Chapter

Markov random fields and Gibbs measures

1. Notation

Let S be a finite index set. (Later the points of S will be called ***sites*** at which random variables are defined.) For each i in S, suppose $\mathcal{X}_i$ is a set equippped with a sigma-field $\mathcal{B}_i$. Let $\Omega = \times_{i \in S} \mathcal{X}_i$. Equip Ω with the product sigma-field $\mathcal{B} = \bigotimes_{i \in S} \mathcal{B}_i$.

For each $A \subseteq S$ write ω_A for $(\omega_i : i \in A)$.

We will mostly be concerned with probability measures dominated by a fixed, sigma-finite product measure $\lambda = \bigotimes_{i \in S} \lambda^i$ on $\mathcal{B}$. For each $A \subseteq S$, the symbol λ^A will denote the product measure $\bigotimes_{i \in A} \lambda^i$ on $\mathcal{B}_A = \bigotimes_{i \in A} \mathcal{B}_i$. I will also write $\lambda^{B,C}$ for $\lambda^{B \cup C}$ if B and C are disjoint subsets of S.

Similarly, for each subset A of S write $\mathbb{P}^A$ for the marginal distribution of ω_A, that is, $\mathbb{P}^A f(\omega_A) = \mathbb{P} f(\omega_A)$, where the second $f(\omega_A)$ is regarded as a function of ω that happens not to depend on the coordinates $\omega_{S \setminus A}$. Note that this notation should not be interpreted to mean that $\mathbb{P}$ is a product measure.

For disjoint subsets A and B of S, the (Kolmogorov) conditional expectation operator will be denoted by $\mathbb{P}_A^B$. That is, at least for bounded, $\mathcal{B}_B$-measurable g, the assertion $G(\omega_A) = \mathbb{P}_A^B g(\omega_B)$ will mean that

$$\mathbb{P} f(\omega_A) g(\omega_B) = \mathbb{P} f(\omega_A) G(\omega_A)$$

at least for bounded $\mathcal{B}_A$-measurable f. If a conditional distribution for ω_B given ω_A exists, I will denote it also by $\mathbb{P}_A^B$. In that case, $\mathbb{P}_A^B(\cdot \mid \omega_A)$ is a probability measure on $\mathcal{B}_B$ for $\mathbb{P}^A$ almost all ω_A.

For these notes, the most important case will have each $\mathcal{X}_i$ finite with λ_i equal to counting measure on $\mathcal{X}_i$. Many, but not all, of the subtleties involving negligible sets with respect to various measures will then disappear.

2. Conditional independence

Suppose $\mathbb{P}$ is a probability measure on the product sigma-field $\mathcal{B}$ for the product space Ω. For disjoint subsets A, B, and C of S, the random elements ω_A and ω_B are said to be ***conditionally independent given*** ω_C if, at least for all bounded, product-measurable real functions f and g,

$$\mathbb{P}\big(f(\omega_A) g(\omega_B) \mid \omega_C\big) = \mathbb{P}\big(f(\omega_A) \mid \omega_C\big)\, \mathbb{P}\big(g(\omega_B) \mid \omega_C\big) \qquad \text{a.e. } [\mathbb{P}].$$

Equivalently, with $F(\omega_C)$ as any version of $\mathbb{P}\big(f(\omega_A) \mid \omega_C\big)$ and $G(\omega_C)$ as any version of $\mathbb{P}\big(g(\omega_B) \mid \omega_C\big)$,

<1> $$\mathbb{P} f(\omega_A) g(\omega_B) h(\omega_C) = \mathbb{P} F(\omega_C) G(\omega_C) h(\omega_C)$$

for all bounded, measurable f, g, and h.

Things are greatly simplified if $\mathbb{P}$ has a density $p(\omega)$ with respect to λ. As in the case of densities with respect to Lebesgue measure, we can obtain versions of the densities for the marginal distributions by integrating out some coordinates. For example, the marginal distribution of $\mathbb{P}^C$ has density

$$p_C(\omega_C) = \lambda^{S\backslash C} p(\omega_C, \omega_{S\backslash C}) \qquad \text{with respect to } \lambda^C.$$

Here the $\lambda^{S\backslash C}$ integrates out over the coordinates $\omega_{S\backslash A}$. The density p_C is unique only up to a λ^C equivalence.

