APPLICATIONES MATHEMATICAE
Online First version

VIVEK S. BORKAR

A SIMPLE PROOF OF THE HAMMERSLEY–CLIFFORD THEOREM

Dedicated to the memory of Prof. Sanjoy K. Mitter

Abstract. Using only elementary facts about conditional independence, a simple proof of the Hammersley–Clifford theorem is given.

1. Introduction. The Hammersley–Clifford theorem about the equivalence of Markov random fields and Gibbs distributions is a cornerstone of the theory of random fields. While the original article of Hammersley and Clifford [5] remained unpublished, several proofs appeared soon afterwards [1, 4, 6, 7, 8, 9] (see also [2]). See [3, 10] for some precursors.

The objective of the present note is to provide a short proof based on a characterization of preservation of conditional independence under an absolutely continuous change of measure.

2. Conditional independence under a change of measure. We begin with some facts about conditional expectations under an absolutely continuous change of probability measures.

Suppose X, Z, Y are random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, taking values in Polish spaces S_1, S_2, S_3 resp. Suppose X, Z are conditionally independent given Y on $(\Omega, \mathcal{F}, P)$. Consider a new probability measure Q on $(\Omega, \mathcal{F})$ defined by

$$Q(A) = \int_A \Lambda \, dP \quad \forall \text{ Borel sets } A \in \mathcal{F},$$

2020 *Mathematics Subject Classification*: Primary 60G60.
Key words and phrases: Gibbs distribution, Markov random field, conditional independence, change of measure, Hammersley–Clifford theorem.
Received 10 June 2025; revised 24 January 2026.
Published online 23 April 2026.

DOI: 10.4064/am2577-1-2026

for a prescribed random variable $\Lambda \geq 0$ on $(\Omega, \mathcal{F}, P)$ satisfying $E[\Lambda] = 1$. Let $E_0[\cdots]$ denote the expectation (or conditional expectation) under Q. Let $\mathcal{F}_0 \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$.

LEMMA 2.1. *For any integrable random variable ξ on $(\Omega, \mathcal{F}, P)$,*

$$E_0[\xi \,|\, \mathcal{F}_0] = \frac{E[\xi\Lambda \,|\, \mathcal{F}_0]}{E[\Lambda \,|\, \mathcal{F}_0]}.$$

Proof. For any $\mathcal{F}_0$-measurable bounded random variable W, we have

$$\begin{aligned} E_0\big[W E_0[\xi \,|\, \mathcal{F}_0]\big] &= E_0[W\xi] = E[W\xi\Lambda] \\ &= E\big[W E[\xi\Lambda \,|\, \mathcal{F}_0]\big] \\ &= E\left[W\left(\frac{E[\xi\Lambda \,|\, \mathcal{F}_0]}{E[\Lambda \,|\, \mathcal{F}_0]}\right)E[\Lambda \,|\, \mathcal{F}_0]\right] \\ &= E\left[W\left(\frac{E[\xi\Lambda \,|\, \mathcal{F}_0]}{E[\Lambda \,|\, \mathcal{F}_0]}\right)\Lambda\right] \\ &= E_0\left[W\left(\frac{E[\xi\Lambda \,|\, \mathcal{F}_0]}{E[\Lambda \,|\, \mathcal{F}_0]}\right)\right]. \end{aligned}$$

Since W was an arbitrary $\mathcal{F}_0$-measurable random variable, the claim follows from the a.s. uniqueness of conditional expectations. ∎

In what follows, $C_b(\cdots) :=$ the space of bounded continuous functions on the Polish space '$\cdots$' with the supremum norm, and $\mathcal{P}(\cdots) :=$ the space of probability measures on the Polish space '$\cdots$' with the Prokhorov topology. Define $\Lambda_{XYZ} := E_0[\Lambda \,|\, X, Y, Z]$, $\Lambda_{XY} := E_0[\Lambda \,|\, X, Y]$, $\Lambda_{YZ} := E_0[\Lambda \,|\, Y, Z]$, $\Lambda_Y := E[\Lambda \,|\, Y]$. Then $\Lambda_\phi = E[\Lambda_\phi \,|\, \phi, Y]$ for $\phi = XYZ, XY, YZ$, and hence $\Lambda_Y = E[E[\Lambda_\phi \,|\, \phi, Y] \,|\, Y] = E[\Lambda_Y \,|\, Y]$ for $\phi = XYZ, XY, YZ$.

