Journal of Multivariate Analysis **73**, 199–220 (2000)
doi:10.1006/jmva.1999.1878, available online at http://www.idealibrary.com on IDEAL®

The Construction of Multivariate Distributions from Markov Random Fields

Mark S. Kaiser and Noel Cressie

Department of Statistics and Statistical Laboratory, Iowa State University

Received September 25, 1996

We address the problem of constructing and identifying a valid joint probability density function from a set of specified conditional densities. The approach taken is based on the development of relations between the joint and the conditional densities using Markov random fields (MRFs). We give a necessary and sufficient condition on the support sets of the random variables to allow these relations to be developed. This condition, which we call the Markov random field support condition, supercedes a common assumption known generally as the positivity condition. We show how these relations may be used in reverse order to construct a valid model from specification of conditional densities alone. The constructive process and the role of conditions needed for its application are illustrated with several examples, including MRFs with multiway dependence and a spatial beta process.
© 2000 Academic Press

AMS subject classifications: 62E99, 62H11, 62H05.

Key words and phrases: beta conditionals, conditional model specification, Hammersley–Clifford theorem, negpotential function, positivity condition.

1. INTRODUCTION

Conditional specification of multivariate statistical models has a long history, with early contributions from Whittle (1963), Brook (1964), Bartlett (1966), and Besag (1974). More recently, Gelman and Meng (1991) have discussed multivariate distributions with Gaussian conditional distributions. Cramer (1998) developed a conditional version of iterative proportional fitting for finding a joint distribution with prescribed conditional and marginal constraints, with emphasis on the Gaussian case. Arnold *et al.* (1992) have presented a comprehensive treatment of bivariate distributions developed from the concept of compatible conditional specifications (Arnold and Press, 1989). Here, we consider an approach originally applied to nearest-neighbor structures by Brook (1964) and later to spatial lattices by Besag (1974). In particular, Besag (1974) gave a necessary form for parameterization of one-parameter exponential family conditional distributions, assuming that dependence was expressed in a "pairwise-only" fashion, and used this result to develop the class of

0047-259X/00 $35.00
Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.

exponential family "auto-models," which are commonly employed in the spatial statistics literature. The joint distributions that correspond to specification of an exponential family auto-model may be readily seen to be special cases of the characterization result for exponential families due to Arnold and Strauss (1991).

The approach used by Besag (1974), in his development of auto-models, begins with specification of the conditional dependencies present among a finite set of random variables that result in a Markov random field (MRF). These conditional dependencies define which of the entries of the multivariate random vector can be considered as neighbors of each other. Although it is generally recognized that there is nothing inherently spatial about MRFs, the approach discussed in this article has been developed largely in spatial statistics.

In this article, we present the basic results needed for a MRF approach to modeling a finite set of random variables. In so doing, we redefine the central quantity of concern, the "negpotential function"; offer a weaker support condition than the often-invoked "positivity condition"; give a new proof of an important theorem due to Hammersley and Clifford (1971); and strengthen a result due to Cressie and Lele (1992) on construction of valid MRFs through specification of conditional distributions. Dobrushin (1968) has considered the countably infinite case where the random variables are defined on the integer-valued d-dimensional lattice, but his results are weaker and not constructive. We believe the results for finite MRFs, collected, modified, and extended here, make the MRF approach to model specification more readily available to statistical modelers than it has been in the past.

We shall deal only with probability measures that admit density or mass functions and will take a MRF to be a stochastic process $\mathbf{Y} \equiv \mathbf{Y}(\mathscr{L})$ defined on the lattice $\mathscr{L}$, a countable subset of $\Re^k$. Henceforth, reference to density functions will be understood to mean functions that return either probability density or mass values. In this paper, we are concerned exclusively with the case of a finite lattice, where a complete listing off all conditional dependencies among elements of $\mathbf{Y}$ is obtainable from a set of dependence index sets (i.e., neighbors), defined more formally in Section 2. To any such MRF corresponds an acyclic algebraic graph with undirected edges (e.g., Whittaker 1990). In fact, graphs in combination with (measurable) random functions on probability spaces may be used to define MRFs (e.g., Speed, 1978).

Our results concern the construction of a joint distribution for $\mathbf{Y}$, given a complete set of full (univariate) conditional distributions. Since these conditional distributions are the building blocks of a MRF (Besag, 1974), we call our development the "Markov random field" approach to model formulation.

