb{Z} \times \mathbb{Z} : -m \leq i, j \leq m \}
\]

and

\[
\partial((i, j)) = \{(k, l) : 0 < (k - i)^2 + (l - j)^2 \leq C \}.
\]

Up to modifications near the boundary, a site \(* \) has for \( C = 1 \) the upper,
lower, left and right site as neighbours; in this case the cliques are

\[
\begin{array}{l}
\emptyset, \ast, \ast\ast \text{ and } |
\end{array}
\]

For \( C = 2 \) and sites \((i, j), i, j \notin \{-m, m\} \), the neighbours \( \circ \) of \(* \) are:

\[
\begin{array}{ccc}
\circ & \circ & \circ \\
\circ & \ast & \circ \\
\circ & \circ & \circ
\end{array}
\]

The corresponding cliques are:
For sites near the boundary the cliques are smaller which may cause some trouble in programming the algorithms.

(d) If there is a pixel and an edge process there may be interaction between pixels, between edges and between pixels and edges. If $S^P$ is a lattice of pixels and $S^E$ the set of microedges then the index set for $(x^P, x^E)$ is $S = S^P \cup S^E$. There may be a neighbourhood system on $S^P$ as in (c) and microedges can be neighbours of pixels and vice versa. For example, pixel * can have

neighbouring edges $| - \text{ like } | * |$.

Now we can formalize local dependence as indicated in Example 3.1.1.

**Definition 3.1.2.** The random field $\Pi$ is a Markov field w.r.t. the neighbourhood system $\partial$ if for all $x \in X$,

$$\Pi (X_s = x_s | X_r = x_r, r \neq s) = \Pi (X_s = x_s | X_r = x_r, r \notin \partial(s)).$$

This definition takes only single site local characteristics into account. The others inherit this property by 3.3.2(b).

**Remark 3.1.2.** For finite product spaces $X$ the above conditions are in principle no restriction since every random field is a Markov field for the neighbourhood system 3.1.2 where all different sites are neighbours. But we are looking for random fields which are Markov for small neighbourhoods. For instance, the Markov property for the neighbourhood system $\partial(s) = \emptyset$ boils down to

$$\Pi (X_s = x_s | X_r = x_r, r \neq s) = \Pi (X_s = x_s).$$

Since for events $E_1, \ldots, E_k$ with nonempty intersection,

$$\Pi (E_1 \cap \ldots \cap E_k) = \Pi (E_1) \cdot \Pi (E_2 | E_1) \cdot \ldots \cdot \Pi (E_k | E_1 \cap \ldots \cap E_{k-1})$$

this implies that the random variables $X_s$ are independent. Large neighbourhoods correspond to long-range dependence.
3.2 Gibbs Fields and Potentials

Now we turn to the representation of random fields in the Gibbsian form (1.1). It is particularly useful for the calculation of (conditional) probabilities. The idea and hence most of the terminology is borrowed from statistical mechanics where Gibbs fields are used as models for the equilibrium states of large physical systems (cf. Example 3.2.1).

Probability measures of the form

$$\Pi(x) = \frac{\exp(-H(x))}{\sum_{z \in \mathcal{X}} \exp(-H(z))}$$

are always strictly positive and hence random fields. $\Pi$ is called the Gibbs field (or measure) induced by the energy function $H$ and the numerator is called the partition function. Every random field $\Pi$ can be written in this form. In fact, setting $H(x) = -\ln \Pi(x) - \ln Z$, one gets $\exp(-H(x)) = \Pi(x)Z$ and $Z$ necessarily is the partition function of $H$. Moreover, the energy function for $\Pi$ is unique up to an additive constant; if $H$ and $H'$ are energy functions for $\Pi$ then

$$H(x) - H'(x) = \ln Z' - \ln Z$$

for every $x \in \mathcal{X}$. It is common to enforce uniqueness choosing some reference or 'vacuum' configuration $o \in \mathcal{X}$ and requiring $Z = \Pi(o)^{-1}$, or, equivalently, $H(o) = 0$.

Hence we restrict attention to Gibbs fields. It is convenient to decompose the energy into the contributions of the configurations on subsets of $\mathcal{S}$. Let $\emptyset$ denote the empty set.

