Chapter

Markov random fields and Gibbs measures

1. Conditional independence

Suppose X_i is a random element of $(\mathcal{X}_i, \mathcal{B}_i)$, for $i = 1, 2, 3$, with all X_i defined on the same probability space $(\Omega.\mathcal{F}, \mathbb{P})$. The random elements X_1 and X_3 are said to be ***conditionally independent given*** X_2 if, at least for all bounded, $\mathcal{B}_i$-measuable real functions f_i,

$$\mathbb{P}\left(f_1(X_1) f_3(X_3) \mid X_2\right) = \mathbb{P}\left(f_1(X_1) \mid X_2\right) \mathbb{P}\left(f_3(X_3) \mid X_2\right) \qquad \text{almost surely.}$$

Equivalently, with $F_i(X_2) = \mathbb{P}\left(f_i(X_i) \mid X_2\right)$ for $i = 1, 3$,

<1>
$$\mathbb{P} f_1(X_1) f_2(X_2) f_3(X_3) = \mathbb{P} F_1(X_2) f_2(X_2) F_3(X_2) \qquad \text{for all } f_i.$$

Notice that equality <1> involves the random elements only through their joint distribution. We could, with no loss of generality, assume $\Omega = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ equipped with its product sigma-field $\mathcal{B} = \mathcal{B}_1 \otimes \mathcal{B}_2 \otimes \mathcal{B}_3$, with $\mathbb{P}$ the joint distribution of (X_1, X_2, X_3) and $X_i(\omega) = \omega_i$, the ith coordinate map.

Things are greatly simplified if $\mathbb{P}$ has a density $p(\omega)$ with respect to a sigma finite product measure $\lambda = \lambda_1 \otimes \lambda_2 \otimes \lambda_3$ on $\mathcal{B}$. As in the case of densities with respect to Lebesgue measure, the marginal distributions have densities obtained by integrating out some coordinates. For example, the marginal distribution of ω_2 has density

$$p_2(\omega_2) = \lambda_1 \otimes \lambda_3 p(\omega_1, \omega_2, \omega_3) \qquad \text{with respect to } \lambda_2.$$

Similarly, the various conditional distributions are given by conditional densities (cf. Pollard 2001, Section 5.4). For example, the conditional distribution of (ω_1, ω_3) given ω_2 has density

$$p_{13|2}(\omega_1, \omega_3 \mid \omega_2) = \frac{p(\omega_1, \omega_2, \omega_3)}{p_2(\omega_2)} \{p_2(\omega_2) > 0\} \qquad \text{with respect to } \lambda_1 \otimes \lambda_3.$$

REMARK. It is traditional to use less formal notation, for example, writing $p(\omega_1, \omega_3 \mid \omega_2)$ instead of $p_{13|2}(\omega_1, \omega_3 \mid \omega_2)$. As long as the arguments are specified symbolically there is no ambiguity. But an expression like $p(a, b \mid c)$ could refer to several different conditional densities evaluated at values (a, b, c).

The conditional expectations are given by integrals involving conditional densities. For example,

$$\mathbb{P}\left(f_1(\omega_1) \mid \omega_2\right) = \lambda_1 f_1(\omega_1) p_{1|2}(\omega_1 \mid \omega_2) \qquad \text{almost surely.}$$

Equality <1> then becomes

$$
\begin{aligned}
&\lambda f_1(\omega_1) f_2(\omega_2) f_3(\omega_3) p(\omega_1, \omega_2, \omega_3) \\
&\quad = \lambda_2 f_2(\omega_2) p_2(\omega_2) \big(\lambda_1 f_1(\omega_1) p_{1|2}(\omega_1 \mid \omega_2)\big) \big(\lambda_1 f_3(\omega_3) p_{3|2}(\omega_3 \mid \omega_2)\big) \\
&\quad = \lambda f_1(\omega_1) f_2(\omega_2) f_3(\omega_3) p_2(\omega_2) p_{1|2}(\omega_1 \mid \omega_2) p_{3|2}(\omega_3 \mid \omega_2)
\end{aligned}
$$

A simple generating-class argument then shows that the conditional independence is equivalent to the factorization