While the construction of graphs concerns directly the deliniation of *conditional independencies* present among the individual random variables of a random field, the goal of statistical modeling is to give (probabilistic) form to the full set of joint *dependencies* present. There are several approaches other than conditional modeling for accomplishing this goal. The construction of models with given marginal distributions has received much attention (e.g., Cohen, 1984; Rüschendorf, 1985; Marshall and Olkin, 1988; Cuadras, 1992; Li *et al.*, 1996), although this approach does not lead to models for specific dependencies but rather to ranges of permissible dependencies. More closely related to our approach is the use of graphical structures that contain greater complexity than the simple undirected graphs of MRFs, such as the "multivariate regression" and "block regression" chain graphs discussed by Cox and Wermuth (1993) and in references therein. One such class of graphical models may be used to describe dependencies in terms of a covariance matrix which, for full parametric modeling with continuous variables, essentially implies the use of a Gaussian model.

We develop the MRF approach to model formulation in three stages. A quantity called the "negpotential function" is defined and several results related to this function are derived, assuming the existence of a joint probability density function for a collection of random variables. This is the topic of Section 2. Then, in Section 3, these results are related to a particular set of conditional density functions under an assumption that the conditional functions exist. Finally, the question of central concern to statistical modelers is addressed, namely, when and how can a joint probability density function be constructed from specification of a set of conditional functions alone? This question is answered in Section 4, and several illustrative examples are presented in Section 5.

2. MRFs AND THE NEGPOTENTIAL FUNCTION

We begin with a finite collection of univariate random variables $\mathbf{Y} = \{ Y(\mathbf{s}_i) : i = 1, ..., n\}$, where $\mathbf{s}_i$ denotes the "location" of $Y(\mathbf{s}_i)$ in a random field. This location may be an actual geographical location, but it may also refer to a time of occurrence in longitudinal data, or a grouping mechanism in a subsampling or repeated-measures study. For example, in a spatial problem, we might take $\mathbf{s}_i \equiv (u_i, v_i)$, where u_i is longitude and v_i is latitude; in a repeated-measures study we might take $\mathbf{s}_i \equiv (k, j)$, where k indexes subject and j indexes observation number; and in a multivariate time-series application we might take $\mathbf{s}_i \equiv (k, t_k(j))$, where k indexes the variable and $t_k(j)$ is the time at which the jth observation of the kth variable is obtained.

Let $N \equiv \{\mathbf{s}_i: i=1, ..., n\}$, $D \equiv (1, 2, ..., n)$, and assume that a joint density $g_{1, ..., n}(\mathbf{y}) = g_D(\mathbf{y})$ exists for $\mathbf{Y}$. Define the *dependence sets* or *neighborhoods* as

$$N_i \equiv \{\mathbf{s}_h: g_i(y(\mathbf{s}_i) \mid \{y(\mathbf{s}_j): j \neq i\}) \text{ depends functionally on } y(\mathbf{s}_h); h \neq i\};$$
$$i = 1, ..., n,$$

where "depends functionally on $y(\mathbf{s}_h)$" means that the functional form of $g_i(y(\mathbf{s}_i) \mid \{y(\mathbf{s}_j): j \neq i\})$ explicitly contains $y(\mathbf{s}_h)$. Corresponding to the (arbitrary) indexing of the locations, one may also define the *dependence index sets* as

$$D_i \equiv \{h: \mathbf{s}_h \in N_i\}; \qquad i = 1, ..., n.$$

It will be convenient in our development to work mostly with $\{D_i: i=1, ..., n\}$, although it should be remembered that the dependencies come from properties of the process described through the more basic $\{N_i: i=1, ..., n\}$. Any finite set of random variables with a joint distribution, and whose conditional probability distributions define a dependence structure through dependence index sets $\{D_i: i=1, ..., n\}$, defines a Markov random field. In terms of the graph associated with the collection of random variables $\mathbf{Y}$, the nodes of the graph are the set $\{\mathbf{s}_i: i=1, ..., n\}$, and $\mathbf{s}_h \in N_i$ if and only if $\{\mathbf{s}_i, \mathbf{s}_h\}$ is an edge of the graph.