Definition 3.2.1. A potential is a family $\{U_A : A \subset \mathcal{S}\}$ of functions on $\mathcal{X}$ such that

(i) $U_\emptyset = 0,$
(ii) $U_A(x) = U_A(y)$ if $X_A(x) = X_A(y)$.

The energy of the potential $U$ is given by

$$H_U = \sum_{A \subset \mathcal{S}} U_A.$$

Given a neighbourhood system $\partial$ a potential $U$ is called a neighbour potential w.r.t. $\partial$ if $U_A = 0$ whenever $A$ is not a clique. If $U_A = 0$ for $|A| > 2$ then $U$ is a pair potential.

Potentials define energy functions and thus random fields.

Definition 3.2.2. A random field $\Pi$ is a Gibbs field or Gibbs measure for the potential $U$, if it is of the form (3.2) and $H$ is the energy $H_U$ of a potential $U$. If $U$ is a neighbour potential then $\Pi$ is called a neighbour Gibbs field.
We give some examples.

Example 3.2.1. (a) The Ising model is particularly simple. But it shows phenomena which are also typical for more complex models. Hence it is frequently the starting point for the study of deep questions about Markov fields. It will be used as an example throughout this text.

$S$ is a finite square lattice and the neighbours of $s \in S$ are the sites with Euclidean distance one (which is the case $C = 1$ in Example 3.1.2(c)). The possible states are $-1$ and $1$ for every site. In the simplest case the energy function is given by

$$H(x) = -\beta \sum_{(s,t)} x_s x_t$$

where $(s,t)$ indicates that $s$ and $t$ are neighbours. Hence $H$ is the energy function of a neighbour potential (in fact, of a pair potential). The configurations of minimal energy are the constant configurations with states $-1$ and $1$, respectively.

Physicists study a slightly more general model: index set, neighbourhood system and state space are the same but the energy function is given by

$$H(x) = -\frac{1}{kT} \left[ J \sum_{(s,t)} x_s x_t - m B \sum_s x_s \right].$$

The German physicist E. Ising (1925; the $I$ pronounced like in eagle and not like in ice) tried to explain theoretically certain empirical facts about ferromagnets by means of this model; it was proposed by Ising's doctoral supervisor W. Lenz in 1920. The lattice is thought of as a crystal lattice, $x_s = \pm 1$ means, that there is a small dipole or spin at the lattice point $s$ which is directed either upwards or downwards. Ising considered only one-dimensional (but infinite) lattices and argued by analogy for higher dimension (unfortunately these conclusions were wrong).

The first term represents the interaction energy of the spins. Only neighbouring spins interact and hence the model is not suited for long-range interactions. $J$ is a matter constant. If $J > 0$ then spins with the same direction contribute low energy and hence high probability. Thus the spins tend to have the same direction and we have a ferromagnet. For $J < 0$ one has an antiferromagnet. The constant $T > 0$ represents absolute temperature and $k$ is the 'Boltzmann factor'. At low temperature (or for large $J$) there is strong interaction and there are collective phenomena; at high temperature there is weak coupling and the spins act almost independently. The second sum represents a constant external field with intensity $B$. The constant $m > 0$ depends again on the material. This term becomes minimal if all spins are parallel to the external field. Besides in physics, similar models were also adopted in various fields like biology, economics or sociology. We used it for smoothing.
The increasing strength of coupling with increasing parameter $\beta$ can be illustrated by sampling from the Ising field at various values of $\beta$. The samples in Fig. 3.1 were taken (from left to right) for values $\beta = 0.1, 0.45, 0.47$ and $4.0$ on a $56 \times 56$ lattice; there is no external field. They range from almost random to ‘nearly constant’.

Fig. 3.1. Typical configurations of an Ising field at various temperatures

The natural generalization to more than two states is

$$H(x) = -\beta \sum_{(s,t)} 1(x_s = x_t).$$

It is called the **Potts model**.