Similarly, versions of the various conditional distributions can be specified by conditional densities (cf. Pollard 2001, Section 5.4). For example, the conditional distribution $\mathbb{P}_C^B$ of ω_B given ω_C has density

$$p_{B|C}(\omega_B \mid \omega_C) = \frac{p(\omega_B, \omega_C)}{p_C(\omega_C)}\{p_C(\omega_C) > 0\} \qquad \text{with respect to } \lambda^B.$$

This density is unique only up to a $\lambda^B \otimes \mathbb{P}^C$-equivalence. For example, we could change the definition arbitrarily on the $\{p_C = 0\}$ without disturbing the defining property of the conditional distribution. If $\lambda^C\{p_C = 0\} > 0$, there can be no hope of having the conditional density unique up to a λ^C-equivalence.

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

The conditional expectations can be written as integrals involving conditional densities. For example,

$$\mathbb{P}\big(f(\omega_A) \mid \omega_C\big) = \lambda^A f(\omega_A) p_{A|C}(\omega_A \mid \omega_C) \qquad \text{a.e. } [\mathbb{P}^C].$$

Similarly,

$$F(\omega_C)G(\omega_C) = \lambda^{A,B} f(\omega_A) g(\omega_B) p_{A|C}(\omega_A \mid \omega_C) p_{B|C}(\omega_B \mid \omega_C) \qquad \text{a.e. } [\mathbb{P}^C]$$

so that equality <1> becomes

$$\begin{aligned}
&\lambda f(\omega_A) g(\omega_B) h(\omega_C) p(\omega_A, \omega_B, \omega_C) \\
&\qquad = \mathbb{P} f(\omega_A) g(\omega_B) h(\omega_C) \\
&\qquad = \mathbb{P}^C F(\omega_C) G(\omega_C) h(\omega_C) \\
&\qquad = \lambda f(\omega_A) g(\omega_B) h(\omega_C) p_C(\omega_C) p_{A|C}(\omega_A \mid \omega_C) p_{B|C}(\omega_B \mid \omega_C).
\end{aligned}$$

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

<2> $$p(\omega_A, \omega_B, \omega_C) = p_C(\omega_C) p_{A|C}(\omega_A \mid \omega_C) p_{B|C}(\omega_B \mid \omega_C) \qquad \text{a.e. } [\lambda]$$

or to

<3> $$p_{A,B|C}(\omega_A, \omega_B \mid \omega_C) = p_{A|C}(\omega_A \mid \omega_C) p_{B|C}(\omega_B \mid \omega_C) \qquad \text{a.e. } [\lambda^{A,B} \otimes \mathbb{P}^C]$$

because λ^C and $\mathbb{P}^C$ have the same negligible subsets of $\{\omega_C : p_C(\omega_C) > 0\}$ and $p(\omega_A, \omega_B, \omega_C) = 0$ a.e. $[\lambda^{A,B}]$ on the set $\{\omega_C : p_C(\omega_C) = 0\}$.

Note that the right-hand side of <2> is a product of a function of (ω_A, ω_C) and a function of (ω_B, ω_C). Actually, any factorization

<4> $$p(\omega_A, \omega_B, \omega_C) = \phi(\omega_A, \omega_C) \psi(\omega_B, \omega_C) \qquad \text{a.e. } [\lambda^{A,B} \otimes \mathbb{P}^C]$$

for nonnegative ϕ and ψ implies the conditional independence of ω_A and ω_B given ω_C. For, if we define $\Phi(\omega_C) = \lambda^A \phi(\omega_A, \omega_C)$ and $\Psi(\omega_C) = \lambda^B \psi(\omega_B, \omega_C)$, we then have

$$\begin{aligned} p_C(\omega_C) &= \lambda^{A,B} p = \Phi(\omega_C)\Psi(\omega_C) && \text{a.e. } [\mathbb{P}^C] \\ p_{A,C}(\omega_A, \omega_C) &= \lambda^B p = \phi(\omega_A, \omega_2)\Psi(\omega_2) && \text{a.e. } [\lambda^A \otimes \mathbb{P}^C] \\ p_{B,C}(\omega_B, \omega_C) &= \lambda^A p = \Phi(\omega_C)\psi(\omega_B, \omega_C) && \text{a.e. } [\lambda^B \otimes \mathbb{P}^C] \end{aligned}$$

from which it follows that, except on a $\lambda^{A,B} \otimes \mathbb{P}^C$-negligible set,

$$\begin{aligned} & p_{A|C}(\omega_A \mid \omega_C) p_{B|C}(\omega_B \mid \omega_C) \\ &\quad = \frac{\phi(\omega_A, \omega_C)}{\Phi(\omega_C)} \frac{\psi(\omega_B, \omega_C)}{\Psi(\omega_C)} \{\Phi > 0, \Psi > 0\} \\ &\quad = p_{A,B|C}(\omega_A, \omega_B \mid \omega_C), \end{aligned}$$

as required by the condition <3> for conditional independence.