<2> $$p(\omega_1, \omega_2, \omega_3) = p_2(\omega_2) p_{1|2}(\omega_1 \mid \omega_2) p_{3|2}(\omega_3 \mid \omega_2) \qquad \text{a.e. } [\lambda]$$

Note that the right-hand side of the last display has the form $\phi(\omega_1, \omega_2)\psi(\omega_2, \omega_3)$. Actually, any factorization

<3> $$p(\omega_1, \omega_2, \omega_3) = \phi(\omega_1, \omega_2)\psi(\omega_2, \omega_3)$$

for nonnegative ϕ and ψ implies the conditional independence. For, if we define $\Phi(\omega_2) = \lambda_1 \phi(\omega_1, \omega_2)$ and $\Psi(\omega_2) = \lambda_3 \psi(\omega_2, \omega_3)$, we then have

$$
\begin{aligned}
p_2(\omega_2) &= \lambda_1 \otimes \lambda_3 p = \Phi(\omega_2)\Psi(\omega_2) \\
p_{1,2}(\omega_1, \omega_2) &= \lambda_3 p = \phi(\omega_1, \omega_2)\Psi(\omega_2) \\
p_{2,3}(\omega_2, \omega_3) &= \lambda_1 p = \Phi(\omega_2)\psi(\omega_2, \omega_3) \\
p_{1|2}(\omega_1 \mid \omega_2) &= \phi(\omega_1, \omega_2)/\Phi(\omega_2)\{p_2 > 0\} \\
p_{3|2}(\omega_3 \mid \omega_2) &= \psi(\omega_2, \omega_3)/\Psi(\omega_2)\{p_2 > 0\}
\end{aligned}
$$

from which <2> follows.

Similar arguments apply when $\mathbb{P}$ is the joint distribution for more than three random elements.

> **Notation.** Let S be a finite index set. (Later the points of S will be called ***sites*** at which random variables are defined.) Consider probability densities $p(\omega_i : i \in S)$ with respect to a product measure $\lambda = \bigotimes_{i \in S} \lambda_i$ on the product sigma-field $\mathcal{B} = \bigotimes_{i \in S} \mathcal{B}_i$ for a product space $\Omega = \times_{i \in S} \mathcal{X}_i$. For each $A \subseteq S$ write ω_A for $(\omega_i : i \in A)$.

If A, B, and C are disjoint subsets with union S, then it follows by an argumnet similar to the one leading from <3> that ω_A and ω_B are conditionally independent given ω_C if and only if the density factorizes as

<4> $$p(\omega) = \phi(\omega_A, \omega_C)\psi(\omega_B, \omega_C)$$

In fact, this condition follows directly from the earlier argument if we take $X_1 = \omega_A$, $X_3 = \omega_B$, and $X_2 = \omega_C$.

Better: The rhs of <4> is a version of p.

If A, B, and C are disjoint subsets whose union is not the whole of S, the extra variables can complicate the checking of conditional independence. However a sufficient condition for conditional independence of ω_A and ω_B given ω_C is existence of a factorization

<5> $$p(\omega) = \phi(\omega_A, \omega_C, \omega_D)\psi(\omega_B, \omega_C, \omega_E)$$

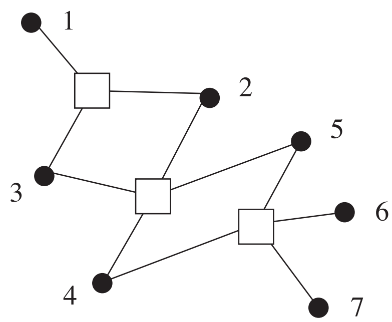
where D and E are disjoint subsets of $S \backslash \big(A \cup B \cup C\big)$. The integration over ω_D and ω_E, which is needed to find the density for $\omega_{A \cup B \cup C}$, preserves the factorization needed for conditional independence.

