Proof of Hammersley-Clifford Theorem

Samson Cheung

February 3, 2008

We start with two definitions.

Definition 1 *An undirected graphical model G is called a Markov Random Field (MRF) if two nodes are conditionally independent whenever they are separated by evidence nodes. In other words, for any node X_i in the graph, the following conditional property holds:*

$$P(X_i|X_{G\backslash i}) = P(X_i|X_{N_i}) \tag{1}$$

where $X_{G\backslash i}$ denotes all the nodes except X_i, and X_{N_i} denotes the neighborhood of X_i – all the nodes that are connected to X_i.

Definition 2 *A probability distribution $P(X)$ on an undirected graphical model G is called a Gibbs distribution if it can be factorized into positive functions defined on cliques that cover all the nodes and edges of G. That is,*

$$P(X) = \frac{1}{Z}\prod_{c\in C_G}\phi_c(X_c) \tag{2}$$

where C_G is a set of all (maximal) cliques in G and $Z = \sum_x \prod_{c\in C_G}\phi_c(X_c)$ is the normalization constant.

The Hammersley Clifford Theorem tells us that the above two definitions are equivalent. In this short note, I will go over the proof of this theorem. Let's start by showing that a Gibbs distribution must satisfy the conditional independence induced by node separation.

1 Backward Direction

Let $D_i = N_i \cup \{X_i\}$ be the set of notes containing the neighborhood of X_i and X_i itself. Starting from the right-hand side of Equation (1), we have

$$\begin{aligned} P(X_i|X_{N_i}) &= \frac{P(X_i, X_{N_i})}{P(X_{N_i})} && (3) \\ &= \frac{\sum_{G\setminus D_i} \prod_{c\in C_G} \phi_c(X_c)}{\sum_{x_i} \sum_{G\setminus D_i} \prod_{c\in C_G} \phi_c(X_c)} && (4) \end{aligned}$$

Now we partition the set of maximal cliques C_G into two parts based on whether the clique contains X_i as a member: $C_i = \{c \in C_G : X_i \in c\}$ and $R_i = \{c \in C_G : X_i \notin c\}$. We can now break the products in (4) into two products on C_i and R_i.

$$\begin{aligned} P(X_i|X_{N_i}) &= \frac{\sum_{G\setminus D_i} \prod_{c\in C_i} \phi_c(X_c) \prod_{c\in R_i} \phi_c(X_c)}{\sum_{x_i} \sum_{G\setminus D_i} \prod_{c\in C_i} \phi_c(X_c) \prod_{c\in R_i} \phi_c(X_c)} && (5) \\ &= \frac{\prod_{c\in C_i} \phi_c(X_c) \sum_{G\setminus D_i} \prod_{c\in R_i} \phi_c(X_c)}{\sum_{x_i} \prod_{c\in C_i} \phi_c(X_c) \sum_{G\setminus D_i} \prod_{c\in R_i} \phi_c(X_c)} && (6) \end{aligned}$$

The summation on $G\setminus D_i$ can be moved pass the product over C_i because all the nodes in the cliques in C_i must be from D_i as they are by definition connected to X_i. Also notice that the factor $\sum_{G\setminus D_i} \prod_{c\in R_i} \phi_c(X_c)$ does not involve X_i at all and thus can be factored out from the summation in the denominator and canceled with the same factor in the numerator. We now have

$$\begin{aligned} P(X_i|X_{N_i}) &= \frac{\prod_{c\in C_i} \phi_c(X_c)}{\sum_{x_i} \prod_{c\in C_i} \phi_c(X_c)} && (7) \\ &= \frac{\prod_{c\in C_i} \phi_c(X_c)}{\sum_{x_i} \prod_{c\in C_i} \phi_c(X_c)} \cdot \frac{\prod_{c\in R_i} \phi_c(X_c)}{\prod_{c\in R_i} \phi_c(X_c)} && (8) \\ &= \frac{\prod_{c\in C_G} \phi_c(X_c)}{\sum_{x_i} \prod_{c\in C_G} \phi_c(X_c)} && (9) \\ &= \frac{P(X)}{P(X_{G\setminus\{i\}})} && (10) \\ &= P(X_i|X_{G\setminus\{i\}}) && (11) \end{aligned}$$

After eliminating the summation over $G\setminus D_i$, we multiply the same factor to both the numerator and the denominator in (8) to re-introduce the potential functions from the remainder of the graph. They allow us to get back $P(X)$ and $P(X_{G\backslash\{i\}})$, resulting in the desired left-hand side. QED.

