Gibbs Random Fields and Markov Random Fields with Constraints

Levent Onural

Abstract

It was shown many times in the literature that a Markov random field is equivalent to a Gibbs random field when all realizations of the field have non-zero probabilities; the proofs are rather complicated. A simpler proof, which is based directly on simple probability theory, is presented. Furthermore, it is shown that the equivalence is still valid when there are constraints (zero probability realizations) of any type. The equivalence extends to infinite size random fields, as well.

Index Terms

Markov random fields, Gibbs random fields, constrained Markov random fields, constrained Gibbs random fields

I. INTRODUCTION

It is well known that a Gibbs random field has an equivalent Markov random field, and vice versa, provided that all realizations have nonzero probabilities; this requirement is known as the “positivity condition” [1]–[4]. There are, however, many open questions when there are constraints, which in turn, impose zero probabilities to some outcomes (i.e., when there are impossible outcomes) [5]. Local constraints in the form of zero conditional probabilities are already incorporated into the theory [6]. Some cases with global constraints are also demonstrated [7], [8]. The positivity condition is sufficient but not necessary; indeed, there are hints that it might not be needed for the equivalence. Here in this paper, the basic steps and the associated proofs to construct Gibbs or Markov random fields are revisited. Simpler alternative proofs are provided for many cases. It is also proven that Gibbs random fields and Markov random fields are equivalent even if there are impossible (zero probability, forbidden) outcomes.

II. PRELIMINARIES

Let us start by the set of all outcomes \( \Omega \), with its elements \( a_i \) having all nonzero probabilities: \( P(a_i) = p_i > 0 \), for all \( a_i \). Let us assume that these probabilities of elements, \( P(a_i) = p_i \) may not be known, but all ratios, \( P(a_i)/P(a_j) = p_i/p_j = r_{ij} \) of such probabilities are known, for all \( i \) and \( j \). It is easy to show that these ratios uniquely specify the unknown \( P(a_i) \)'s for all \( i \). To show this, let us choose an arbitrary \( i \), and consider all ratios \( p_j/p_i \), for all \( j \). Clearly,

\[
\sum_j \frac{p_j}{p_i} = \frac{1}{p_i},
\]

which implies

\[
p_i = \frac{1}{\sum_j \frac{p_j}{p_i}},
\]

and therefore,

\[
p_j = \frac{p_i}{r_{ij}} = \frac{1}{r_{ij} \sum_j \frac{p_j}{p_i}}, \quad \forall j, \quad \text{and for any } i,
\]

where the sum over \( j \) indicates that the summation is running over all elements \( a_j \in \Omega \). Therefore, all unknown probabilities can be found from the known ratios of those probabilities. The reason to explicitly include Eq. (3) is to

L. Onural is with the Dept. of Electrical and Electronics Eng. of Bilkent University.
stress the fact that only one arbitrary element \( a_i \) may be chosen as the reference, and all ratios of probabilities are given with respect to \( p_i \) as \( p_j/p_i \), for all \( j \). Indeed, it is sufficient to know only such a set of ratios of probabilities, instead of all ratios, to find probabilities, \( p_j \), for all \( j \).

Any constraint partitions the sample space \( \Omega \) into two sets, \( C_k \), and its complement \( C_k^c \). There may be many such constraints, labeled by the index \( k \), and in such a case, we are interested in the intersection \( C = \bigcap_k C_k \). Therefore, any \( a_i \in C \) satisfies all constraints, and again, the sample space is partitioned into \( C \) and \( C^c \).

Now, let us concentrate on the conditional probabilities \( P(a_i \mid C) \). Simply, from the definition of the conditional probabilities,

\[
P(a_i \mid C) = \begin{cases} \frac{P(a_i \cap C)}{P(C)} = \frac{P(a_i)}{P(C)} & \text{if } a_i \in C \\ 0 & \text{else} \end{cases}
\]

Clearly, \( P(C) > 0 \) provided that \( C \) is not the empty set, since all \( a_i \)'s have positive probabilities.

