University of Washington
Dept. of Electrical Engineering

Handout 2: Hammersley Clifford

Prof: J. Bilmes <bilmes@ee.washington.edu>

April 15, 2006

This handout includes a complete description and proof of the Möbius Inversion Lemma, and also the Hammersley-Clifford theorem which states that for positive distributions we have that (F) $\equiv$ (G). This consequential theorem states that for positive distributions, we are at liberty to reason about Markov random fields using any of the three main Markov properties on a graph, and also the factorization property.

Lemma 2.1. *(Möbius Inversion Lemma) Let ψ and ϕ be functions defined on the set of all subsets of a finite set V, taking values in an Abelian group (i.e., a group (closure, associativity, identity, and inverse) for which the elements also commute, the reals being just one example). The following two equations imply each other.*

$$\forall a \subseteq V : \psi(a) = \sum_{b:b\subseteq a} \phi(b) \tag{2.1}$$

$$\forall a \subseteq V : \phi(a) = \sum_{b:b\subseteq a} (-1)^{|a\setminus b|}\psi(b) \tag{2.2}$$

Proof. We first plug equation 2.2 into equation 2.1 and show that equation 2.1 follows.

$$\begin{aligned}
\sum_{b:b\subseteq a} \phi(b) &= \sum_{b:b\subseteq a} \sum_{c:c\subseteq b} (-1)^{|b\setminus c|}\psi(c) && (2.3)\\
&= \sum_{c:c\subseteq a} \sum_{b:c\subseteq b \& b\subseteq a} \psi(c)(-1)^{|b\setminus c|} && \text{we rearrange the order of summation} \quad (2.4)\\
&= \sum_{c:c\subseteq a} \psi(c) \sum_{b:c\subseteq b \& b\subseteq a} (-1)^{|b\setminus c|} && (2.5)\\
&= \sum_{c:c\subseteq a} \psi(c) \sum_{h:h\subseteq a\setminus c} (-1)^{|h|} && (2.6)
\end{aligned}$$

The last step follows because the set of subsets $b : c \subseteq b\&b \subseteq a$ is like the set of subsets of $a \setminus c$ since, by requiring $c \subseteq b$, we are essentially reducing the number of possible subsets by a factor of $2^{|c|}$. That is, in each case there are a total of $2^{|a|-|c|} = 2^{|a\setminus c|}$ possible subsets, and the cardinalities of the set of subsets are the same. Also, for each of those subsets we are raising (-1) to the number of elements in that subset, but not including c (this is the exponent $|b \setminus c|$). An easy way to see this is to think of a bit vector to select subsets. In any event, the result then becomes the last equation.

Also, note that

$$\sum_{h:h\subseteq a\setminus c} (-1)^{|h|}$$

is zero for all $a \setminus c$ except for the case when $a \setminus c = \emptyset$ (i.e., for any non empty set, there are the same number of even and odd subsets). Also, $a \setminus c = \emptyset$ only when $a = c$, leading to

$$\sum_{c:c\subseteq a} \psi(c) \sum_{h:h\subseteq a\setminus c} (-1)^{|h|} = \psi(a)$$

thus proving the theorem. The other direction is very similar. $\square$

Theorem 2.2. *(Hammersley and Clifford) A probability distribution P with positive and continuous density f satisfies the pairwise Markov property with respect to an undirected graph $\mathcal{G}$ if and only if it factorizes according to $\mathcal{G}$. I.e., (F) ≡ (G)*

Proof. Recall our notation. X is the collection of random variables and X_A is the subset of random variables for some set $A \subseteq V$ where V is the set of nodes in the graph. An assignment to the set of random variables is denoted by $X = x$ or just x, and an assignment to a subset of the random variables is denoted as $X_A = x_A$ or just x_A.

Also, recall the definition of factorization.

$$f(x) = \prod_{a:a \text{ is complete}} \psi_a(x) \tag{2.7}$$

It suffices to prove that (P) ⇒ (F) and that (F) ⇒ (G) since we have already shown that (G) ⇒ (L) ⇒ (P).

(P) ⇒ (F)

A quick preliminary and definition: since the density $f(x)$ is positive, we may take logarithms of both sizes of Equation 2.7 and get:

$$\log f(x) = \sum_{a:a \subseteq V} \phi_a(x)$$

where $\phi_a(x) = \log \psi_a(x)$ and where we must have that $\phi_a(x) \equiv 0$ unless a is a complete subset of V.