2 Forward Direction

The proof of the other direction is constructive – we are going to show how to express $\phi_c(X_c)$ given the joint probability and the Markov property. For any subset $s \subset G$, we define a candidate potential function as follows:

$$f_s(X_s = x_s) = \prod_{z \subset s} P(X_z = x_z, X_{G\backslash z} = 0)^{-1^{|s|-|z|}} \tag{12}$$

The product on the right-hand side of (12) is over all subsets of s. Given a subset z of s, $P(X_z = x_z, X_{G\backslash z} = 0)$ means that the nodes over z matches the corresponding portion of X_s while the remainder of the graph is set to a default configuration denoted as "0". The power is 1 if the difference in size between s and z is even and -1 otherwise. Believe it or not, this is actually what we are looking for. f_s is obviously positive. To show the rest, we need to demonstrate:

1. $\prod_{s \subset G} f_s(X_s) = P(X)$
2. $f_s(X_s) = 1$ if s is not a clique.

To show 1., let us first state a really stupid identity:

$$0 = (1-1)^K = C_0^K - C_1^K + C_2^K + \ldots + (-1)^K C_K^K \tag{13}$$

where C_N^K denotes the number of combinations when choosing N elements from a K-element set. We want to show that all factors in $\prod_{s \subset G} f_s(X_s)$ cancel themselves out except for $P(X)$. Consider any subset $z \subset G$ and the corresponding factor $\triangle = P(X_z, X_{G\backslash z} = 0)$: it occurs in $f_z(X_z)$ as $\triangle^{-1^0} = \triangle$. It also occurs in the functions over subsets containing z and one additional element – there are $C_1^{|G|-|z|}$ such functions, and the additional element creates the inverse factor $\triangle^{-1^1} = \triangle^{-1}$. Moving onto the functions over subsets containing z and two additional elements – there are $C_1^{|G|-|z|}$ such functions, and the two additional element creates the factor itself $\triangle^{-1^2} = \triangle$. Continuing this process and using the equality (13), it is easy to see that everything cancels each other out except the boundary case when $z = G$. In that case, you have a single factor which is $P(x)$.

To show fact 2, we need to use the MRF property. Since s is not a clique, we must be able to find two nodes a and b that are not connected to each other. So we rewrite $f_s(X_s)$ as follows:

$$
\begin{aligned}
& f_s(X_s = x_s) \\
&= \prod_{z \subset s} P(X_z = x_z, X_{G\setminus z} = 0)^{-1^{|s|-|z|}} && (14) \\
&= \prod_{w \subset s\setminus\{a,b\}} \left[\frac{P(X_w, X_{G\setminus w} = 0) P(X_{w\cup\{a,b\}}, X_{G\setminus w\cup\{a,b\}} = 0)}{P(X_{w\cup\{a\}}, X_{G\setminus w\cup\{a\}} = 0) P(X_{w\cup\{b\}}, X_{G\setminus w\cup\{b\}} = 0)} \right]^{-1^*} && (15)
\end{aligned}
$$

What (15) does is to classify each subset $z \in s$ into four types: $z = w$, $z = w \cup \{a\}$, $z = w \cup \{b\}$ and $z = w \cup \{a,b\}$, and to explicitly write out those factors. Notice that the relative positions of the four factors in (15) are correct and we are going to show that they cancel each other. Thus, the precise power is not important and is represented as -1^*. Using Bayes rules, we have

$$
\begin{aligned}
& \frac{P(X_w, X_{G\setminus w} = 0)}{P(X_{w\cup\{a\}}, X_{G\setminus w\cup\{a\}} = 0)} \\
&= \frac{P(X_a = 0 | X_b = 0, X_w, X_{G\setminus w\cup\{a,b\}} = 0) P(X_b = 0, X_w, X_{G\setminus w\cup\{a,b\}} = 0)}{P(X_a | X_b = 0, X_w, X_{G\setminus w\cup\{a,b\}} = 0) P(X_b = 0, X_w, X_{G\setminus w\cup\{a,b\}} = 0)} && (16) \\
&= \frac{P(X_a = 0 | X_b, X_w, X_{G\setminus w\cup\{a,b\}} = 0) P(X_b, X_w, X_{G\setminus w\cup\{a,b\}} = 0)}{P(X_a | X_b, X_w, X_{G\setminus w\cup\{a,b\}} = 0) P(X_b, X_w, X_{G\setminus w\cup\{a,b\}} = 0)} && (17) \\
&= \frac{P(X_{w\cup\{b\}}, X_{G\setminus w\cup\{b\}} = 0)}{P(X_{w\cup\{a,b\}}, X_{G\setminus w\cup\{a,b\}} = 0)} && (18)
\end{aligned}
$$

First, we notice that the priors (the second factor) in both the denominator and numerator of (16) are the same. Instead of canceling them, the default $X_b = 0$ is changed to the actual value X_b in (17). We can change $X_b = 0$ to X_b in the conditional probabilities because X_a and X_b are conditional independent given the rest of the graph as they are not connected. As such we can replace the condition X_b with any value we like. The result follows.