Markov Random Fields and Gibbs Measures

Oskar Sandberg

December 15, 2004

1 Introduction

A *Markov random field* is a name given to a natural generalization of the well known concept of a Markov chain. It arrises by looking at the chain itself as a very simple graph, and ignoring the directionality implied by "time". A Markov chain can then be seen as a chain graph of stochastic variables, where each variable has the property that it is independent of all the others (the future and past) given its two neighbors.

With this view of a Markov chain in mind, a Markov random field is the same thing, only that rather than a chain graph, we allow any graph structure to define the relationship between the variables. So we define a set of stochastic variables, such that each is independent of all the others given its neighbors in a graph.

Markov random fields can be defined both for discrete and more complicated valued random variables. They can also be defined for continuous index sets, in which case more complicated neighboring relationships take the place of the graph. In what follows, however, we will look only at the most approachable cases with discrete index sets, and random variables with finite state spaces (for the most part, the state space will simply be $\{0, 1\}$). For a more general treatment see for instance [Rozanov].

2 Definitions

2.1 Markov Random Fields

Let $X_1, ...X_n$ be random variables taking values in some finite set S, and let $G = (N, E)$ be a finite graph such that $N = \{1, ..., n\}$, whose elements will sometime be called *sites*. For a set $A \subset N$ let ∂A define its neighbor (or boundary) set: all elements in $N \backslash A$ that have a neighbor in A. For $i \in N$ let $\partial i = \partial\{i\}$.

The random variables are said to define a Markov random field if, for any vector $x \in S^N$:

$$\Pr(X_i = x_i \,|\, X_j = x_j, j \in N \backslash i) = \Pr(X_i = x_i \,|\, X_j = x_j, j \in \partial i) \tag{1}$$

2.2 Potentials

A *potential* is a function indexed by subsets of N on the space S^N. We will write potentials as $V_A(\omega)$ for $A \subset N$, $\omega \in S^N$.

Given a full set of potentials, the *energy* of a configuration ω will be defined as:

$$U(\omega) = -\sum_{A \subset N} V_A(\omega).$$

(Somewhat confusingly, this is called a potential in some texts (eg [Preston]), in which case V_A is called an "interaction potential". I will not use those terms here.)

Using the energy, we can define a probability measure, P, from a set of potentials by:

$$P(\omega) = \frac{\exp(-U(\omega))}{Z}$$

Where Z is the normalizer given by:

$$Z = \sum_{\omega \in S^N} \exp(-U(\omega)).$$

P is called a *Gibbs measure* (or *Gibbs state*).

The parallel between these random fields defined by potentials and Markov random fields, comes when one limits the type of potential to what is called a *nearest neighbor potential*. The parallel to the condition of (1) above that defines this property is that:

$$V_A(\omega) = 0 \tag{2}$$

whenever A is not a clique in the graph G. That is to say, V_A is identically zero when the subgraph induced by the A is not complete. As we shall see below, there is an equivalence between probability measures defined by such potentials and the laws of Markov random fields.

The concept of a *canonical potential* is important in this respect. Because of the normalization, there is not a unique energy or set of potentials associated with a probability measure, but by defining things correctly, we can define special potentials that are uniquely associated with a particular measure. Given some preferred element in S, denoted by here by 0, we define the *canonical energy* of a probability measure P as:

$$\tilde{U}(\omega) = -(\log P(\omega) - \log(P(\bar{0})))$$

where $\bar{0}$ is the zero vector over N. The canonical potential of the measure is then defined by:

$$\tilde{V}_A(\omega) = \sum_{B \subset A} -1^{|A-B|} \tilde{U}(\omega^B) \tag{3}$$

where ω^B is the configuration which takes the same values as ω on B, but sets all values to 0 elsewhere. We will show that the set of potentials defined by (3) really do correspond the probability measure P, and that they are nearest neighbor potentials.

3 Theory

Lemma 3.1 *If V_A for $A \subset N$ are nearest neighbors potentials, then:*

$$P(\omega) = \frac{\exp(-U(\omega))}{Z}$$

with U and Z as above, defines the law of a Markov random field.

