1. Markov random fields and Gibbs distributions

Let $\{X_t : t \in T\}$ be a finite collection of random variables—a stochastic process—with X_t taking values in a finite set $\mathcal{S}_t$. For simplicity of notation, suppose the index set T is $\{1, 2, \ldots, n\}$ and $\mathcal{S}_t = \{0, 1, \ldots, m_t\}$. The joint distribution of the variables is

$$\mathbb{Q}\{\mathbf{x}\} = \mathbb{P}\{X_t = x_t \text{ for } t \in T\} \qquad \text{where } \mathbf{x} = (x_1, \ldots, x_n),$$

with $0 \le x_t \le m_t$. More formally, the vector $\mathbf{X} = (X_1, \ldots, X_n)$ takes values in $\mathcal{X} = \prod_{t \in T} \mathcal{S}_t$, the set of all n-tuples $\mathbf{x} = (x_1, \ldots, x_n)$ with $x_t \in \mathcal{S}_t$ for each t.

Suppose T is the set of nodes of a graph. Let $\mathcal{N}_t$ denote the set of nodes (the neighbors of t) for which (t, s) is an edge of the graph.

<1> **Definition.** *The process is said to be a **Markov random field** if*

(i) $\mathbb{Q}\{\mathbf{x}\} > 0$ *for every* $\mathbf{x}$ *in* $\mathcal{X}$

(ii) for each t *and* $\mathbf{x}$,

$$\mathbb{P}\{X_t = x_t \mid X_s = x_s \text{ for } s \neq t\} = \mathbb{P}\{X_t = x_t \mid X_s = x_s \text{ for } s \in \mathcal{N}_s\}.$$

Property (ii) is equivalent to the requirement:

(ii)′ *the conditional probability* $\mathbb{P}\{X_t = x_t \mid X_s = x_s \text{ for } s \in \mathcal{N}_s\}$ *depends only on* x_s *for* $s \in \{t\} \cup \mathcal{N}_t$.

A subset A of T is said to be ***complete*** if each pair of vertices in A defines an edge of the graph. Write $\mathcal{C}$ for the collection of all complete subsets. clique?

<2> **Definition.** *The probability distribution* $\mathbb{Q}$ *is called a **Gibbs distribution** for the graph if it can be written in the form*

$$\mathbb{Q}\{\mathbf{x}\} = \prod\nolimits_{A \in \mathcal{C}} V_A(\mathbf{x}),$$

where each V_A *is a positive function that depends on* $\mathbf{x}$ *only through the coordinates* $\{x_t : t \in A\}$.

The Hammersley-Clifford Theorem asserts that the process $\{X_t : t \in T\}$ is a Markov random field if and only if the corresponding $\mathbb{Q}$ is a Gibbs distribution.

It is mostly a matter of bookkeeping to show that every Gibbs distribution defines a Markov random field.

<3> **Example.** With only a slight abuse of notation, we may write $V_A(\mathbf{x})$ as $V_A(x_{i_1}, \ldots, x_{i_k})$ if $A = \{i_1, \ldots, i_k\}$, ignoring the arguments that do not affect V_A. Suppose $\mathcal{N}_1 = \{2, 3\}$. Consider the variables x_j that actually appear in the conditional probability

$$\begin{aligned}
&\mathbb{P}\{X_1 = x_1 \mid X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n\} \\
&\qquad = \frac{\mathbb{P}\{X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n\}}{\mathbb{P}\{X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n\}} \\
&\qquad = \frac{\prod_{A \in \mathcal{C}} V_A(x_1, x_2, \ldots, x_n)}{\sum_w \prod_{A \in \mathcal{C}} V_A(w, x_2, \ldots, x_n)}
\end{aligned}$$

For example, which terms actually involve the value x_4? By assumption, V_A depends on x_4 only if $4 \in A$. For such an A, we cannot also have $1 \in A$, because then we would have $(1, 4)$ as an edge of the graph, contradicting the assumption that $4 \notin \mathcal{N}_1$. For concreteness, suppose $A = \{4, 7, 19\}$. Then $V_A(x_4, x_7, x_{19})$ appears once as a factor in the numerator and once as a factor in each summand in the denominator. It cancels from the ratio.

The only factors that do not cancel are those for which $1 \in A$. By definition of a complete subset, those factors can depend only on x_j values for $j \in \{1\} \cup \mathcal{N}_1$. □

It is slightly harder to show that every Markov random field corresponds to some Gibbs distribution. The simplest proof that I know depends on a general representation of a function as a sum of simpler functions.