Now for the proof. Assume that P is pairwise Markov and choose a fixed but arbitrary assignment x^* to the set of random variables X. For all $a \subseteq V$, define the following function

$$H_a(x) = \log f(x_a, x^*_{a^c})$$

where $a^c \triangleq V \setminus a$, and where we use the notation $(x_a, x^*_{a^c})$ to indicate a particular assignment of values to all the random variables X. Note that $H_V(x) = \log f(x)$. This assignment is defined as follows: for the random variables X_a, the assigned values comes from the values contained in x, the argument of $H_a(x)$. For the random variables $X_{a^c} = X_{V \setminus a}$, the assigned values come from the arbitrary assignment we choose earlier, x^*. Another way of saying this is that $(x_a, x^*_{a^c})$ is an assignment to all the random variables, which we denote by $\hat{x}$, such that $\hat{x}_\gamma = x_\gamma$ for $\gamma \in a$ and where $\hat{x}_\gamma = x^*_\gamma$ for $\gamma \notin a$.

Here's an example that will make things clear if they are not already. Suppose $V = \{1, 2, 3, 4\}$ which means we have the four random variables $X = \{X_1, X_2, X_3, X_4\}$. We might choose some x^* as one assignment. I.e., specifying x^* could mean that $X_1 = 3, X_2 = 5, X_3 = 10, X_4 = 15$. Suppose that $a = \{1, 3\}$. Then the function $H_a(x)$ becomes $H_a(x) = H_a(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) = \log f(X_1 = x_1, X_2 = 5, X_3 = x_3, X_4 = 15)$. Note that $H_a(x)$ doesn't depend on x_2 and x_4 since those are already assigned via x^*.

More generally, since x^* is fixed, $H_a(x)$ depends on x only via the values x_a and not via x_{a^c}.

We further define for all $a \subseteq V$

$$\phi_a(x) = \sum_{b:b \subseteq a} (-1)^{|a \setminus b|} H_b(x)$$

Since $\phi_a(x)$ depends on x only via $H_b(x)$ and since b is chosen to be a subset of a, $\phi_a(x)$ also depends on x only through x_a.

Next, apply the Möbius inversion lemma to obtain that

$$H_V(x) = \sum_{a:a \subseteq V} \phi_a(x)$$

(note that the lemma says that this is also true for all subsets of V but we need only this case).

You might have noticed one important thing, namely that

$$\log f(x) = H_V(x)$$

because of the definition of $H_a(x)$. We therefore have that

$$\log f(x) = H_V(x) = \sum_{a:a\subseteq V} \phi_a(x)$$

But if we look back at our preliminary definition, we see that we are "almost" done. To show that the distribution factorizes according to the complete subsets of V (our goal, if you recall), we just have to ensure that this new $\phi_a(x) \equiv 0$ whenever a is not a complete subset of V

So, assume that $\alpha, \beta \in a$ and that $\alpha \not\sim \beta$ (i.e., they are not joined, so a is not complete). Also, define the set $c = a \setminus \{\alpha, \beta\}$. Then we can expand the definition of $\phi_a(x)$ as follows:

$$\begin{aligned}
\phi_a(x) &= \sum_{b:b\subseteq a} (-1)^{|a\setminus b|} H_b(x) && (2.8)\\
&= \sum_{b:b\subseteq (c\cup\{\alpha,\beta\})} (-1)^{|a\setminus b|} H_b(x) && (2.9)\\
&= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_b(x) + \sum_{b:b\subseteq c} (-1)^{|a\setminus (b\cup\{\alpha\})|} H_{b\cup\{\alpha\}}(x) + \sum_{b:b\subseteq c} (-1)^{|a\setminus (b\cup\{\beta\})|} H_{b\cup\{\beta\}}(x) + \sum_{b:b\subseteq c} (-1)^{|a\setminus (b\cup\{\alpha,\beta\})|} H_{b\cup\{\alpha,\beta\}}(x) && (2.10)\\
&= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_b(x) - \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_{b\cup\{\alpha\}}(x) - \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_{b\cup\{\beta\}}(x) + \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_{b\cup\{\alpha,\beta\}}(x) && (2.11)\\
&= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} \left(H_b(x) - H_{b\cup\{\alpha\}}(x) - H_{b\cup\{\beta\}}(x) + H_{b\cup\{\alpha,\beta\}}(x)\right) && (2.12)\\
&= \sum_{b:b\subseteq c} (-1)^{|c\setminus b|} \left(H_b(x) - H_{b\cup\{\alpha\}}(x) - H_{b\cup\{\beta\}}(x) + H_{b\cup\{\alpha,\beta\}}(x)\right) && (2.13)
\end{aligned}$$