Notice that we also have,

$$\begin{aligned} p_{A|C}(\omega_A \mid \omega_C) &= \frac{\phi(\omega_A, \omega_C)}{\Phi(\omega_C)} \{\Phi > 0\} && \text{a.e. } [\lambda^A \otimes \mathbb{P}^C] \\ p_{B|C}(\omega_A \mid \omega_C) &= \frac{\psi(\omega_B, \omega_C)}{\Psi(\omega_C)} \{\Psi > 0\} && \text{a.e. } [\lambda^A \otimes \mathbb{P}^C]. \end{aligned}$$

The extra indicator functions are not needed when we do not divide by the corresponding Φ or Ψ factors. By similar reasoning we have

$$p_{B,C}(\omega_B, \omega_C) = \Phi(\omega_C)\psi(\omega_B, \omega_C) \qquad \text{a.e. } [\lambda^{B,C}]$$

and

$$\begin{aligned} & p_{A|B,C}(\omega_A \mid \omega_B, \omega_C) \\ &\quad = \frac{p(\omega)}{p_{B,C}(\omega_b, \omega_C)} \{p_{B,C} > 0\} && \text{a.e. } [\lambda^A \otimes \mathbb{P}^{B,C}] \\ &\quad = \frac{\phi(\omega_A, \omega_C)}{\Phi(\omega_C)} \{\Phi(\omega_C) > 0\} && \text{a.e. } [\lambda^A \otimes \mathbb{P}^{B,C}] \end{aligned}$$

<5> $$\quad = p_{A|C}(\omega_A \mid \omega_C) \qquad \text{a.e. } [\lambda^A \otimes \mathbb{P}^C]$$

because $\mathbb{P}^{B,C}\{\psi(\omega_B, \omega_C) = 0\} = \lambda^{B,C}\psi\{\psi = 0\} = 0$.

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

<6> $$p(\omega) = \phi(\omega_A, \omega_C, \omega_D)\psi(\omega_B, \omega_C, \omega_E) \qquad \text{a.e. } [\lambda^{A,B} \otimes \lambda^{C,D} \otimes \mathbb{P}^C]$$

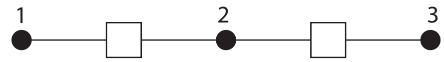
where D and E are disjoint subsets of $S \backslash \left(A \cup B \cup C\right)$. The integration over ω_D and ω_E, which is needed to find the density for $(\omega_A, \omega_B, \omega_C)$, preserves the factorization needed for conditional independence.

3. 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, measurable 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*** $\mathbb{P}$ by means of the density

<7> $$\frac{d\mathbb{P}}{d\lambda} = p(\omega) = \frac{1}{Z} \prod_{a \in \mathbb{F}} \Psi_a(\omega_a) \qquad \text{with respect to } \lambda.$$

<8> **Example.** Consider the very simple case where $S = \{1, 2, 3\}$ and λ^i is counting measure on $\mathcal{X}_i = \{0, 1\}$.



The measure $\mathbb{P}$ defined by

$$\Psi_{1,2}(\omega_1, \omega_2) = \{\max(\omega_1, \omega_2) = 1\} \qquad \text{and} \qquad \Psi_{2,3}(\omega_2, \omega_3) = \{\omega_3 = 1\}$$

has density

$$p(\omega) = \tfrac{1}{3}\{\omega_1 = 1, \omega_2 = 1, \omega_3 = 1\} + \tfrac{1}{3}\{\omega_1 = 1, \omega_2 = 0, \omega_3 = 1\}$$
$$+ \tfrac{1}{3}\{\omega_1 = 0, \omega_2 = 1, \omega_3 = 1\} \qquad \text{with respect to } \lambda.$$