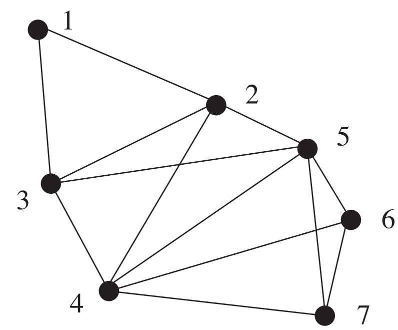
2. Gibbs distributions

There is a particularly easy way to create probability measures with recognizable conditional independence properties. Let $\mathbb{F}$ be a collection of subsets of S. For each a in $\mathbb{F}$, let $\Psi_a(\omega) = \Psi_a(\omega_a)$ be a nonnegative function that depends on ω only through ω_a. Provided the number $Z := \lambda \prod_{a\in\mathbb{F}} \Psi_a$ is neither zero nor infinite, we can define a ***Gibbs measure*** by means of the density

<6> $$p(\omega) = \frac{1}{Z} \prod_{a\in\mathbb{F}} \Psi_a(\omega_a) \qquad \text{with respect to } \lambda.$$

The conditional independence properties are easily seen from the corresponding ***factor graph***, which has vertex set $\mathbb{V} = S \cup \mathbb{F}$ with edges drawn only between those $i \in S$ and $a \in \mathbb{F}$ for which $i \in a$. For example, if $S = \{1, 2, 3, 4, 5, 6, 7\}$ and $\mathbb{F} = \{\{1, 2, 3\}, \{2, 3, 4, 5\}, \{4, 5, 6, 7\}\}$:





The same connectivities could be represented, less economically, without the $\mathbb{F}$ vertices by joining all sites i, j for which there is some $a \in \mathbb{F}$ with $\{i, j\} \subseteq a$, as shown in the graph on the right of the display.

If i and j share a common neighbor a in the factor graph (equivalently, if i and j are neighbors in the representation without $\mathbb{F}$), write $i \sim j$. Write $\mathcal{N}_i = \{j \in S\backslash\{i\} : i \sim j\}$.

<7> **Theorem.** *Suppose A, B, and C are disjoint subsets of S with the separation property: each path joining a site in A to a site in B must pass through some site in C. Then ω_A and ω_B are conditionally independent given ω_C under the Gibbs measure given by the density <6>.*

Proof. Define $\mathbb{V}_A$ as the set of vertices from $\mathbb{V}$ that can be connected to some site in A by a path that passes through no sites from C. By the separation assumption, $A \subseteq \mathbb{V}_A$ and $B \subseteq \mathbb{V}_A^c$. Sites in C might belong to either $\mathbb{V}_A$ or $\mathbb{V}_A^c$. Note that

$$p(\omega) = \phi(\omega)\psi(\omega) \qquad \text{where } \phi = \prod_{a\in\mathbb{V}_A} \Psi_a \text{ and } \psi = \prod_{a\notin\mathbb{V}_A} \Psi_a.$$

The function ϕ depends on ω only through the sites in $\mathbb{V}_A$. Conservatively, we could write it as $\phi(\omega_A, \omega_C, \omega_D)$, where D denotes the set of all sites in $\mathbb{V}_A\backslash\left(A \cup C\right)$. Similarly, $\psi(\omega) = \psi(\omega_B, \omega_C, \omega_E)$ where $E \subseteq S\backslash\left(B \cup C\right)$ consists of all those sites i connected to some $a \in \mathbb{V}_A^c$. There can be no path joining i to a point k in A without passing through C, for otherwise there would be a path from k to a via i that would force a to belong to $\mathbb{V}_A$. In other words, D and E are disjoint subsets of $S\backslash\left(A \cup B \cup C\right)$. That is, we have a factorization of p as in <5>, which implies the asserted conditional independence. □

<8> **Exercise.** For each site i , show that the conditional distribution of ω_i given $\omega_{S\backslash i}$ depends only on $(\omega_j; j \in \mathcal{N}_i)$. Specifically, show that

$$p_{i|S\backslash i}(\omega_i \mid \omega_{S\backslash i}) =$$

□ This fact is called the ***Markov property*** for the Gibbs distribution.

3. Hammersley-Clifford

When $\mathbb{P}$ is defined by a strictly positive density $p(\omega)$ with respect to λ, there is a representation for $\mathbb{P}$ as a Gibbs measure with a factor graph that represents any Markov properties that $\mathbb{P}$ might have. The representation will follow from an application of the following Lemma to $\log p(\omega)$.

<9> **Lemma.** *Let g be any real-valued function on $\Omega = \times_{i\in S}\mathcal{X}_i$. Let $\widetilde{\omega}$ be any arbitrarily chosen, but fixed, point of Ω. There exists a representation $g(\omega) = \sum_{a\subseteq S}\Phi_a(\omega)$ with the following properties.*

(i) $\Phi_a(\omega) = \Phi_a(\omega_a)$ depends on ω only through the coordinates ω_a. In particular, $\Phi_\emptyset$ is a constant.

