Notes on the Gibbs-Markov equivalence

Tibério Caetano

August 3, 2005

This note is an attempt to describe in a detailed way how the Gibbs-Markov equivalence (also known as Hammersley-Clifford theorem) can be obtained. We start by proving the M{"o}bius inversion lemma and then proceed to the Grimmer’s proof of the theorem.

Notation: A capital letter ($A, B, \text{etc.}$) will denote a set of variables while $x_i$, where $i$ is a lowercase letter, will correspond to a specific instantiation on this set. Two or more capital letters together (e.g. $AB$) correspond to the union of the individual sets (e.g. $A \cup B$). When indexed by more than a single lowercase letter, $x$ represents a joint instantiation on the combined sets (e.g. $x_{ab}$).

Lemma 1 (M{"o}bius inversion) Let $F$ and $G$ be two arbitrary real-valued functions defined on every finite subset$^1$ of a finite set $A$. So, $F : \mathcal{P}(A) \mapsto \mathbb{R}$ and $G : \mathcal{P}(A) \mapsto \mathbb{R}$, where $\mathcal{P}(A)$ is the power set$^2$ of $A$. Then the following equivalence holds:

$$F(A) = \sum_{B : B \subseteq A} G(B) \iff G(B) = \sum_{C : C \subseteq B} (-1)^{|B|-|C|} F(C)$$  \hspace{1cm} (1)

or, equivalently:

$$F(A) \equiv \sum_{B : B \subseteq A} \left( \sum_{C : C \subseteq B} (-1)^{|B|-|C|} F(C) \right)$$  \hspace{1cm} (2)

Proof We start from (2). Swapping sums we have:

$$F(A) \equiv \sum_{C : C \subseteq B} \left[ \sum_{B : B \subseteq A} (-1)^{|B|-|C|} F(C) \right]$$  \hspace{1cm} (3)

Let’s concentrate on the bracketed expression now. Notice that $C$ is fixed in this expression, and the sum runs over $B$. Notice that the expression in brackets is a sum over all the subsets ($B$’s) of $A$ which contain a fixed subset $C$ of $B$. So, keep in mind here that $C \subseteq B \subseteq A$ always holds. As a result, $|B|$ in the bracketed expression varies from $|C|$ (when $B = C$) to $|A|$ (when $B = A$). There are $|A| - |C| + 1$ possible different sizes for the subset $B$ in this bracketed expression (ranging from $|C|$ when $B = C$ to $|A| - |C|$ when $B = A$). For each of these sizes, and denoting $n = |A| - |C|$ and $p = |B| - |C|$, there are

$${n \choose p} := \frac{n!}{p!(n-p)!}$$

$^1$We will assume that they are defined on the empty set as well, but we make $F(\emptyset) = 0$ and $G(\emptyset) = 0$.

$^2$the set of all subsets. Recall that the number of elements in the power set is $2^{|A|}$. 

1
As in the second, if the same growth is in which subsets. As a result, we may write (3) as
\[
F(A) \equiv \sum_{C:C \subseteq B} \left[ \sum_{p=0}^{n} (-1)^p \binom{n}{p} \right] F(C).
\] (4)

Now, let’s separate the outer sum in (4), which runs over all subsets \( C \) of a fixed subset \( B \), into three parts: one in which \( C = \emptyset \), another in which \( C = A \) and another including all other \( C \)'s:
\[
F(A) \equiv \sum_{C:C=\emptyset} \left[ \right] F(C) + \sum_{C:C=A} \left[ \right] F(C) + \sum_{C:\emptyset \subseteq C \subseteq A} \left[ \right] F(C).
\] (5)

Due to the assumption that \( F(\emptyset) = 0 \), the first of the three sums (which contains a single term) vanishes. The second sum also has a single term (for \( C = A \)): the bracketed expression of eqs. (4) and (5) in this case becomes
\[
\sum_{p=0}^{0} (-1)^{0} \binom{0}{0},
\] (6)
which is 1. Therefore, the second sum in (5) is simply \( F(A) \). The third sum in (5) is seen to be zero: simply write the bracketed expression in the form recognizable in the binomial theorem:
\[
\sum_{p=0}^{n} (-1)^p \binom{n}{p} = \sum_{p=0}^{n} (1)^{n-p} (-1)^p \binom{n}{p} = (1 - 1)^n = 0.
\] (7)