You should check that

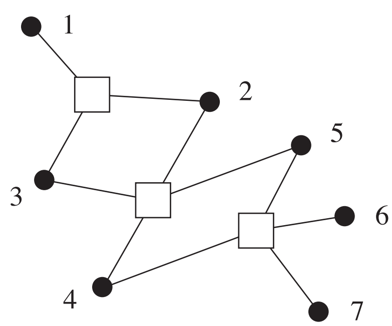
$$p_{1|2}(\omega_1 \mid \omega_2) = \tfrac{1}{2}\{\omega_1 = 0, \omega_2 = 1\} + \tfrac{1}{2}\{\omega_1 = 1, \omega_2 = 0\} + \{\omega_1 = 1, \omega_2 = 0\}.$$

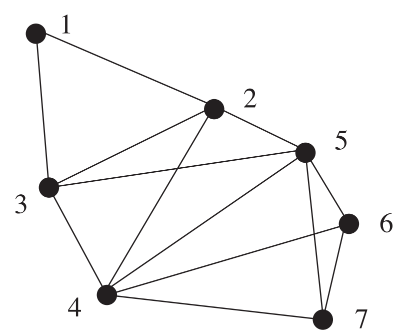
Why is there no need for a qualification involving $\lambda^1 \otimes \mathbb{P}^2$? You should also check that

$$p_{1|2,3}(\omega_1 \mid \omega_2, \omega_3) = p_{1|2}(\omega_1 \mid \omega_2) \qquad \text{a.e. } [\mathbb{P}^3]$$

□ What happens on the set $\{\omega_3 = 0\}$?

In general, the conditional independence properties for a Gibbs measure 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.

<9> **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 <7>.*

Proof. Define $\mathbb{V}_A$ as the set of all vertices from $\mathbb{V}$ that can be connected to a 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

$$p(\omega) = \phi(\omega)\psi(\omega) \qquad \text{where } \phi = \prod\nolimits_{a \in \mathbb{F}\mathbb{V}_A} \Psi_a \text{ and } \psi = \prod\nolimits_{a \in \mathbb{F}\mathbb{V}_A^c} \Psi_a.$$

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 $\mathbb{V}_A \backslash \big(A \cup C\big)$. Similarly, $\psi(\omega) = \psi(\omega_B, \omega_C, \omega_E)$ where $E \subseteq S \backslash \big(B \cup C\big)$ consists of all those sites i direcctly 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 $S\backslash\big(A \cup B \cup C\big)$. That is, we have a factorization of p as in <6>, which implies the asserted conditional independence. □

<10> **Definition.** *If i and j share a common neighbor a in the factor graph (that is, if $\{i, j\} \subseteq a$), write $i \sim j$. Write $\mathcal{N}\{i\} = \{j \in S\backslash\{i\} : i \sim j\}$. More generally, if $A \subset S$, write $\mathcal{N}(A)$ for $\{j \in S\backslash A : i \sim j$ for some $i \in A\ \}$.*

<11> **Example.** For a subset A of S let $C = \mathcal{N}(A)$ and $B = S\backslash(A \cup C)$. The set C separates A from B. It follows that ω_A is conditionally independent of ω_B given $\omega_{\mathcal{N}(A)}$. Equivalently, via <5>,

$$p_{A|\,A^c}(\omega_A \mid \omega_{A^c}) = p_{A|\mathcal{N}(A)}(\omega_A \mid \omega_{\mathcal{N}(A)}) \qquad \text{a.e. } [\lambda^A \otimes \mathbb{P}^{A^c}].$$

The special case where A consists of a single site is called the ***Markov property*** for the Gibbs distribution. □

4. 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 the Markov properties for $\mathbb{P}$.

The argument depends on the $\lambda^{i,j} \otimes \mathbb{P}^C$-almost sure factorization

$$p_{i,j|C}(\omega_i, \omega_j \mid \omega_C) = p_{i|C}(\omega_i \mid \omega_C)p_{j|C}(\omega_j \mid \omega_C) \qquad \text{where } C = S\backslash\{i, j\}.$$

Remark. I am worried about the effects on the arguments given below if there are values of ω for which the factorization can fail. I will therefore assume that the factorizations hold everywhere, in which case the representation will follow from an application of the following Lemma to $\log p(\omega)$.

I would be interested to see whether the arguments still work under weaaker assumptions.