(b) More generally, each term in the sum may be weighted individually, i.e.

$$H(x) = \sum_{(s,t)} a_{st} x_s x_t + \sum_s a_s x_s$$

where $x_s = \pm 1$. If $a_{st} = 1$ then $x_s = x_t$ is favourable and, conversely, $a_{st} = -1$ encourages $x_s = -x_t$. For the following pictures, we set all $a_s$ to 0 and almost all $a_{st}$ to $+1$ like in the Ising model but some to $-1$ (the reader may guess which!). The samples from the associated Gibbs field were taken at the same parameter values as in Fig. 3.1. With increasing $\beta$ the samples contain larger and larger portions of the image in Fig. 2.3(a) or of its inverse much like the
samples in 3.1 contain larger and larger patches of black and white. Fig. 3.2 may look nicer than 3.1 but it does not tell us more about Gibbs fields.

(c) Nearest neighbour binary models are lattice models with the same neighbourhood structure as before but with values in \( \{0,1\} \):

\[
H(x) = \sum_{\langle s,t \rangle} b_{st} x_s x_t + \sum_s b_s x_s, \; x_s \in \{0,1\}.
\]

In the 'autologistic model', \( b_{st} = b_h \) for horizontal and \( b_{st} = b_v \) for vertical bonds; sometimes the general form is also called autologistic. In the isotropic case \( b_{st} = a \) and \( b_s = b \); it looks like an Ising model and in fact, the models in (b) and the nearest neighbour binary models are equivalent by the transformation \( \{0,1\} \rightarrow \{-1,1\}, \; x_s \mapsto 2x_s - 1 \).

Plainly, models of the form (b) or (c) can be defined on any finite undirected graph with a set \( S \) of nodes and \( \langle s,t \rangle \) if and only if there is a bond between \( s \) and \( t \) in the graph. Such models play a particularly important role in neural networks (cf. KAMP and HASLER (1990)). In imaging, these and related models are used for description, synthesis and classification of binary textures (cf. Chapter 15). Generalizations (cf. the Potts model) apply to textures with more than two colours.

(d) Spin glass models do not fit into this framework but they are natural generalizations. The coefficients \( a_{st} \) and \( a_s \) are themselves random variables. In the physical context they model the 'random environment' in which the
particles with states $x_s$ live. Spin glasses become more and more popular in
the Neural Network community, cf. the work of van Hemmen and others.

If a Markov field is given by a potential then the local characteristics may
easily be calculated. For us this is the main reason to introduce potentials.

**Proposition 3.2.1.** Let the random field $\Pi$ be given by some neighbour
potential $U$ for the neighbourhood system $\partial$, i.e.

$$
\Pi(x) = \frac{\exp \left( - \sum_{C \in \mathcal{C}} U_C(x) \right)}{\sum_y \exp \left( - \sum_{C \in \mathcal{C}} U_C(y) \right)}
$$

where $\mathcal{C}$ denotes the set of cliques of $\partial$. Then the local characteristics are
given by

$$
\Pi \left( X_s = x_s, s \in A \mid X_s = x_s, s \in S \setminus A \right) = \\
\frac{\exp \left( - \sum_{C \in \mathcal{C}, C \cap A \neq \emptyset} U_C(x) \right)}{\sum_{y_A \in X_A} \exp \left( - \sum_{C \in \mathcal{C}, C \cap A \neq \emptyset} U_C(y_A | x_{S \setminus A}) \right)}
$$

(For a general potential, replace $\mathcal{C}$ on the right-hand side by the power set of
$S$.) Moreover,

$$
\Pi \left( X_s = x_s, s \in A \mid X_s = x_s, s \in S \setminus A \right) = \\
= \Pi \left( X_s = x_s, s \in A \mid X_s = x_s, s \in \partial(A) \right)
$$

for every subset $A$ of $S$. In particular, $\Pi$ is a Markov field w.r.t. $\partial$.