Proof: By definition it follows that:

$$\begin{aligned} P(\omega(i) \,|\, \omega(j), j \in N \backslash \{i\}) &= \frac{P(\omega)}{\sum_{\eta(j)=\omega(j), j \neq i} P(\eta)} \\ &= \frac{\exp\left(\sum_{A \in N} V_A(\omega)\right)}{\sum_{\eta(j)=\omega(j), j \neq i} \exp\left(\sum_{A \in N} V_A(\eta)\right)} \end{aligned}$$

Now because V_A is a nearest neighbor potential, it is zero except when A is a clique. Any clique that does not contain i can be factored out identically in the nominator and divisor, and thus the last expression depends only the value of ω at i and its neighbors.

It follows that $P(\omega(i)|\omega(j), j \in N\backslash\{i\}) = P(\eta(i)|\eta(j), j \in N\backslash\{i\})$ if ω and η agree on i and ∂i. It follows from a straightforward computation of basic probability that:

$$P(\omega(i)|\omega(j), j \in N\backslash\{i\}) = P(\omega(i)|\omega(j), j \in \partial i)$$

and so P is the law of a Markov Random Field.

$\square$

If we instead start from a with the random field and its law P, we can find a potential:

Lemma 3.2 *For every measure P that is the law of a Markov random field on S^N there is a canonical potential that has P as its Gibbs measure.*

Proof: To prove this, all we need to do is to show that the canonical potential defined by (3) actually is a potential for P. That is to prove:

$$\begin{aligned}\tilde{U}(\omega) &= \sum_{A\subset N} \tilde{V}_A(\omega) \\ &= \sum_{A\subset N}\sum_{B\subset A} -1^{|A-B|}\tilde{U}(\omega^B).\end{aligned}$$

But reversing the order of summation gives that the latter expression is:

$$\begin{aligned}&= \sum_{B\subset N} \tilde{U^B}(\omega) \sum_{A\supset B}(-1)^{|A-B|} \\ &= \sum_{B\subset N} \tilde{U^B}(\omega) \sum_{j=0}^{|N-B|} \binom{|N-B|}{j}(-1)^j \\ &= \tilde{U}(\omega^N) = \tilde{U}(\omega)\end{aligned}$$

where last line follows since the inner sum on the line before is 0 whenever $|N - B| > 0$. This shows that:

$$P(\omega) = \frac{\exp\left(\sum_{A\in N} \tilde{V}_A(\omega)\right)}{Z}$$

as expected.

$\square$

So given any probability measure P, then we have a corresponding set of potentials. To set up the full correspondence between Markov random fields and potentials, we need then only prove that:

Lemma 3.3 *If P is the law of a Markov Random Field $(X_i)_{i \in N}$, then corresponding canonical potential $\tilde{V}_A$ is a nearest neighbor potential.*

Proof: We need to show that for any $A \subset N$ such that A is not a clique in the graph, $\tilde{V}_A(\omega) = 0$ for all ω.

Consider such a subset A, and take $i, j \in N$ as two sites that are not neighbors (ie $x \notin \partial y$). For brevity set $B = A \backslash \{i, j\}$. Then:

$$\begin{aligned} V_A(\omega) &= \sum_{C \subset A} (-1)^{|A-C|} \tilde{U}(\omega^C) \\ &= \sum_{D \subset B} \Big((-1)^{|A-D|+2} \tilde{U}(\omega^{D \cup \{i,j\}}) + (-1)^{|A-D|+1} \tilde{U}(\omega^{D \cup \{i\}}) \\ &\qquad + (-1)^{|A-D|+1} \tilde{U}(\omega^{D \cup \{j\}}) + (-1)^{|A-D|} \tilde{U}(\omega^D) \Big) \\ &= \sum_{D \subset B} (-1)^{|A-D|} \Big(\tilde{U}(\omega^{D \cup \{i,j\}}) - \tilde{U}(\omega^{D \cup \{i\}}) - \tilde{U}(\omega^{D \cup \{j\}}) + \tilde{U}(\omega^D) \Big). \end{aligned}$$