Define the support of $g_D(\cdot)$ as $\Omega \equiv \{\mathbf{y}: g_D(\mathbf{y}) > 0\}$. We choose a particular value $\mathbf{y}^* \equiv (y^*(\mathbf{s_1}), ..., y^*(\mathbf{s_n}))^T \in \Omega$ such that $g_D(\mathbf{y}^*)$ is finite and define the negpotential function,

$$Q(\mathbf{y}) \equiv \log\{g_D(\mathbf{y})/g_D(\mathbf{y}^*)\}; \qquad \mathbf{y} \in \Omega. \tag{1}$$

In the past (e.g., Besag, 1974), $\mathbf{y}^*$ has been chosen to be $\mathbf{0} = (0, ..., 0)^T$, that is, $\mathbf{y}^* = (y^*(\mathbf{s}_1), ..., y^*(\mathbf{s}_n))^T = \mathbf{0}$. Notice that our choice of $\mathbf{y}^*$ is completely general and requires only that $\mathbf{y}^* \in \Omega$ and $g_D(\mathbf{y}^*)$ is finite. Thus, (1) represents a departure from the usual definition of the negpotential function; see Moussouris (1974) for a definition similar to ours. It will be seen below that the extra generality of (1) is needed. For example, in discussing exponential family auto-models, Besag (1974) noted that the use of $\mathbf{0}$ causes problems for many gamma conditional distributions but indicated that transformations, such as the log transformation, could be made such that transformed variables also have exponential family distributions with $\mathbf{0} \in \Omega$. This is true, but it is not true that the distributional forms remain unchanged under such transformations. Thus, one cannot use the approach of Besag to justify MRF formulation of a model with gamma conditionals.

Several fundamental results of the MRF approach to statistical model formulation, proved by Besag (1974) using (1) with $\mathbf{y}^* = \mathbf{0}$, continue to hold for the more general definition with any particular value $\mathbf{y}^* \in \Omega$. The first result is that the negpotential function (1) may be written as the expansion

$$\begin{aligned} Q(\mathbf{y}) = &\sum_{1 \leqslant i \leqslant n} H_i(y(\mathbf{s_i})) + \sum\sum_{1 \leqslant i < j \leqslant n} H_{i,j}(y(\mathbf{s_i}), y(\mathbf{s_j})) \\ &+ \sum\sum\sum_{1 \leqslant i < j < k \leqslant n} H_{i,j,k}(y(\mathbf{s_i}), y(\mathbf{s_j}), y(\mathbf{s_k})) \\ &+ \cdots \\ &+ H_{1,2,\ldots,n}(y(\mathbf{s_1}), y(\mathbf{s_2}), \ldots, y(\mathbf{s_n})); \qquad \mathbf{y} \in \Omega. \end{aligned} \tag{2}$$

In what is to follow, we give a set of H-functions that yield $Q(\cdot)$ via (2).

Now, let $\mathbf{j}_m \equiv (j(1), j(2), \ldots, j(m))^T$ denote a generic index of $m \leqslant n$ distinct index values; that is, $\mathbf{j}_m$ consists of a particular ordering of any m distinct indices chosen from $\{1, \ldots, n\}$. Define the sets

$$T_m(p) \equiv \{\text{all distinct } p\text{-tuples formed from elements of } \mathbf{j}_m\}; \qquad p \leqslant m.$$

Then, for all $m = 1, \ldots, n$ and all $\mathbf{j}_m$, it is straightforward to show that H-functions that yield (2) can be calculated via the formula

$$\begin{aligned} &H_{j(1),\ldots,j(m)}(y(\mathbf{s}_{j(1)}), \ldots, y(\mathbf{s}_{j(m)})) \\ &\quad\equiv \sum_{t=0}^{m-1} \sum_{\mathbf{j}_{m-t} \in T_m(m-t)} (-1)^t\, Q(\{\mathbf{y}: y(\mathbf{s}_h) = y^*(\mathbf{s}_h); h \notin \mathbf{j}_{m-t}\}), \end{aligned} \tag{3}$$