**Proof.** By assumption,

$$
\Pi \left( X_A = x_A \mid X_{S \setminus A} = x_{S \setminus A} \right) = \\
\frac{\Pi \left( X = x_A | x_{S \setminus A} \right)}{\Pi \left( X_{S \setminus A} = x_{S \setminus A} \right)} = \\
= \frac{\exp \left( - \sum_{C \in \mathcal{C}} U_C \left( x_A | x_{S \setminus A} \right) \right)}{\sum_{y_A \in X_A} \exp \left( - \sum_{C \in \mathcal{C}} U_C \left( y_A | x_{S \setminus A} \right) \right)}
$$

Divide now the set of cliques into two classes:

$$
\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 = \{ C \in \mathcal{C} : C \cap A \neq \emptyset \} \cup \{ C \in \mathcal{C} : C \cap A = \emptyset \}.
$$
Letting $R = S \setminus (A \cup \partial A)$ where $\partial A = \cup_{s \in A} \partial(s) \setminus A$ and introducing a reference element $o \in X$,

$$U_C (z_{Az \partial AzR}) = U_C (o_{Az \partial AzR}) \quad \text{if} \quad C \in C_2,$$

and similarly,

$$U_C (z_{Az \partial AzR}) = U_C (z_{Az \partial AoR}) \quad \text{if} \quad C \in C_1.$$

Rewrite the sum as

$$\sum_{C \in C} \ldots = \sum_{C \in C_1} \ldots + \sum_{C \in C_2} \ldots,$$

and use the multiplicativity of exponentials to check that in the above fraction the terms for cliques in $C_2$ cancel out. Let $x_{\partial A}$ denote the restriction of $x_{S \setminus A}$ to $\partial A$. Then

$$\Pi (X_A = x_A | X_{S \setminus A} = x_{S \setminus A}) = \frac{\exp \left( - \sum_{C \in C_1} U_C (x_{Az \partial AoR}) \right)}{\sum_{y_A \in x_A} \exp \left( - \sum_{C \in C_1} U_C (y_{Az \partial AoR}) \right)}$$

which is the desired form since $U_C$ does not depend on the configurations on $R$.

The last expression equals

$$= \exp \left( - \sum_{C \in C_1} U_C (x_{Az \partial AoR}) \right) \cdot \sum_{y_A} \exp \left( - \sum_{C \in C_2} U_C (o_{Ax \partial AyR}) \right)$$

$$= \sum_{y_A} \exp \left( - \sum_{C \in C_1} U_C (y_{Az \partial AoR}) \right) \cdot \sum_{y_R} \exp \left( - \sum_{C \in C_2} U_C (o_{Ax \partial AyR}) \right)$$

$$= \sum_{y_A} \sum_{y_R} \exp \left( - \sum_{C \in C_1} U_C (y_{Az \partial AyR}) \right) \cdot \exp \left( - \sum_{C \in C_2} U_C (y_{Ax \partial AyR}) \right)$$

$$= \Pi (X_A = x_A | X_{\partial A} = x_{\partial A}).$$

Specializing to sets of the form $A = \{s\}$ shows that $\Pi$ is a Markov field for $\partial$. This completes the proof.
3.3 More on Potentials

The following results are not needed for the next chapters. They will be used in later chapters and may be skipped in a first reading. On the other hand, they are recommended as valuable exercises on random fields.

For technical reasons, we fix in each component $X_t$ a reference element $o_t$ and set $o = (o_t)_{t \in S}$. For a configuration $x$ and a subset $A$ of $S$ we denote by $^A x$ the configuration which coincides with $x$ on $A$ and with $o$ off $A$.

**Theorem 3.3.1.** Every random field $\Pi$ is a Gibbs field for some potential. We may choose the potential $V$ with $V_o = 0$ and which for $A \neq \emptyset$ is given by

$$V_A(x) = - \sum_{B \subset A} (-1)^{|A-B|} \ln \left( \Pi \left( ^B x \right) \right).$$

(3.1)

For all $A \subset S$ and every $a \in A$,

$$V_A(x) = - \sum_{B \subset A} (-1)^{|A-B|} \ln \left( \Pi \left( X_a = ^B x_a \vert X_s = ^B x_s, s \neq a \right) \right).$$

(3.2)

For the potential $V$ one has $V_A(x) = 0$ whenever $x_a = o_a$ for some $a \in A$.