So our job is done if we show that the quantity

$$\left(H_b(x) - H_{b\cup\{\alpha\}}(x) - H_{b\cup\{\beta\}}(x) + H_{b\cup\{\alpha,\beta\}}(x)\right)$$

is zero. We do this as follows. First, for notational simplicity, define $d = V \setminus \{\alpha, \beta\}$. Then the following equations follow where we use positivity and continuity of the distributions:

$$\begin{aligned}
H_{b\cup\{\alpha,\beta\}}(x) - H_{b\cup\{\alpha\}}(x) &= \log \frac{f(x_b, x_\alpha, x_\beta, x^*_{d\setminus b})}{f(x_b, x_\alpha, x^*_\beta, x^*_{d\setminus b})} && (2.14)\\
&= \log \frac{f(x_\alpha|x_\beta, x_b, x^*_{d\setminus b}) f(x_\beta, x_b, x^*_{d\setminus b})}{f(x_\alpha|x^*_\beta, x_b, x^*_{d\setminus b}) f(x^*_\beta, x_b, x^*_{d\setminus b})} && (2.15)\\
&= \log \frac{f(x_\alpha|x_b, x^*_{d\setminus b}) f(x_\beta, x_b, x^*_{d\setminus b})}{f(x_\alpha|x_b, x^*_{d\setminus b}) f(x^*_\beta, x_b, x^*_{d\setminus b})} && \text{by the pairwise Markov property} \quad (2.16)\\
&= \log \frac{f(x^*_\alpha|x_b, x^*_{d\setminus b}) f(x_\beta, x_b, x^*_{d\setminus b})}{f(x^*_\alpha|x_b, x^*_{d\setminus b}) f(x^*_\beta, x_b, x^*_{d\setminus b})} && \text{Since the first ratios is just unity} \quad (2.17)\\
&= \log \frac{f(x_b, x_\beta, x^*_\alpha, x^*_{d\setminus b})}{f(x_b, x^*_\alpha, x^*_\beta, x^*_{d\setminus b})} && \text{by pairwise Markov property and chain rule} \quad (2.18)\\
&= H_{b\cup\{\beta\}}(x) - H_b(x) && (2.19)
\end{aligned}$$

therefore everything is zero.

(F) ⇒ (G)

Let (A, B, S) be any triple of disjoint subsets such that S separates A from B. Also, let $\tilde{A}$ be the connectivity components of $\mathcal{G}_{V \setminus S}$ containing A. I.e.,

$$\tilde{A} = \bigcup_{\alpha \in A} [\alpha]_{V \setminus S}$$

and define $\tilde{B}$ as $\tilde{B} = V \setminus (\tilde{A} \cup S)$, so that $\tilde{B}$ is the remainder of the graph and it contains B.

Since A is separated from B by S, A and B are in different connectivity components of $\mathcal{G}_{V \setminus S}$. This implies that any clique of $\mathcal{G}$ is either a subset of $\tilde{A} \cup S$ or of $\tilde{B} \cup S$ but not both (i.e., if we had a clique in both, some elements would bypass S connecting $\tilde{A}$ to $\tilde{B}$, but this can't be since S separate A from B).

Let $\mathcal{C}_A$ be the cliques in $\tilde{A} \cup S$. Then we have that $\mathcal{C}$ (the cliques in the graph) are such that

$$\mathcal{C} = \mathcal{C}_A \cup \mathcal{C}_B$$

with $\mathcal{C}_B = \mathcal{C} \setminus \mathcal{C}_A$. But because of factorization, this means we can represent the joint distribution as:

$$f(x) = \prod_{c \in \mathcal{C}} \psi_c(x) = \prod_{c \in \mathcal{C}_A} \psi_c(x) \prod_{c \in \mathcal{C} \setminus \mathcal{C}_A} \psi_c(x) = h(x_{\tilde{A} \cup S}) k(x_{\tilde{B} \cup S})$$

which implies that $\tilde{A} \perp\!\!\!\perp \tilde{B} | S$ but this then implies that $A \perp\!\!\!\perp B | S$ which is (G), the global Markov property.

Note that this direction of the proof does not require positivity.

□