where $y^*(\mathbf{s}_h)$ is the hth element of $\mathbf{y}^*$, and assuming that all arguments at which $Q(\cdot)$ is evaluated are elements of the joint support Ω. Relations (2) and (3) are analogous to the well known Möbius inversion formulas (e.g., Lauritzen, 1996, p. 239). We shall now investigate properties of the H-functions given by (3). Note that we have defined functions $H_{j(1),\ldots,j(m)}(\cdot)$ for any $1 \leqslant m \leqslant n$ and any $j(1), \ldots, j(m) \in \{1, \ldots, n\}$, although only some of these functions appear in the expansion (2). For example, while (3) results in a valid definition for $H_{1,2,3}(\cdot)$, $H_{2,1,3}(\cdot)$, and so forth, only $H_{1,2,3}(\cdot)$ appears in (2). If we begin with a joint density $g_D(\cdot)$, all of these H-functions will be the same. That is, the functions defined in (3) are invariant to permutation of the indices contained in each $\mathbf{j}_m$. Given a joint, this fact is seemingly trivial, since any joint is invariant to permutation of its arguments, and it is the joint that defines $Q(\cdot)$ in (1), and thus also the H-functions in (3). However, the point will be of central importance in Section 4, where we wish to verify the existence of a joint using a set of H-functions that are *constructed* from a set of specified conditionals; this construction is accomplished using the results contained in Section 3. If we use such constructed H-functions in (2) to try and recover a joint distribution, it will be crucial that they are invariant to permutation of their indices.

Besag (1974) assumed a restriction on the support of $g_D(\mathbf{y})$ called the "positivity condition," which states that $\Omega = \Omega_1 \times \cdots \times \Omega_n$, where Ω_i is the support of the marginal density of $Y(\mathbf{s}_i)$; $i = 1, ..., n$. The positivity condition is sufficient to ensure that all components of the summations in all expressions given by (3) are defined, that is, that all of the arguments at which $Q(\cdot)$ is evaluated are contained in Ω. However, it is not necessary, as evidenced by examples given in Moussouris (1974) and by the following result. First, we define $\Omega_{j(1), ..., j(m)}$ as the support of the marginal density $g_{j(1), ..., j(m)}(\cdot)$ and Φ_i as the support of $g_{D\backslash i}(\cdot)$.

THEOREM 1. *Let $g_D(\mathbf{y})$ be a joint density for $\mathbf{Y}$ and suppose that a value $\mathbf{y}^* \in \Omega$ has been chosen to define the negpotential function* (1) *on Ω. Then functions $\{H_{j(1), ..., j(m)}(\cdot)\}$ given by* (3) *for all $m \in \{1, ..., n\}$, all distinct $j(1), ..., j(m) \in \{1, ..., n\}$, and all $(y(\mathbf{s}_{j(1)}), ..., y(\mathbf{s}_{j(m)}))^T \in \Omega_{j(1), ..., j(m)}$, can be combined to yield* (2), *defined on Ω, if and only if*

$$\{y_i^*(\mathbf{s}_i)\} \times \Phi_i \subseteq \Omega; \qquad i = 1, ..., n. \tag{4}$$

Proof. First, suppose that (4) does not hold. Then, for some $q \in \{1, ..., n\}$ there exists a value $\mathbf{y}^@ \equiv (y^@(\mathbf{s}_1), ..., y^@(\mathbf{s}_{q-1}), y^@(\mathbf{s}_{q+1}), ..., y^@(\mathbf{s}_n))^T$ and a value $y^\circ(\mathbf{s}_q)$ such that $(y^\circ(\mathbf{s}_q), \mathbf{y}^@)^T \in \Omega$, but $(y^*(\mathbf{s}_q), \mathbf{y}^@)^T \notin \Omega$. Then,

$$g_{q, D\backslash q}(y^*(\mathbf{s}_q), \{y^@(\mathbf{s}_h) : h \neq q\}) = 0,$$

so that all $H_{j(1), ..., j(n-1)}(\mathbf{y}^@)$ in (3) will be undefined, since they each will contain terms of the form $\log(u/v)$ in which $u = 0$. One of these functions must appear in the right hand side of expansion (2) for $Q(y^\circ(\mathbf{s}_q), \mathbf{y}^@)$, so that even though the left-hand side of (2) is defined, the right-hand side is not. Therefore, (4) is necessary for the expansion (2) to be defined for all $\mathbf{y} \in \Omega$.