**Remark 3.3.1.** If a potential $V$ fulfills $V_A(x) = 0$ whenever $x_a = o_a$ for some $a \in A$ then it is called normalized. We shall prove that $V$ from the theorem is the only normalized potential for $\Pi$ (cf. Theorem 3.3.3 below). The proof below will show that the vacuum $o$ has probability $\Pi(o) = (\sum_z \exp(-H_V(z)))^{-1} = Z^{-1}$ which is equivalent to $H_V(o) = 0$. This explains why a normalized potential is also called a vacuum potential and the reference configuration $o$ is called the vacuum (in physics, the ‘real vacuum’ is the natural choice for $o$). If $\Pi$ is given in the Gibbsian form by any potential then it is related to the normalized potential by the formula in Theorem 3.3.3.

**Example 3.3.1.** Let $x_s \in \{0, 1\}$, $V_{\{s\}}(x) = b_s x_s$, $V_{\{s,t\}}(x) = b_{st} x_s x_t$ and $V_A \equiv 0$ whenever $|A| \geq 3$. Then $V$ is a normalized potential. Such potentials are of interest in texture modeling and neural networks.

For the proof of Theorem 3.3.1 we need the Moebius inversion formula, which is of independent interest.

**Lemma 3.3.1.** Let $S$ be a finite set and $\Phi$ and $\Psi$ real-valued functions on the power set of $S$. Then

$$\Phi(A) = \sum_{B \subset A} (-1)^{|A-B|} \Psi(B) \quad \text{for every} \quad A \subset S$$

if and only if

$$\Psi(A) = \sum_{B \subset A} \Phi(B) \quad \text{for every} \quad A \subset S.$$
Proof (of the lemma). For the above theorem we need that the first condition implies the second one. We rewrite the right-hand side of the second formula as

\[ \sum_{B \subset A} \Phi(B) = \sum_{B \subset A} \sum_{D \subset B} (-1)^{|B-D|} \Psi(D) \]

\[ = \sum_{D \subset A, C \subset A \setminus D} (-1)^{|C|} \Psi(D) \]

\[ = \sum_{D \subset A} \Psi(D) \sum_{C \subset A \setminus D} (-1)^{|C|} = \Psi(A). \]

Let us comment on the last equation. We note first that the inner sum equals 1 if \( A \setminus D = \emptyset \). If \( A \setminus D \neq \emptyset \), then we have setting \( n = |A \setminus D| \),

\[ \sum_{C \subset A \setminus D} (-1)^{|C|} = \sum_{k=0}^{n} \{ C \subset A \setminus D : |C| = k \} (-1)^k \]

\[ = \sum_{k=0}^{n} \binom{n}{k} (-1)^k = (1 - 1)^n = 0. \]

Thus the equation is clear.

For the converse implication assume that the second condition holds. Then the same arguments show

\[ \sum_{B \subset A} (-1)^{|A-B|} \Psi(B) = \sum_{D \subset B \subset A} (-1)^{|A-B|} \Phi(D) \]

\[ = \sum_{D \subset A} \Phi(D) \sum_{C \subset A \setminus D} (-1)^{|C|} = \Phi(A) \]

which proves the lemma. \( \Box \)

Now we can prove the theorem. We shall write \( B + a \) for \( B \cup \{a\} \).

Proof (of Theorem 3.3.1). We use the Möbius inversion for

\[ \Phi(B) = -V_B(x), \quad \Psi(B) = \ln \left( \frac{\Pi(B,x)}{\Pi(o)} \right). \]

Suppose \( A \neq \emptyset \). Then \( \sum_{B \subset A} (-1)^{|A-B|} = 0 \) (cf. the last proof) and hence

\[ \Phi(A) = -V_A(x) \]

\[ = \sum_{B \subset A} (-1)^{|A-B|} \ln \left( \Pi(B,x) \right) - \ln \left( \Pi(o) \right) \sum_{B \subset A} (-1)^{|A-B|} \]