<4> **Lemma.** *Let g be any real-valued function on $\mathcal{X}$. For each subset $A \subseteq \mathcal{S}$ define*

$$g_A(\mathbf{x}) = g(\mathbf{y}) \qquad \text{where } y_i = \begin{cases} x_i & \text{if } i \in A \\ 0 & \text{if } i \in A^c \end{cases}$$

and

$$\Psi_A(\mathbf{x}) = \sum\nolimits_{B \subseteq A} (-1)^{\#(A\backslash B)} g_B(\mathbf{x}).$$

Then

(i) the function Ψ_A depends on $\mathbf{x}$ only through those coordinates x_j with $j \in A$ (in particular, $\Psi_\emptyset$ is a constant)

(ii) for $A \neq \emptyset$, if $x_i = 0$ for at least one i in A then $\Psi_A(\mathbf{x}) = 0$

(iii) $g(\mathbf{x}) = \sum_{A \subseteq T} \Psi_A(\mathbf{x})$ x can be looked as a constant in this equal

Proof. Assertion (i) is trivial: every g_B appearing in the definition of $\Psi_A(\mathbf{x})$ does not depend on the variables $\{x_j : j \notin A\}$.

For (ii), divide the subsets of A into two subcollections: those that contain i and those that do not contain i. For each B of the first type there is a unique set, $\tilde{B} = \{i\} \cup B$, of the second type. Note that $g_B(\mathbf{x}) = g_{\tilde{B}}(\mathbf{x})$ because $x_i = 0$. The contributions to $\Psi_A(\mathbf{x})$ from the sets $B, \tilde{B}$ cancel, because one of the two numbers $\#A\backslash B$ and $\#A\backslash\tilde{B}$ is odd and the other is even.

For (iii), note that the coefficient of g_B in the double sum

<5> $$\sum\nolimits_{A \subseteq T} \Psi_A(\mathbf{x}) = \sum\nolimits_{A \subseteq T} \sum\nolimits_{B \subseteq A} (-1)^{\#(A\backslash B)} g_B(\mathbf{x})$$

equals

$$\sum\nolimits_A \{B \subseteq A \subseteq T\}(-1)^{\#(A\backslash B)} = \sum\nolimits_{E \subseteq B^c} (-1)^{\#E}.$$ for any B < T, the sum = 1 + (-1)C_{#B^c}^1 +

For B equal to T, the last sum reduces to $(-1)^0$ because $\emptyset$ is the only subset of T^c. For $B^c \neq \emptyset$, half of the subsets E have $\#E$ even and the other half have $\#E$ odd, which reduces the coefficient to 0. Thus the double sum <5> simplifies to $g_T(\mathbf{x}) = g(\mathbf{x})$. □

Applying the Lemma with $g(\mathbf{y}) = \log \mathbb{Q}\{\mathbf{y}\}$ gives

<6> $$\mathbb{Q}\{\mathbf{x}\} = \exp\left(\sum\nolimits_{A \subseteq \mathbb{I}} \Psi_A(\mathbf{x})\right)$$

To show that the expression on the right-hand side is a Gibbs distribution, we have only to prove that $\Psi_A(\mathbf{x}) \equiv 0$ when A is not a complete subset of the graph.

<7> **Theorem.** *For a Markov random field, the term Ψ_A in <6> is identically zero if A is not a complete subset of T.*

Proof. For simplicity of notation, suppose $1, 2 \in A$ but nodes 1 and 2 are not connected by an edge of the graph, that is, they are not neighbors. Consider the contributions to $\Psi_A(\mathbf{x})$ from pairs $B, \tilde{B}$, where $1 \notin B$ and $\tilde{B} = B \cup \{1\}$. The numbers $\#A\backslash B$ and $\#A\backslash\tilde{B}$ differ by 1; the pair contributes $\pm\left(g_{\tilde{B}}(\mathbf{x}) - g_B(\mathbf{x})\right)$ to the sum. Define

$$y_i = \begin{cases} x_i & \text{if } i \in B \\ 0 & \text{if } i \in B^c \end{cases}$$

Then

$$g_{\tilde{B}}(\mathbf{x}) - g_B(\mathbf{x}) = \log \frac{\mathbb{P}\{X_1 = x_1, X_2 = y_2, \dots, X_n = y_n\}}{\mathbb{P}\{X_1 = 0, X_2 = y_2, \dots, X_n = y_n\}}$$
$$= \log \frac{\mathbb{P}\{X_1 = x_1 \mid X_2 = y_2, \dots, X_n = y_n\}}{\mathbb{P}\{X_1 = 0 \mid X_2 = y_2, \dots, X_n = y_n\}}$$

A common factor of $\mathbb{P}\{X_2 = y_2, \dots, X_n = y_n\}$ has cancelled from numerator and denominator.

The Markov property ensures that the conditional probabilities in the last ratio do not depend on the value y_2. The ratio is unchanged if we replace y_2 by 0. The same argument works for every $B, \tilde{B}$ pair. Thus $\Psi_A(\mathbf{x})$ is unchanged if we put x_2 equal to 0. From Lemma <4> (ii), deduce that $\Psi_A(\mathbf{x}) = 0$ for all $\mathbf{x}$, as asserted. □

References

Griffeath, D. (1976), *Introduction to Markov Random Fields*, Springer. Chapter 12 of *Denumerable Markov Chains* by Kemeny, Knapp, and Snell (2nd edition).