THEOREM 2.2. *X, Z are conditionally independent given Y as random variables on $(\Omega, \mathcal{F}, Q)$ if and only if Λ_{XYZ} factorizes as $\Lambda_{X,Y,Z} = \Lambda_1(X, Y)\Lambda_2(Y, Z)$ for suitable $\Lambda_1(X, Y), \Lambda_2(Y, Z)$, where the choice of the latter is not unique.*

Proof. ($\Leftarrow$) Let $f \in C_b(S_1)$, $g \in C_b(S_2)$. If $\Lambda_0 = \Lambda_1\Lambda_2$, then

$$\begin{aligned} E_0[f(X)g(Z) \,|\, Y] &= \frac{E[f(X)g(Z)\Lambda_1(X, Y)\Lambda_2(Y, Z) \,|\, Y]}{E[\Lambda_1(X, Y)\Lambda_2(Y, Z) \,|\, Y]} \\ &= \frac{E[f(X)\Lambda_1(X, Y) \,|\, Y]E[g(Z)\Lambda_2(Y, Z) \,|\, Y]}{E[\Lambda_1(X, Y) \,|\, Y]E[\Lambda_2(Y, Z) \,|\, Y]} \\ &= \left(\frac{E[f(X)\Lambda_1(X, Y) \,|\, Y]}{E[\Lambda_1(X, Y) \,|\, Y]}\right) \times \left(\frac{E[g(Z)\Lambda_2(Y, Z) \,|\, Y]}{E[\Lambda_2(Y, Z) \,|\, Y]}\right) \\ &= E_0[f(X) \,|\, Y]E_0[g(Z) \,|\, Y], \end{aligned}$$

where the first and the fourth equalities follow from Lemma 2.1 and the second equality follows from the conditional independence of X, Z given Y

under P. Since f, g were arbitrary in $C_b(S_1), C_b(S_3)$ resp., it follows that X, Z are conditionally independent given Y under Q.

($\Rightarrow$) Suppose that $\Lambda' := \Lambda_1\Lambda_2$ for $\Lambda_1 = \Lambda_{XY}, \Lambda_2 = \Lambda_{YZ}$ defined above. A priori, it is not obvious that $E[\Lambda'] = 1$. If not, we normalize one of them by Λ_Y, e.g., by replacing Λ_1 with Λ_1/Λ_Y. With abuse of notation, we retain the notation Λ_1, Λ_2 regardless. Define a probability measure Q' on $(\Omega, \mathcal{F})$ by

$$Q'(B) = \int_B \Lambda' dP \qquad \forall \text{ Borel sets } B \in \mathcal{F}.$$

Denote by $E'[\cdots]$ the expectation (or conditional expectation) under Q'. Then we have

$$\begin{aligned} E_0[f(X)g(Z)\,|\,Y] &= E_0[f(X)\,|\,Y]E_0[g(Z)\,|\,Y] \\ &= \left(\frac{E[f(X)\Lambda_1\,|\,Y]}{E[\Lambda_1\,|\,Y]}\right)\left(\frac{E[g(Z)\Lambda_2\,|\,Y]}{E[\Lambda_2\,|\,Y]}\right) \\ &= \left(\frac{E[f(X)\Lambda_1\,|\,Y]E[g(Z)\Lambda_2\,|\,Y]}{E[\Lambda_1\,|\,Y]E[\Lambda_2\,|\,Y]}\right) \\ &= \left(\frac{E[f(X)g(Z)\Lambda_1\Lambda_2\,|\,Y]}{E[\Lambda_1\Lambda_2\,|\,Y]}\right) \\ &= \left(\frac{E[f(X)g(Z)\Lambda'\,|\,Y]}{E[\Lambda'\,|\,Y]}\right) \\ &= E'[f(X)g(Z)\,|\,Y]. \end{aligned} \tag{2.1}$$

Here the first equality follows from conditional independence of X, Z given Y under Q, the second from Lemma 2.1, the fourth from the conditional independence of X, Z given Y under Q' (which follows from the first part of this proof), the fifth from the definition of Λ', and the last from Lemma 2.1. It follows that Λ_{XYZ}, Λ' can differ at most by a factor depending on Y alone, which can be absorbed in Λ_1 or Λ_2. The claim follows. ■

3. The Hammersley–Clifford theorem. Consider an undirected graph $\mathcal{G}$ with node set $\mathcal{V}$ and edge set $\mathcal{E}$. Let

$$\mathcal{V}_i := \{j \in \mathcal{V} : \text{edge } (i,j) \in \mathcal{E}\}.$$

Since there is an edge (i,j) if and only if $j \in \mathcal{V}_i$, we have $j \in \mathcal{V}_i$ if and only if $i \in \mathcal{V}_j$. Assign to each node i of $\mathcal{G}$ an S-valued random variable X_i such that the law of $X = [X_1, \ldots, X_N]$ is $\gamma \in \mathcal{P}(S^N)$. Suppose that for each i, the conditional law of X_i given X_j, $j \neq i$, is the same as its conditional law given its neighbours X_j, $j \in \mathcal{V}_i$. This generalizes the Markov property from random processes indexed by discrete time to more general collections of random variables indexed by the nodes of a graph.