\[ = \sum_{B \subset A} (-1)^{|A-B|} \Psi(B). \]
Furthermore,
\[ \Phi(\emptyset) = -V_o(x) = 0 = \ln \left( \frac{\Pi(x)}{\Pi(\emptyset)} \right) = \Phi(\emptyset). \]
Hence the assumptions of the lemma are fulfilled. We conclude
\[ \ln \left( \frac{\Pi(x)}{\Pi(\emptyset)} \right) = \Phi(S) = \sum_{B \subseteq S} \Phi(B) = -\sum_{B \subseteq S} V_B(x) = -H_V(x) \]
and thus
\[ \Pi(x) = \Pi(\emptyset) \exp(-H_V(x)). \]
Since \( \Pi \) is a probability distribution, \( \Pi(\emptyset)^{-1} = Z \) where \( Z \) is the normalization constant in (3.2). This proves the first part of the theorem.

For \( a \in A \) the formula (reformulation for \( V \)) becomes:
\[ V_A(x) = -\sum_{B \subseteq A \setminus a} (-1)^{|A-B|} \left[ \ln \left( \frac{\Pi(B_x)}{\Pi(B^{+a}x)} \right) \right. \] (3.3)
and this shows that \( V_A(x) = 0 \) if \( x_a = o_a \).

Now the local characteristics enter the game; for \( B \subset A \setminus \{a\} \) we have
\[ \frac{\Pi(x_a = B \, x_a \mid X_s = B \, x_s, \ s \neq a)}{\Pi(x_a = B + a \, x_a \mid X_s = B + a \, x_s, \ s \neq a)} = \frac{\Pi(B \, x)}{\Pi(B^{+a} \, x)}. \] (3.4)
In fact, the denominators of both conditional probabilities on the left-hand side coincide since only \( x_s \) for \( x \neq a \) appear. Plugging this relation into (3.3) yields (3.3.1). This completes the proof.

By (3.3.1),

**Corollary 3.3.1.** A random field is uniquely determined by the local characteristics for singletons.

A random field can now be represented as a Gibbs field for a suitable potential. A Markov field is even a neighbour Gibbs field for the original neighbourhood system. Given \( A \subset S \) the set \( \partial A \) of neighbours is the set \( \bigcup_{s \in A} \partial(s) \setminus A \).

**Theorem 3.3.2.** Let a neighbourhood system \( \partial \) on \( S \) be given. Then the following holds:
(a) A random field is a Markov field for \( \partial \) if and only if it is a neighbour Gibbs field for \( \partial \).
(b) For a Markov random field \( \Pi \) with neighbourhood system \( \partial \),
\[ \Pi \left( X_s = x_s, \ s \in A \mid X_s = x_s, \ s \in S \setminus A \right) = \Pi \left( X_s = x_s, \ s \in A \mid X_s = x_s, \ s \in \partial(A) \right) \]
for every subset \( A \) of \( S \).
In western literature, this theorem is frequently referred to as the Hammersley-Clifford theorem or the equivalence theorem. One early version is HAMMERSLEY and CLIFFORD (1968), but there are several independent papers in the early 70’s on this topic; cf. the literature in GRIMMETT (1975), AVERINTSEV (1978) and GEORGII (1988). The proof using Moebius inversion is due to G.R. GRIMMETT (1975).

**Proof (of the theorem).** A neighbour Gibbs field for \( \partial \) is a Markov field for \( \partial \) by proposition 3.2.1. This is one implication of (a). The same proposition covers assertion (b) for neighbour Gibbs fields.

To complete the proof of the theorem we must check the remaining implication of (a). Let \( \Pi \) be Markovian w.r.t \( \partial \) and let \( V \) be a potential for \( \Pi \) in the form (3.3.1). We must show that \( V_A \) vanishes whenever \( A \) is not a clique. To this end, suppose that \( A \) is not a clique. Then there is \( a \in A \) and \( b \in A \setminus \partial(a) \). Using (3.3), we rewrite the sum in (3.3.1) in the form

\[
V_A(x) = \sum_{B \subset A \setminus \{a,b\}} (-1)^{|A-B|} \ln \left( \frac{\Pi(X_a = B, X_s = B, s \neq a) \Pi(X_a = B + b, X_s = B + b, s \neq a)}{\Pi(X_a = B + a + b, X_s = B + a + b, s \neq a)} \right).
\]

Consider the first fraction in the last line: Since \( a \neq b \) we have \( \{X_a = B, X_s = B\} = \{X_a = B + b, X_s = B + b\} \); moreover, since \( b \notin \partial(a) \), the numerator and the denominator coincide by the very definition of a Markov random field. The same argument applies to the second fraction and hence the argument of the logarithm is 1 and the sum vanishes. This completes the proof of the remaining implication of (a) and thus the proof of the theorem. \( \square \)

We add some more information about potentials.