Now, let us focus only on those outcomes in \( C \); by the way, conditional probabilities are probabilities, i.e., they satisfy all axioms and properties of a probability structure. Therefore, we simply state that, i) \( P(a_i \mid C) > 0 \) for \( a_i \in C \), and, ii)

\[
\frac{P(a_j \mid C)}{P(a_i \mid C)} = \frac{P(a_j)}{P(a_i)} = \frac{p_j}{p_i} = r_{ij}.
\]

Therefore, given two outcomes where both satisfy the constraints, the ratio of their conditional probabilities is still the same as the ratio of their (unconditional) probabilities, as expected. This observation leads us to the conclusion that for any subset \( C \), if all the ratios of probabilities (or equivalently, conditional probabilities, given that both realizations are in \( C \), \( r_{ij} \)'s, are known, then the conditional probabilities \( P(a_j \mid C) \) for all \( j \neq a_j \in C \) are also known as they are induced by these ratios, following the same steps are in Eqs. (1-3); and furthermore, this will also induce the (unconditional) probabilities of those elements, as well, if needed.

### III. GRF-MRF Equivalence in the Presence of Constraints

Indeed, the equivalence of a GRF to its corresponding MRF, and vice versa, is proven using these ratios of probabilities; the invariance of these ratios, whether for the unconditional case or the conditional case, assures the GRF-MRF equivalence even if there are also constraints. And this is true for any constraint, as long as the set of outcomes satisfying the constraints form a non-empty set. As already stated, constraints mean zero probabilities as also indicated by Eq. (4).

Starting from the general, and simple, observations above, we can now turn our attention to GRFs and MRFs. (We will use undertilde to describe random variables; no undertilde will be used for numerical values that these random variables take. We will use bold fonts to represent vectors (arrays). For example, \( P(\mathbf{x} = \mathbf{x}) \) means the “probability that the vector random variable \( \mathbf{x} \) takes the specific vector value \( \mathbf{x} \). We will also use the notation \( P_{\mathbf{x}}(\mathbf{x}) \) for the same purpose, whenever we feel this notation is more appropriate. Indeed, whenever there is no ambiguity in the meaning, we will also use the shortened notation \( P(\mathbf{x}) \) to represent the same probability as described above. There is no specific meaning attached to lower case or upper case symbols.) As usual, we assume a set of indexed random variables, \( x_l; \) the number of elements in the set may be finite of infinite. The index could be called the “site”, but actually it may or may not be associated with a physical location. The collection of all of those random variables \( x_l \) for all \( l \), is a vector random variable \( \mathbf{X} \); we will call \( \mathbf{X} \) as the “random pattern”. A realization of \( \mathbf{X} \), is denoted by \( \mathbf{x} \) which is a pattern over all sites.

As proven many times, conditional probabilities given for the MRF induce joint probabilities for all outcomes \( \mathbf{X} = \mathbf{X} \), under the positivity condition \([1][4][5]\). However, the provided proofs are unnecessarily complicated and lengthy. Instead, a simple proof is a direct consequence of the trivial discussion on the ratios of probabilities, as presented above:

Let us first prove that an MRF is also a GRF. As before, the set \( C \) contains all outcomes that satisfy the constraints; therefore, the elements in \( C \) all have nonzero probabilities. So, we can safely form the ratios of such probabilities, as,

\[
\frac{P(\mathbf{X} = \mathbf{X}_j \mid C)}{P(\mathbf{X} = \mathbf{X}_i \mid C)} = \frac{P(\mathbf{X} = \mathbf{X}_j)}{P(\mathbf{X} = \mathbf{X}_i)} = r_{ij} \quad \forall \; i,j.
\]
Let us choose patterns $X$ from $C$ such that their values at all sites, except the specific but arbitrary site $l$, are the same. The number of such patterns is at least one; we are interested in cases where this number is greater than one since otherwise the rest of the discussion is trivial. By the way, the number of such patterns could be quite small as a consequence of the constraints, and each such distinct pattern has a different realization for $x_l$. Assuming that there are more than one such patterns, we can write,

$$ P (X = X_j | C) = P (x_l = v_j | x_m = x_m \forall m \neq l, C) P (x_m = x_m \forall m \neq l, C) ,$$