REMARK 3.1. In principle this identification is always possible for *some* graph by identifying the edge set with the prescribed symmetric relation $j \in \mathcal{V}_i$ and vice versa. Also, note the obvious analogy with the equivalent definition of Markov property for a Markov chain $\{X_n\}$:

$$P(X_n = i \,|\, X_k,\, k < n \text{ or } > n) = P(X_n = i \,|\, X_{n-1}, X_{n+1}) \quad \forall i, n.$$

Collections of random variables indexed by the nodes of a graph satisfying the aforementioned Markov property are called *Markov random fields.* In fact, more generally, a Markov random field is a collection of random variables indexed by points in any abstract topological space satisfying: collections of these random variables indexed by points in the interior of a set in this space and those indexed by points in the complement of its closure are conditionally independent given the values of those that are indexed by points on the boundary of the set. In most applications, one does not have to venture into such a level of generality, but random fields indexed by points of a euclidean space are not uncommon – think of the temperature tomorrow as a function of time and location.

Returning to graph-indexed random fields, let $\mathcal{G}$ be a connected undirected graph as above with $|\mathcal{V}| = N$. Without loss of generality, we write $\mathcal{V} = \{1, \ldots, N\}$. We shall consider distributions of the type

$$P(X_i = s_i,\, i \in \mathcal{V}) = \left(\frac{e^{H(s_1,\ldots,s_N)}}{\int_{S^N} e^{H(\bar{s}_1,\ldots,\bar{s}_N)} \prod_{k=1}^N \mu(d\bar{s}_i)} \right) \prod_{k=1}^N \mu(ds_i), \tag{3.1}$$

where $\mu \in \mathcal{P}(S)$ is a prescribed base measure on S (e.g., the uniform distribution on a finite set S or the Lebesgue measure for $S = \mathcal{R}^d$). Distributions of this form are called *Gibbs distributions* after the famous statistical physicist J. Willard Gibbs. This distribution is quite general, but the most interesting class is when the function $H(\cdot) : S^N \to \mathbb{R}$ can be written as a sum

$$H(s) = \sum_{k=1}^M H_k(s), \tag{3.2}$$

where each $H_k(\cdot)$ depends only on the variables s_i corresponding to *some* clique of the graph. At one level, (3.1) too is quite general, because given any such distribution on a finite set $\mathcal{V}$, we can construct a graph by adding an edge (i, j) whenever i, j occur together in the same H_k. However, with $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ already prescribed, the celebrated Hammersley–Clifford theorem says the following:

THEOREM 3.2 (Hammersley–Clifford theorem). *There is a one-to-one correspondence between Gibbs distributions on $\mathcal{G}$ and Markov random fields.*

In other words, given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, $\mathbf{X} := \{X_1, \ldots, X_N\}$ indexed by $\mathcal{V}$ constitutes a Markov random field if and only if its law is a Gibbs distribution on $\mathcal{V}$ where H is a sum of functions of subsets of $\mathbf{X}$ corresponding to the cliques of $\mathcal{G}$. In hindsight, the result is quite intuitive. We give here a simple proof based on the results on conditional independence from Section 2.

Proof of Theorem 3.2. We first prove that given a Markov random field, its distribution is a Gibbs distribution. Consider the probability space $(\Omega, \mathcal{F}, P)$ wherein $\Omega := \prod_{i\in\mathcal{V}} \hat{S}_i$ where the $\hat{S}_i$'s are copies of S, $\mathcal{F} :=$ its Borel σ-field, and $P :=$ the prescribed Gibbs distribution. Define the X_i, $i \in \mathcal{V}$, 'canonically' via the evaluation map $X_i(\omega_1, \ldots, \omega_N) = \omega_i$, $i \in \mathcal{V}$, $\omega := [\omega_1, \ldots, \omega_N] \in \Omega$. Let P_0 denote another probability measure on $(\Omega, \mathcal{F})$ given by $P_0(d\omega) := \prod_{i=1}^N \mu(d\omega_i)$. Then $P \ll P_0$ with the Radon–Nikodym derivative

$$\Lambda_{\mathcal{V}}(X_1(\omega), \ldots, X_N(\omega)) := \frac{dP}{dP_0}(\omega) = \frac{e^{H(X_1(\omega),\ldots,X_N(\omega))}}{\sum_{\bar{s}\in S^N} e^{H(\bar{s})}},$$