What (7) says is basically that, in the power set \( \mathcal{P}(A) \) of some finite set \( A \), the amounts of elements with odd and even cardinalities are the same. For example, if \( A = \{1, 2, 3\} \), we have that \( \mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\} \) and \( |\emptyset| = 0, |\{1, 2\}| = 2, |\{1, 3\}| = 2, |\{2, 3\}| = 2, |\{1\}| = 1, |\{2\}| = 1, |\{3\}| = 1 \) and \( |\{1, 2, 3\}| = 3 \), i.e. 4 even and 4 odd.

Therefore, (5) can be rewritten as
\[
F(A) \equiv 0 + F(A) + 0 \quad \Rightarrow \quad F(A) \equiv F(A),
\]
which is true.

We will in a few moments use the Möbius inversion lemma to prove the Hammersley-Clifford theorem. Before that, we fix some definitions required in order to state this theorem:

**Definition 2 (Markov property)** Let \( G = (V, E) \) be a graph and \( A, S \) and \( B \) disjoint subsets of \( V \). If \( S \) separates \( A \) from \( B \) (such that any path from \( A \) to \( B \) passes through \( S \)), then \( A \) and \( B \) are conditionally independent given \( S \): \( A \perp B|S \), or \( P(x_{ab}) = P(x_a)P(x_b)/P(x) := h(x_a)h(x_b) \).

**Definition 3 (Gibbs distribution, Gibbs potential)** A density over a graph \( G = (V, E) \) is said to be Gibbs if it is of the form
\[
P(x_v) = \prod_{K \in K_v} \psi_K(x_k)
\]
where \( K_v \) is the set of all complete subgraphs of \( G \) and \( \psi' \)'s are arbitrary\(^3\) nonnegative real-valued functions. Equivalently, by applying logarithms in both sides, we obtain the Gibbs potential

\[
\Phi(x_v) = \sum_{K \in K_v} \phi_K(x_k),
\]

(8)

where \( \phi_K = -\log \psi_K \) are the so-called “clan-potentials” (Moussouris terminology), or potentials over subsets of cliques.

Now we state and prove the two theorems for Gibbs-Markov equivalence: the Gibbs \( \Rightarrow \) Markov and the Markov \( \Rightarrow \) Gibbs parts. The first is easier:

**Theorem 4 (Gibbs \( \Rightarrow \) Markov)** Every Gibbs system is Markovian.

**Proof** Assume \( G, A, B \) and \( S \) as in definition 2. Assume also, without loss of generality, that only the maximal complete subgraphs (cliques) are included in the factorization of the Gibbs distribution in definition 3 (this can always be done by appropriately selecting the \( \psi' \)'s to include the subsets of each clique). Since \( S \) separates \( A \) and \( B \), \( A \) and \( B \) are in two different connectivity components of \( G \), which we call \( \bar{A} \) and \( \bar{B} \), such that \( \bar{A} \cup \bar{B} \cup S = V \). However, notice that the cliques of \( G \) can only be in \( \bar{A} \cup S \) or in \( \bar{B} \cup S \), since there is no edge between \( \bar{A} \) and \( \bar{B} \). If \( C \) denotes the set of cliques in \( G \), \( C_A \) denotes the set of cliques in \( \bar{A} \cup S \) and \( C \) denotes a specific clique, then the Gibbs distribution takes the form

\[
P(x_v) = \prod_{C \in C_A} \psi_C(x_C) \prod_{C \in C \setminus C_A} \psi_C(x_C)
\]

what implies that \( \bar{A} \perp \bar{B} | S \), according to definition 2. Since \( A \) and \( B \) are subsets of, respectively, \( \bar{A} \) and \( \bar{B} \), \( \bar{A} \perp \bar{B} | S \) also holds.

\[\Box\]

Now we present the converse, i.e. the Markov \( \Rightarrow \) Gibbs relation, which is less trivial.