where $v_j$ is the realization (value) of the random variable $x_l$ at site $l$ within the pattern $X_j$. Therefore, the ratio in Eq.6

$$ \frac{P (X = X_j | C)}{P (X = X_i | C)} = \frac{P (x_l = v_j | x_m = x_m \forall m \neq l, C) P (x_m = x_m \forall m \neq l, C)}{P (x_l = v_i | x_m = x_m \forall m \neq l, C) P (x_m = x_m \forall m \neq l, C)} = \frac{P (x_l = v_j | x_m = x_m \forall m \neq l, C) P (x_m = x_m \forall m \neq l, C)}{P (x_l = v_i | x_m = x_m \forall m \neq l, C) P (x_m = x_m \forall m \neq l, C)} = \frac{P (x_l = v_j | x_m = x_m \forall m \neq l, C)}{P (x_l = v_i | x_m = x_m \forall m \neq l, C)} = r_{ij} ,$$

where we used the Markovianity, and $\eta_l$ is the neighborhood associated with site $l$ (see, for example, [5], for the definition and the properties of the neighborhood). Therefore, if all conditional probabilities,

$$ P (x_l = v_l | x_m = x_m \forall m \in \eta_l, C) ,$$

for all $l$ and for all allowed values $v_l$ for that location are known when there are constraints, or no constraints (i.e., when $C = \Omega$, the sample space), and if the Markovianity as indicated by

$$ P (x_l = v_l | x_m = x_m \forall m \neq l) = P (x_l = v_l | x_m = x_m \forall m \neq l)$$

holds for all $l$ and $v_l$ (i.e., if the field is a MRF), we can go backwards through the arguments and obtain the result:

Given all conditional probabilities $P (x_l = v_l | x_m = x_m \forall m \in \eta_l)$ and the set indicated by constraints $C$, all ratios $r_{ij}$ for all $X_i$ and $X_j$ in $C$ are known; and from that (due to Eqs.(10)) all conditional probabilities $P (X_i | C)$ for all $X_i \in C$ are known and they are positive; and therefore, we can always write $P (X_i | C) = ke^{-U(X_i)}$ due to nonzero value of $P (X_i | C)$, where $k$ is just a normalization constant to have the sum of all probabilities equal to one, i.e.,

$$ k = \sum_{all i : X_i \in C} e^{-U(X_i)} .$$

Furthermore, noting that,

$$ P(C) = \sum_{all i : X_i \in C} k e^{-U(X_i)} ,$$

we also know $P(C)$, and therefore, $P(X_i) = P (X_i | C) P(C)$ is also known, and is equal to $kP(C)e^{-U(X_i)}$, for all $X_i$, where $kP(C)$ is just the new normalization constant to have the sum of all probabilities equal to one. By the way, $U(X)$ is called the “energy” of the pattern $X$. Therefore, a conditional (constraint satisfying) MRF is also a conditional GRF. The terms “conditional Markov random fields (CMRF)” and “conditional Gibbs random
fields (CGRF)” are also used in [7], [8]. Please note that the proof presented above is also valid for the case where there are no constraints, i.e., when $C = \Omega$, and therefore, we have also provided an alternative proof that a MRF is equivalent to a GRF under the positivity condition, as well. We believe that this proof is a lot simpler and straightforward than other known proofs, as given, for example in [11–13].

Now we will prove that every conditional GRF is also a conditional MRF. If $X$ is a GRF, then each $X_i \in C$ has a probability in the form $ke^{-U(X_i)}$ as a consequence of the definition of a GRF; we have shown above that the (unconditional) probabilities of these patterns are also in the form $ke^{-U(X_i)}$, where $kP(C)e^{-U(X_i)}$ is just the related normalization constant. We can always decompose any function $U(X_i)$ into an additive form as,

$$U(X_i) = \sum_{c \in Q} V_c(X_i)$$  \hspace{1cm} (12)

where $c$ is a “clique” which simply means a subset of indices (“sites”) $l$. The set $Q$ is the set of all cliques. The number of elements in a clique ranges from zero up to the maximum number of indices (sites) in a pattern; therefore, that maximum is infinity for infinite size patterns. Therefore, the ratio of the probabilities of two patterns, $X_j$ and $X_i$, becomes,

$$\frac{P(X = X_j | C)}{P(X = X_i | C)} = \frac{P(X = X_j)}{P(X = X_i)} = \frac{e^{-U(X_j)}}{e^{-U(X_i)}} = \frac{\sum_{c \in Q} V_c(X_j)}{\sum_{c \in Q} V_c(X_i)}$$  \hspace{1cm} (13)