(ii) If $a \neq \emptyset$ and if $\omega_i = \widetilde{\omega}_i$ for any i in a then $\Phi_a(\omega) = 0$.

Proof. For each subset $a \subseteq S$ define

$$g_a(\omega_a) = g_a(\omega) = g(\omega_a, \widetilde{\omega}_{S\backslash a})$$

and

$$\Phi_a(\omega) = \sum\nolimits_{b\subseteq a}(-1)^{\#(a\backslash b)}g_b(\omega_b).$$

Clearly Φ depends on ω only through ω_a.

To prove (ii), for simplicity suppose $1 \in a$ and $\omega_1 = \widetilde{\omega}_1$. Divide the subsets of a into a set of pairs: those b for which $1 \notin b$ and the corresponding $b' = b \cup \{1\}$. The subsets b and b' together contribute

$$(-1)^{\#(a\backslash b)}\left(g(\omega_b, \widetilde{\omega}_1, \widetilde{\omega}_{S\backslash b'}) - g(\omega_b, \omega_1, \widetilde{\omega}_{S\backslash b'})\right) = 0 \qquad \text{because } \omega_1 = \widetilde{\omega}_1.$$

Finally, to establish the representation for g, consider the coefficient of g_b in

$$\sum\nolimits_{a\subseteq S}\Phi_a(\omega) = \sum\nolimits_{a,b}\{b \subseteq a \subseteq S\}(-1)^{\#(a\backslash b)}g_b(\omega_b)$$

When $b = a = S$, we recover $(-1)^0 g_S(\omega) = g(\omega)$. When b is a proper subset of S, the coefficient of g_b equals

$$\sum\nolimits_a\{b \subseteq a \subseteq S\}(-1)^{\#(a\backslash b)} = \sum\nolimits_{D\subseteq S\backslash b}(-1)^{\#D},$$

which equals zero because half of the subsets D of $S\backslash b$ have $\#D$ even and the other half have $\#D$ odd. □

Applying the Lemma with $g(\omega) = \log p(\omega)$ gives

$$p(\omega) = \prod\nolimits_{a\subseteq S}\exp\Phi_a(\omega_a),$$

which is a trivial sort Gibbs density. We could, of course, immediately discard any terms for which $\Phi_a(\omega) = 0$ for all ω. The factorizations that follow from any conditional independndce properties of $\mathbb{P}$ lead in this way to more interesting representations.

> The following argument definitely works when Ω is finite and λ is counting measure. I am not so confident for the general case because I haven't thought carefully enough about the role of negligible sets and versions.

<10> **Lemma.** *Suppose ω_i and ω_j are conditionally independent given ω_C, where $C = S\backslash\{i, j\}$. Then $\Phi_a \equiv 0$ for every a with $\{i, j\} \subseteq a \subseteq S$.*

Proof. Invoke equality <4> with $A = \{i\}$ and $B = \{j\}$ to get

$$\log p(\omega) = f(\omega_i, \omega_C) + h(\omega_j, \omega_C) \qquad \text{where } f = \log\phi \text{ and } h = \log\psi.$$

Notice that the representation for Φ_a from Lemma <9> is linear in g. Thus $\Phi_a = \Phi_a^f + \Phi_a^h$, the sum of the terms by applying the same construction to f then to h. For example,

$$\Phi_a^f(\omega_a) = \sum\nolimits_{b \subseteq a} (-1)^{\#a \backslash b} f(\omega_b, \widetilde{\omega}_{S \backslash b})$$

If $j \in a$, the right-hand side is unaffected if we replace ω_j by $\widetilde{\omega}_j$, because f does not depend on ω_j. However, the left-hand side is zero when $\omega_j = \widetilde{\omega}_j$. It follows that $\Phi_a^f(\omega_a) \equiv 0$ when $j \in a$. A similar argument shows that $\Phi_a^h(\omega_a) \equiv 0$ when $i \in a$. $\square$

References

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

Pollard, D. (2001), *A User's Guide to Measure Theoretic Probability*, Cambridge University Press.