**Theorem 5 (Markov \( \Rightarrow \) Gibbs)** Every positive Markov system is Gibbsian, with clan potentials given by

\[
\phi(x_k) = \sum_{K' : K' \subseteq K} (-1)^{|K'|-|K|} \Phi(x'_k).
\]

**Proof** We apply here the Möbius inversion equivalence:

\[
F(A) = \sum_{B : B \subseteq A} G(B) \iff G(B) = \sum_{C : C \subseteq B} (-1)^{|B|-|C|} F(C),
\]

(11)

which in our case becomes

\[
\Phi(x_v) = \sum_{B : B \subseteq V} \Gamma(x_b) \iff \Gamma(x_b) = \sum_{C : C \subseteq B} (-1)^{|B|-|C|} \Phi(x_c).
\]

(12)

In order to prove the theorem, it suffices to show that \( \Gamma(x_b) = 0 \) whenever \( B \) is not a clan. This is so because then the sum over all subsets \( B : B \subseteq V \) in

\[^3\text{They must be such that } P \text{ integrates to 1 of course.}\]
\[ \Phi(x_v) = \sum_{B : B \subseteq V} \Gamma(x_b) \]

will then become effectively a sum only over the clans, and the Gibbs potential (8) is then recovered

\[ \Phi(x_v) = \sum_{K : K \in K_v} \phi_K(x_b). \]  

(13)

To show that \( \Gamma(x_b) = 0 \) when \( B \) is not a clan, we start by noting that there are at least two nodes \( Z_1 \) and \( Z_2 \) that are not connected in \( B \). So let’s denote \( B \) by \( Z_1S Z_2 \) (i.e. the union of a “separator” \( S \) and the two unconnected nodes). Analogously, the joint assignment is \( x_b = x_{z_1z_2} \). Then, we note that any instantiation \( x_v \subseteq x_b \) must be obtained from \( x_{c'} \subseteq s \) by adjoining one, none of both of the instantiations \( z_1 \) and \( z_2 \). As a result, we can develop \( \Gamma(x_b) \)

\[ \Gamma(x_b) = \sum_{C : C \subseteq B} (-1)^{|B| - |C|} \Phi(x_c) \]  

(14)

in the following way

\[
\begin{align*}
\Gamma(x_b) &= \sum_{C : C \subseteq B} (-1)^{|B| - |C'|} \Phi(x_{c'}) + \\
&+ \sum_{C : C \subseteq S} (-1)^{|B| - |Z_1C'|} \Phi(x_{z_1c'}) + \\
&+ \sum_{C : C \subseteq S} (-1)^{|B| - |Z_1Z_2|} \Phi(x_{z_1z_2}) + \\
&+ \sum_{C : C \subseteq S} (-1)^{|B| - |Z_1C'Z_2|} \Phi(x_{z_1c'z_2})
\end{align*}
\]  

(15)

The exponents of the \((-1)\)’s can be made equal by correcting the signs of the correspondent \( \Phi \)’s (since two exponents have odd and two have even parity):

\[
\Gamma(x_b) = \sum_{C : C \subseteq S} (-1)^{|B| - |C'|} [\Phi(x_{c'}) + \Phi(x_{z_1c'z_2}) - \Phi(x_{z_1c'}) - \Phi(x_{c'z_2})]
\]  

(16)

Now we will show that the term in brackets is zero (what makes \( \Gamma(x_b) = 0 \) and thus proves the theorem). By exponentiating it, we obtain

\[
\exp[.] = \frac{P(x_{c'})}{P(x_{z_1c'})} \frac{P(x_{z_1c'z_2})}{P(x_{z_1c'})} \frac{P(x_{z_1z_2})}{P(x_{z_1c'})} \frac{P(x_{z_1z_2})}{P(x_{z_1z_2})} = 1
\]  

(17)

(18)

So, \([.] = 0 \). Note that the Markov property was used: \( P(z_1|c'z_2) = P(z_1|c') \). Also, note that the positivity of \( P \) is essential for the Markov ⇒ Gibbs part (and not necessary for the converse part).