where H is of the form (3.2). For $B \subset \mathcal{V}$, we shall write $X_B := \{X_i,\ i \in B\}$ and $S_B := \prod_{i\in B} \hat{S}_i$, suitably ordered. Also define the interior of B, denoted by $\mathring{B}$, as the nodes in B that do not have a neighbour in B^c, and the boundary of B, denoted by ∂B, as the set of nodes in B that have at least one neighbour in B^c. Let $\tilde{B} := B^c \cup \partial B$. Thus $\partial B = B \cap \tilde{B} = \partial\tilde{B}$. We shall denote by P_B, P_{0B} the restrictions of P, P_0 to S_B with the corresponding Borel σ-fields, and write $\Lambda_B(X_B) := \frac{dP_B}{dP_{0B}}$. Then, because $X_{\mathcal{V}}$ is a Markov random field, by Theorem 2.2, we have $\Lambda(X_{\mathcal{V}}) = \Lambda_B(X_B)\Lambda_{\tilde{B}}(X_{\tilde{B}})$. Now consider X_B. It is easy to check that this too is a Markov random field, now indexed by nodes in B, and therefore for any $C \subset B$, we can write $\Lambda_B(X_B) = \Lambda_C(X_C)\Lambda_{\tilde{C}}(X_{\tilde{C}})$, where $\tilde{C} := B \setminus \mathring{C}$. We can repeat this procedure till it is no longer possible to do so, i.e., when Ω has been split into sets B_i such that $B_i = \partial B_i$, $i = 1, \ldots, k$ (say). Then it is easy to see that the B_i's must be cliques.

Next we prove the converse, i.e., that given a Gibbs distribution, there is a Markov random field that corresponds to it. Let $\mathcal{C}_i$ denote the set of cliques containing a prescribed $i \in \mathcal{V}$ and let H_C^i denote the value of $H_C, C \in \mathcal{C}_i$, with the value of X_i frozen at s_i (say). Let H_C^{-i} denote the value of $H_C, C \in \mathcal{C}_i$, with the value of X_i frozen at some $s \neq s_i$. Then from (3.1)–(3.2),

$$\begin{aligned} P(X_i = s_i \mid X_j = s_j,\ j \neq i) &= \frac{e^{\sum_{C\in\mathcal{C}_i} H_C^i + \sum_{C\notin\mathcal{C}_i} H_C}}{e^{\sum_{C\in\mathcal{C}_i} H_C + \sum_{C\notin\mathcal{C}_i} H_C}} \\ &= \frac{e^{\sum_{C\in\mathcal{C}_i} H_C^i}}{e^{\sum_{C\in\mathcal{C}_i} H_C}} = \frac{e^{\sum_{C\in\mathcal{C}_i} H_C^i}}{e^{\sum_{C\in\mathcal{C}_i} H_C^i + \sum_{C\in\mathcal{C}_i} H_C^{-i}}} \\ &= P(X_i = s_i \mid X_j = s_j,\ j \neq i,\ j \in \mathcal{C}_i). \end{aligned}$$

This proves the converse and completes the proof. ∎

REMARK 3.3. The resulting clique decomposition at the end of the above procedure is not unique and can depend, e.g., on the exact sequence of splittings used.

REMARK 3.4. The usual 'positivity condition' assumed in proofs of the equivalence of a Gibbs ensemble and a Markov random field is built into the absolute continuity hypothesis in the background.

Funding. Work supported in part by a grant from Google Research Asia.

References

[1] J. Besag, *Spatial interaction and the statistical analysis of lattice systems*, J. Roy. Statist. Soc. Ser. B 36 (1974), 192–236.
[2] P. Clifford, *Markov random fields in statistics*, in: Disorder in Physical Systems: A Volume in Honour of John M. Hammersley, Oxford Univ. Press, 1990, 19–32.
[3] R. L. Dobrushin, *The description of a random field by means of conditional probabilities and conditions of its regularity*, Theory Probab. Appl. 13 (1968), 197–224.
[4] G. R. Grimmett, *A theorem about random fields*, Bull. London Math. Soc. 5 (1973), 81–84.
[5] J. M. Hammersley and P. Clifford, *Markov fields on finite graphs and lattices*, unpublished manuscript, 1971.
[6] J. Moussouris, *Gibbs and Markov random systems with constraints*, J. Statist. Phys. 10 (1974), 11–33.
[7] C. J. Preston, *Generalized Gibbs states and Markov random fields*, Adv. Appl. Probab. 5 (1973), 242–261.
[8] S. Sherman, *Markov random fields and Gibbs random fields*, Israel J. Math. 14 (1973), 92–103.
[9] T. P. Speed, *A note on nearest-neighbour and Gibbs probabilities*, Sankhyā 4 (1979), 184–197.
[10] F. Spitzer, *Markov random fields and Gibbs ensembles*, Amer. Math. Monthly 78 (1973), 142–154.

Vivek S. Borkar
Department of Electrical Engineering
Indian Institute of Technology Bombay
Powai, Mumbai 400076, India
E-mail: borkar.vs@gmail.com