<12> **Lemma.** *Let f 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 $f(\omega) = \sum_{a\subseteq S}\Phi_a(\omega)$ 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

$$f_a(\omega_a) = f_a(\omega) = f(\omega_a, \widetilde{\omega}_{S\backslash a})$$

and

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

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

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

Finally, to establish the representation for f, consider the coefficient of f_b in

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

When $b = a = S$, we recover $(-1)^0 f_S(\omega) = f(\omega)$. When b is a proper subset of S, the coefficient of f_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. □

Alternative proof. Suppose $\mu = \otimes_{i\in S}\mu^i$ and $\nu = \otimes_{i\in S}\nu^i$ are probability measures on $\mathcal{B}$. Write Δ^i for the signed measure $\mu^i - \nu^i$. Then

$$\begin{aligned}\mu f &= \otimes_{i\in S}\left(\nu^i + \Delta^i\right) f \\ &= \sum\nolimits_{a\subseteq S} \nu^{S\backslash a}\Delta^a f(\omega_a, \omega_{S\backslash a}) \\ &= \sum\nolimits_{a\subseteq S} \Delta^a f_a(\omega_a) \qquad \text{where } f_a := \nu^{S\backslash A} f\end{aligned}$$

If $i \in a$ and $\mu^i = \nu^i$ then Δ^a is the zero measure; the contribution from f_a to μf then disappears from the sum.

Specialize to the case where μ^i is a point mass at x_i and ν^i is a point mass at $\widetilde{\omega}_i$ to get $f(x) = \sum_{a\subseteq S}\Delta^a f_a(\omega_a)$, with $\Delta^a f_a = 0$ if $x_i = \widetilde{\omega}_i$ for any i in a. □

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. If the factorizations of p are controlled by a factor graph then a muuch more precise assertion is possible.

Say that a function f on Ω ***respects a factor graph*** if, for each pair of sites with $i \nsim j$, there is a decomposition

$$f(\omega) = g(\omega_{-i}) + h(\omega_{-j}).$$

That is, g does not depend on ω_i annd h does not depend on ω_j.

<13> **Lemma.** *Suppose $p(\omega) > 0$ for all ω. Suppose also that $\log p$ respects a given factor graph. Then p has a factorization $p(\omega) = \prod_{i\in S}\psi_i(\omega_i)\prod_{a\in\mathbb{F}}\psi_a(\omega_a)$.*

REMARK. The ψ_i factors can be absorbed into the other product if each i in S is connected to at least one a in $\mathbb{F}$.

Proof. Suppose $i \nsim j$. Then $f(\omega) := \log p(\omega) = g(\omega_{-i}) + h(\omega_{-j})$. Notice that the representation for Φ_a from Lemma <12> is linear in f. Thus $\Phi_a = \Phi_a^g + \Phi_a^h$, the sum of the terms by applying the same construction to g then to h. For example,

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

If $i \in a$, the right-hand side is unaffected if we replace ω_i by $\widetilde{\omega}_i$, because g does not depend on ω_i. However, the left-hand side is zero when $\omega_i = \widetilde{\omega}_i$. It follows that $\Phi_a^g(\omega_a) \equiv 0$ when $i \in a$. A similar argument shows that $\Phi_a^h(\omega_a) \equiv 0$ when $j \in a$. Thus $\Phi_a(\omega_a) = 0$ for each subset a of S for which $\{i, j\} \subseteq a$. □

5. Conditional distributions for finite sets of sites

The assertion of Example <11> can be proved directly by noting that the $\Psi_a(\omega_a)$ terms that involve any components of ω_A correspond to $\partial A := \{a \in \mathbb{F} : aA \neq \emptyset\}$.

Thus

$$p_{A^c}(\omega_{A^c}) = \lambda^A p(\omega_A, \omega_{A^c}) = \frac{Z_{-A}}{Z} \prod_{a\in\mathbb{F}\backslash\partial A} \psi_a(\omega_a)$$
$$\text{where } Z_{-A} = Z_{-A}(\omega_{\mathcal{N}(A)}) := \lambda^A \prod_{a\in\partial A} \Psi_a(\omega_a)$$

and

$$p_{A\,|\,A^c}(\omega_A \mid \omega_{A^c}) = \tilde{p}_{\mathcal{N}(A)}(\omega_A \mid \omega_{\mathcal{N}(A)})$$
$$:= \frac{\{Z_{-A} \neq 0\}}{Z_{-A}} \prod_{a\in\partial A} \Psi_a(\omega_a) \qquad \text{a.e. } [\lambda^A \otimes \mathbb{P}^{A^c}].$$