Now, let us assume that the patterns $X_i$, and $X_j$ differ only at one specific site $l$, and they have the same values at all other sites other than $l$. In that case, the above ratio becomes,

$$\frac{\sum_{c \in Q} V_c(X_j)}{\sum_{c \in Q} V_c(X_i)} = \frac{\left(\sum_{e \in D_j} V_e(X_j)\right)\left(\sum_{e \in D_j^c} V_e(X_i)\right)}{\left(\sum_{e \in D_j} V_e(X_i)\right)\left(\sum_{e \in D_j^c} V_e(X_j)\right)} = \frac{\sum_{c \in Q} V_c(X_j)}{\sum_{c \in Q} V_c(X_i)}$$  \hspace{1cm} (14)

where $D_j$ is the set of only those cliques which have different realizations for $X_i$, and $X_j$ over them, for a given $l$. In other words,

$$D_j = \{c \mid x_{mi} \neq x_{mj} \text{ for at least one } m \in c\}$$  \hspace{1cm} (15)

where $x_{mi}$ and $x_{mj}$ are the realizations at site $m$ of the patterns $X_i$, and $X_j$, respectively. Since $X_i$, and $X_j$ may differ only at site $l$, as a consequence of the above assumption, an equivalent definition of $D_l$ can be given by

$$D_l = \{c \mid l \in c\}$$  \hspace{1cm} (16)

Please note that $D_l^c$ is the complement of $D_l$ in $Q$: $D_l^c = Q \setminus D_l$. Noting, from Eq.(14), that

$$\frac{P(X = X_j | C)}{P(X = X_i | C)} = \frac{P(x_l = v_j | x_m = x_m \forall m \neq l, C)}{P(x_l = v_i | x_m = x_m \forall m \neq l, C)} = \frac{e^{-U(X_j)}}{e^{-U(X_i)}} = \frac{\left(\sum_{e \in D_l^c} V_e(X_j)\right)\left(\sum_{e \in D_l} V_e(X_i)\right)}{\left(\sum_{e \in D_l^c} V_e(X_i)\right)\left(\sum_{e \in D_l} V_e(X_j)\right)} = \frac{\sum_{c \in Q} V_c(X_j)}{\sum_{c \in Q} V_c(X_i)}$$  \hspace{1cm} (17)

We conclude that $x_m$, $m \in \eta_l$ are sufficient to compute each term of the last expression above, to yield,

$$\frac{P(x_l = v_j | x_m = x_m \forall m \neq l, C)}{P(x_l = v_i | x_m = x_m \forall m \neq l, C)} = \frac{P(x_l = v_j | x_m = x_m \text{ for } m \in \eta_l, C)}{P(x_l = v_i | x_m = x_m \text{ for } m \in \eta_l, C)}$$  \hspace{1cm} (18)

and complete the proof that every GRF is also a MRF. Please note that the proof is also valid if there are no constraints; i.e., when $C = \Omega$. 