So the lemma holds if we can show that:

$$\Big(\tilde{U}(\omega^{D \cup \{i,j\}}) - \tilde{U}(\omega^{D \cup \{i\}}) - \tilde{U}(\omega^{D \cup \{j\}}) + \tilde{U}(\omega^D) \Big) = 0$$

for all D. This is equivalent to:

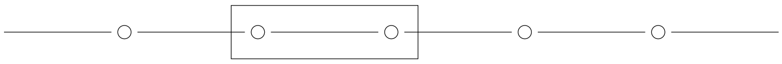
$$\frac{P(\omega^{D \cup \{i,j\}})}{P(\omega^{D \cup \{i\}})} = \frac{P(\omega^{D \cup \{j\}})}{P(\omega^D)}$$

which in turn is equivalent to:

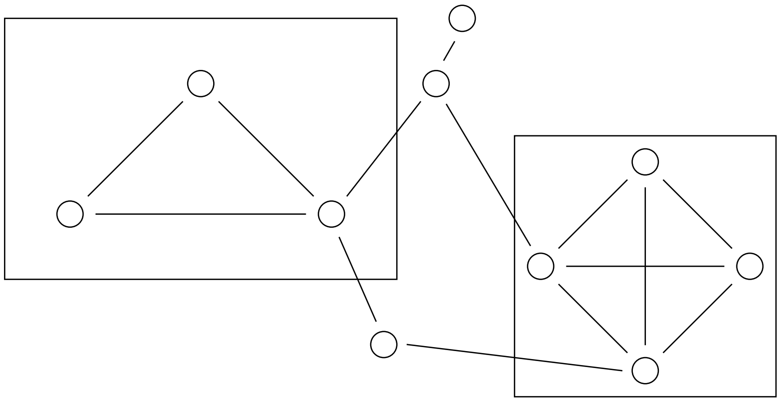
$$\frac{P(\omega^{D \cup \{i,j\}})}{P(\omega^{D \cup \{i,j\}}) + P(\omega^{D \cup \{i\}})} = \frac{P(\omega^{D \cup \{j\}})}{P(\omega^{D \cup \{j\}}) + P(\omega^D)}.$$

But this is conditional probability statement regarding the probability of site j being set to 0 or $\omega(j)$ conditioned on the configuration elsewhere. Since only the value at site i differs between the two sides, and i and j were not neighbors, it must be true if P is the law of a Markov Random Field.

$\square$



Figur 1: The structure of a Markov chain seen as a graph. Only pairs define cliques.



Figur 2: A more complicated graph on which a Markov random field could be defined. Two larger cliques are boxed.

So putting the above lemma's together yields the following theorem, originally due to J. M. Hammersley and P. Clifford. The proof presented above mostly follows that of [Preston] which also contains a discussion of its development.

Theorem 3.4 *For a probability measure P defined on a space S^N for finite sets S and N, where there is a graph defined as $G = (N, E)$. Then the following are equivalent:*

- *P is the law of a Markov Random Field.*
- *P is associated with a unique canonical potential $\tilde{V}_A(\omega)$, which is a nearest neighbor potential.*

4 Example

The simplest example of a Markov random field is to return to where we started, and look at a Markov chain in the above vocabulary. If we choose the simplest possible situation, that of a two state discrete time Markov chain on $\{0, 1\}$. We will look at the chain $(X_i)_{i=1...n}$, with an even initial distribution, and the transition matrix:

$$\begin{pmatrix} p & (1-p) \\ (1-p) & p \end{pmatrix}$$

For an outcome $x_1, ..., x_n$ of $X_1, ..., X_n$ of, set n_t to the number of 0-1 or 1-0 transitions, and n_r the number of times the process remains in its current state. We then have:

$$\begin{aligned} \Pr(X_1 = x_1, ..., X_n = x_n) &= \frac{1}{2}p^{n_t}(1-p)^{n_r} \\ &= \frac{1}{2}p^{n-1-n_r}(1-p)^{n_r} \\ &= \frac{1}{Z}\exp(\sum_{i=1}^{n} c\delta_{x_i}(x_i+1)) \end{aligned}$$

where Z and c are constants, and $\delta_y(x) = 1$ exactly when $y = x$. Since the pairs i, $i+1$ are the only cliques in the graph (see Figure 1), this shows that the law of Markov chain can be written as a Gibbs measure.