Notice that $\tilde{p}_{\mathcal{N}(A)}$ depends on ω only through the coordinates $\omega_{A\cup\mathcal{N}(A)}$. In fact, it defines a new Markov kernel, a family of measures $Q_{\mathcal{N}(A)}(\cdot \mid \omega_{\mathcal{N}(A)})$ on $\mathcal{B}_A$: for each $\omega_{\mathcal{N}(A)}$,

$$\frac{dQ_{\mathcal{N}(A)}}{d\lambda^A} = \tilde{p}_{\mathcal{N}(A)}(\omega_A \mid \omega_{\mathcal{N}(A)})$$

Notice that $Q_{\mathcal{N}(A)}(\cdot \mid \omega_{\mathcal{N}(A)})$ is a probability measure at each $\omega_{\mathcal{N}(A)}$ for which $Z_{-A}(\omega_{\mathcal{N}(A)}) \neq 0$.

For $D \subseteq A$, write $Q^D_{\mathcal{N}(A)}$ for the marginal distribution of ω_D under $Q_{\mathcal{N}(A)}$. The new Markov kernels fit together in an elegant way. For example, suppose $B = A \cup \mathcal{N}(A)$. Then, at least for nonnegative, $\mathcal{B}_A$-measurable f,

<14> $$\boxed{Q^{\mathcal{N}(A)}_{\mathcal{N}(B)} Q^A_{\mathcal{N}(A)} f(\omega_A) = Q^A_{\mathcal{N}(B)} f(\omega_A)}$$

Proof. Notice that both sides of the asserted equality are zero at $\omega_{\mathcal{N}(B)}$ where Z_{-B} is zero. We need only consider an $\omega_{\mathcal{N}(B)}$, which will stay fixed thoughout the argument, at which Z_{-B} is nonzero. To simplify notation, drop the subscript $\mathcal{N}(B)$ from Q and $\tilde{p}$, and write D for $\mathcal{N}(A)$. Note that

$$\frac{dQ^D}{d\lambda^D} = \frac{1}{Z_{-B}} \lambda^A \tilde{p}(\omega_A \mid \omega_{\mathcal{N}(B)}) = \frac{Z_{-A}(\omega_D)}{Z_{-B}} \prod_{a\in\partial B\backslash\partial A} \Psi_a(\omega_a),$$

because the factors Ψ_a for $a \in \partial B\backslash\partial A$ do not depend on ω_A. The asserted equality then becomes

$$Q^D Q^A_D f(\omega_A) = Q^A f(\omega_A)$$

When multiplied by Z_{-B}, the left-hand side becomes

$$\lambda^D Z_{-A}(\omega_D) \left(\prod_{a\in\partial B\backslash\partial A} \Psi_a\right) \lambda^A \frac{\{Z_{-A}(\omega_D) \neq 0\}}{Z_{-A}(\omega_D)} \left(\prod_{a\in\partial A} \Psi_a\right) f(\omega_A)$$
$$= \lambda^D \lambda^A \{Z_{-A}(\omega_D) \neq 0\} \left(\prod_{a\in\partial B} \Psi_a\right) f(\omega_A)$$

On the set $\{Z_{-A}(\omega_D) = 0\}$ we have $\prod_{a\in\partial A} \Psi_a = 0$ a.e. $[\lambda^A]$. The last iterated integral is unchanged if we omit the indicator function $\{Z_{-A} \neq 0\}$, leaving

$$\lambda^B \left(\prod_{a\in\partial B} \Psi_a\right) f(\omega_A) = Z_{-B} \times \left(\text{rhs of <14>}\right)$$

□ The asserted equality follows.

Everything from here onwards is under revision.

6. Conditional distributions for countable sets of sites

- Example to show difficulty of defining $\mathbb{P}$ via density when S is countable.

- Example to show dangers of careless specification of condit distns.

7. Notes

See Griffeath (1976) for the main idea (the so-called Möbius inversion, as presented in Lemma <12>) behind the proof of the Hammersley-Clifford result.

Georgii (1988, page 16) called a family of Markov kernels that satisfies a condition like <14> a ***specification***.

REFERENCES

Georgii, H.-O. (1988), *Gibbs Measures and Phase Trasitions*, Walter de Gruyter.

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.