Based on the discussions above, we can conclude that,

* The only requirement to write a probability in exponential form is the positivity of that probability; therefore, any field with positive probability realizations is a GRF.
* A GRF has an equivalent MRF, and vice versa, when all the realizations (outcomes) have positive probabilities.
* At this point, one could get the impression that MRFs, or equivalently GRFs, are mathematically so simple: provided that outcomes (patterns) in a set of outcomes all have non-zero probabilities, the set forms a MRF (GRF). This is true. However, the MRF model becomes useful only when the neighborhoods \( \eta_l \) are simple (small in size; i.e., have few indices (sites) in it for every \( l \)); equivalently, GRF model becomes useful when the set of cliques \( Q \) contains only simple and few components. An equivalent statement is that the clique potentials, \( V_c(X)'s \), are zero for most cliques (i.e., for most subsets \( c \) of sites). In other words, the benefits prevail only when the direct statistical interactions among the random variables at different sites are sparse. Obviously, every field with non-zero probability outcomes is a MRF when the sizes of the neighborhoods \( \eta_l 's \) reach the size of the entire field; or equivalently, every field with non-zero probability outcomes is a GRF when the sizes of \( c's \) in \( Q \) are allowed to reach the size of the entire field (i.e., \( D^c = \emptyset \)). In such cases, one may still call the field as a MRF with the neighborhood size equal to the size of the entire field; or may choose to leave those extreme cases out of the definition and say that those cases are not MRFs (GRFs); we prefer the first alternative in this paper.
* The neighborhood of a site is not necessarily near (here we assume that the sites form a lattice, and therefore, a distance measure is applicable) to that site; indeed, the simplicity (sparsity) of a neighborhood scheme is based on the number of sites in a neighborhood, and not where they are located.
* Extension of the proofs for the Markov case where the “interior” is no longer a single site, but more than one site, is straightforward.
* The discussion is valid both for causal fields, as well as non-causal ones; the causality is a direct consequence of the neighborhood shape (see, for example [9] for the definition of “causal” and, “non-causal” random fields).
* GRFs are usually simpler to handle than MRFs.
* Inclusion of constraints, which in turn means zero probability outcomes, does not violate the MRF-GRF equivalence or the basic structures of GRFs and the MRFs. This is a consequence of the observation that the constraints result in a smaller set of non-zero probability realizations, and within that smaller set, all other features that leads to the equivalence is still valid and the same as the unconstrained case.
* These conclusions are valid both for the finite size or infinite size random fields.
* Simple proofs are given for above statements; the proofs are direct consequences of elementary probability theory, and techniques.

The presented discussions clarify and provide answers to those long open questions related to constraints in MRFs or GRFs.

**REFERENCES**

1. J. M. Hammersley and P. Clifford, “Markov random fields on finite lattices and graphs” Unpublished, 1971. [Online] Available: [http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf](http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf)
2. J. Besag, “Spatial interaction and the statistical analysis of lattice systems” Journal of the Royal Statistical Society Series B (Methodological), vol. 36, no. 2, pp 192–236, 1974. [Online] Available: [http://www.jstor.org/stable/2984812](http://www.jstor.org/stable/2984812)
3. F. Spitzer, “Markov random fields and Gibbs ensembles” The American Mathematical Monthly, vol. 78, no. 2, pp 142–154, 1971. [Online] Available: [http://www.jstor.org/stable/2317621](http://www.jstor.org/stable/2317621)
4. V. Isham, “An introduction to spatial point processes and Markov random fields” International Statistical Review / Revue Internationale de Statistique, vol. 49, no. 1, pp 21–43, 1981. [Online] Available: [http://www.jstor.org/stable/1403035](http://www.jstor.org/stable/1403035)
5. H. Derin and P. A. Kelly, “Discrete-index Markov-type random processes” Proceedings of the IEEE, vol. 77, no. 10, pp 1485–1510, 1989. [Online] Available: [http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=40665](http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=40665)
6. J. Moussouris, “Gibbs and Markov random systems with constraints” Journal of Statistical Physics, vol. 10, no. 1, pp 11–33 1974. [Online] Available: [http://dx.doi.org/10.1007/BF01011714](http://dx.doi.org/10.1007/BF01011714)
7. L. Onural, “Conditional Markov random fields applied to textured fractal pattern generation” In Preprint Booklet, The First International Conference on Statistical Computing, pp. 666, Çeşme, İzmir, Turkey, 1987. (Published abstract, unpublished presented paper).
8. L. Onural, “Generating connected textured fractal patterns using Markov random fields” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 13, no. 8, pp 819–825 1991. [Online] Available: [http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=85673](http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=85673)
9. C. S. Won and R. M. Gray, Stochastic Image Processing, Springer, 2004.