**Theorem 3.3.3.** The potential \( V \) given by (3.3.1) is the unique normalized potential for the Gibbs field \( \Pi \). A potential \( U \) for \( \Pi \) is related to \( V \) by

\[
V_A(x) = \sum_{B \subset A \setminus \partial_D \subset S} (-1)^{|A-B|} U_D(B, x).
\]

This shows for instance that normalization of pair potentials gives pair potentials.

**Proof.** Let \( U \) and \( W \) be normalized potentials for \( \Pi \). Since two energy functions for \( \Pi \) differ by a constant only and since \( H_U(o) = 0 = H_W(o) \) the two energy functions coincide. Let now any \( x \in X \) be given. For every \( s \in S \), we have
\[ U_{\{s\}}(x) = U_{\{s\}}(x^s) = H_U(x^s) = H_W(x^s) = W_{\{s\}}(x^s) = W_{\{s\}}(x). \]

Furthermore, for each pair \( s, t \in S, s \neq t, \)
\[ U_{\{s,t\}}(x) = U_{\{s,t\}}(x^{s,t}) = H_U(x^{s,t}) = U_{\{s\}}(x^{s,t}) - U_{\{t\}}(x^{s,t}). \]

The same holds for \( W. \) Since \( H_U(x^{s,t}) = H_W(x^{s,t}) \) and \( U_{\{s\}}(x^{s,t}) = U_{\{s\}}(x^{s,t}) \) we conclude that \( U_A = W_A \) whenever \( |A| = 2. \) Proceeding by induction over \( |A| \) shows that \( U = W. \)

Let now \( U \) be any potential for \( \Pi. \) Then for \( B \subset S \) and \( a \in S, \)
\[ \ln \frac{\Pi(B^x)}{\Pi(B^x + a)} = \sum_{D \subset S} (U_D(A^x) - U_D(A + a^x)). \]

Choose now \( A \subset S \) and \( a \in A. \) Then
\[ V_A(x) = \sum_{B \subset A - a} (-1)^{|A - B|} \ln \frac{\Pi(B^x)}{\Pi(B^x + a^x)} \]
\[ = \sum_{D \subset S} \sum_{B \subset A - a} (-1)^{|A - B|} U_D(B^x) - U_D(B^x + a^x) \]
\[ = \sum_{D \subset S} \sum_{B \subset A} (-1)^{|A - B|} U_D(B^x) \sum_{B' \subset A - D} (-1)^{|A - D| - |B'|}. \]

The first equality is (3.3.1), then the above equality is plugged in. Observing
\[ U_D(B^x) = U_D(B^D, x) \] gives the next identity. The last term vanishes except for \( A - D = \emptyset, \) i.e. \( A \subset D. \) This proves the desired identity. \( \square \)

**Corollary 3.3.2.** Two potentials \( U \) and \( U' \) determine the same Gibbs field if and only if
\[ \sum_{B \subset A \subset D \subset S} (-1)^{|A| - |B|} (U_D(B^x) - U'_D(B^x)) = 0 \]
for every \( A \neq \emptyset. \)

**Proof.** By uniqueness of normalized potentials, two potentials determine the same Gibbs field if and only if they have the same normalized potential. By the explicit representation in the theorem this is equivalent to the above identities. \( \square \)

The short survey by D. Griffieath (1976) essentially covers the previous material. Random fields for countable index sets \( S \) are introduced as well. S.D. Kindermann and J.L. Snell (1980) informally introduce to the physical ideas behind. French readers may consult PRUM (1986); there is also an English version PRUM and FORT (1991). Presently, the most comprehensive treatment is GEORGII (1988).