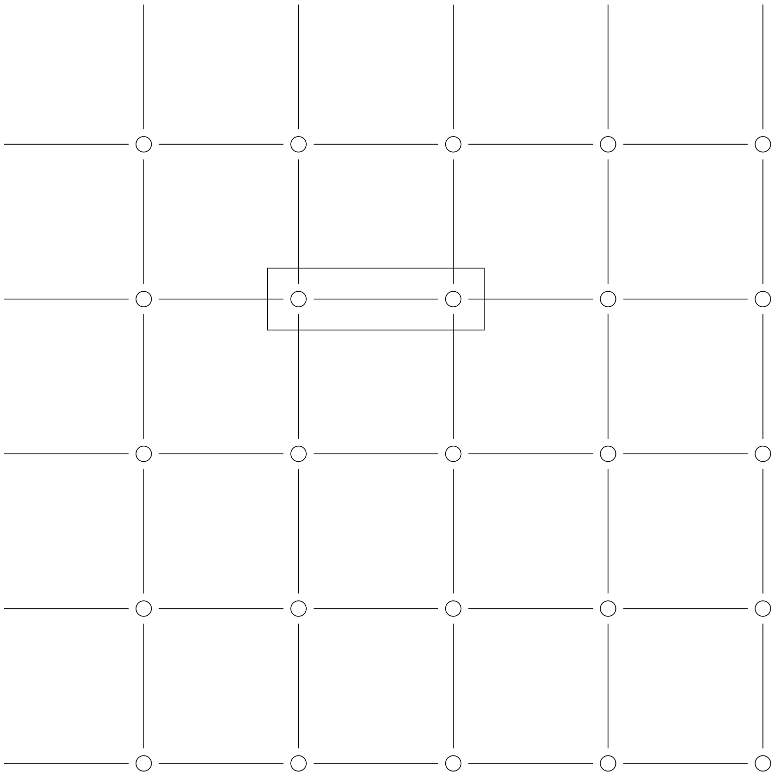
The potential defined by the delta functions above is, however, not the canonical potential. That can, however, be calculated directly from (3), over the different types of cliques:

$$\begin{aligned} V_{i,i+1}(x_1, x_n) &= \sum_{B \in \{2,3\}} \log \Pr(x_i = 1, i \in B, x_j = 0, j \notin B) \\ &= \log \Pr(\bar{0}) - \log \Pr(x_i = 1, x_j = 0, j \neq i) \\ &\quad - \log \Pr(x_{i+1} = 1, x_j = 0, j \neq i+1) \\ &\quad\quad + \Pr(x_i = 1, x_{i+1} = 1, x_j = 0, j \neq i, i+1) \\ &= 2\log\left(\frac{1-p}{p}\right) \end{aligned}$$

where the last line follows by direct computation. This assumes the value 1 at both sites, otherwise the value can easily be seen to be zero: it is in fact a (defining) property of the canonical potential it is zero whenever any site on the index set is 0. Through similar calculations one can see that:

$$V_i(x_1, ..., x_n) = -2\log\left(\frac{1-p}{p}\right)$$

for $i = 2, ..., n-1$ when the site takes the value one, and $V_0(x_1, ..., x_n) = V_1(x_1, ..., x_n) = -\log((1-p)/p)$. Adding up these potentials leads to the



Figur 3: The grid can define the graph of a Markov Random Field. Again the only cliques are pairs of elements. In the case of $S = \{-1, +1\}$ this the famous Ising model for magnetic spins.

energy and the Gibbs measure in the standard fashion. This can be seen to be same measure as above through calculation and noting that the potentials for two site cliques count the number of times we remain in state 1, and the potentials for one site clique count the number of times we are in state 1, whence their difference counts the number of transitions.

5 Bibliography

- R. Kinderman, J. L. Snell, *Markov Random Fields and their Applications*, AMS, Providence, 1980.
- C. J. Preston, *Gibbs States on Countable Sets*, Cambrige University Press, Cambrige, 1974.

- Yu. A. Rozanov, *Markov Random Fields*, Springer-Verlag, New York, 1982.