To prove sufficiency, note that the condition that all arguments of $Q(\cdot)$ in the right hand side of (3) be elements of the joint support Ω, is equivalent to a condition that, for all $m = 1, ..., (n-1)$,

$$\{(y^*(\mathbf{s}_{j(1)}), ..., y^*(\mathbf{s}_{j(m)}))\} \times \Phi_{j(1), ..., j(m)} \subseteq \Omega, \tag{5}$$

where $j(1), ..., j(m)$ are any m distinct location indices and $\Phi_{j(1), ..., j(m)}$ is the support of the density function for $\{Y(\mathbf{s}_h) : h \neq j(1), ..., j(m)\}$. The case $m = n$ is covered by the assumption that $\mathbf{y}^* \in \Omega$. Now, suppose that (4) holds, and choose any $m \in \{1, ..., (n-1)\}$, and any $j(1), ..., j(m) \in \{1, ..., n\}$. Let $\mathbf{y}^\dagger \equiv \{y^\dagger(\mathbf{s}_h) : h \neq j(1), ..., j(m)\}$ be any value contained in $\Phi_{j(1), ..., j(m)}$.

Then there must be at least one value $\mathbf{y}^\circ \equiv \{ y^\circ(\mathbf{s}_k) : k = j(1), ..., j(m)\}$ contained in $\Omega_{j(1), ..., (jm)}$ such that $(\mathbf{y}^\circ, \mathbf{y}^\dagger) \in \Omega$.

Now, $(\mathbf{y}^\circ, \mathbf{y}^\dagger) \in \Omega$ implies that $(\{ y^\circ(\mathbf{s}_k) : k = j(2), ..., j(m)\}, \mathbf{y}^\dagger) \in \Phi_{j(1)}$, so that, from (4), $(y^*(\mathbf{s}_{j(1)}), \{ y^\circ(\mathbf{s}_k) : k = j(2), ..., j(m)\}, \mathbf{y}^\dagger) \in \Omega$. Then it follows that $(y^*(\mathbf{s}_{j(1)}), \{ y^\circ(\mathbf{s}_k) : k = j(3), ..., j(m)\}, \mathbf{y}^\dagger) \in \Phi_{j(2)}$ from which (4) implies that $(y^*(\mathbf{s}_{j(1)}), y^*(\mathbf{s}_{j(2)}), \{ y^\circ(\mathbf{s}_k) : k = j(3), ..., j(m)\}, \mathbf{y}^\dagger) \in \Omega$. Continuing in this fashion for $j(3), ..., j(m)$ shows that $(y^*(\mathbf{s}_{j(1)}), ..., y^*(\mathbf{s}_{j(m)}), \mathbf{y}^\dagger) \in \Omega$. Since the choices of $m, j(1), ..., j(m)$, and $\mathbf{y}^\dagger$ are arbitrary, condition (6) holds, and all arguments at which $Q(\cdot)$ is evaluated in (3) are elements of Ω. Thus, all H-functions in the expansion (2) are defined. ∎

We call condition (4) the *MRF support condition* and note that while it is not needed for the negpotential function (1) to be defined (for this, one only needs a $\mathbf{y}^* \in \Omega$), it is needed for the validity of the expansion (2), in terms of the H-functions defined by (3). In discussion of a paper on Markov Chain Monte Carlo simulation, Besag (1994, Lemma 0.1) shows that the positivity condition is not needed for what he calls the "Brook expansion" and what Cressie (1993, p. 412) calls the "factorization theorem." The condition given by Besag (1994) concerns primarily the question of when, given a joint distribution and its full set of conditionals, one may simulate from that joint distribution using the conditionals. Besag's (1994) condition is weaker than the MRF support condition given in Theorem 1, as the simple example in Section 5.2 illustrates. There, we give a full set of conditionals that do not allow *construction* of a corresponding joint using the methods of this paper since the MRF support condition is not satisfied, yet the correct joint could easily be *simulated* from using the Lemma of Besag (1994).

We now adapt another useful result of Besag (1974) to our formulation. Assume the MRF support condition and write $\mathbf{y}_i \equiv (y(\mathbf{s}_1), ..., y(\mathbf{s}_{i-1}), y^*(\mathbf{s}_i), y(\mathbf{s}_{i+1}), ..., y(\mathbf{s}_n))$; then,

$$\exp(Q(\mathbf{y}) - Q(\mathbf{y}_i)) = \frac{g_D(\mathbf{y})}{g_D(\mathbf{y}_i)} = \frac{g_i(y(\mathbf{s}_i) \mid \{ y(\mathbf{s}_j) : j \neq i\})}{g_i(y^*(\mathbf{s}_i) \mid \{ y(\mathbf{s}_j) : j \neq i\})}, \tag{6}$$

for any $\mathbf{y} \in \Omega$ and $\{ y(\mathbf{s}_j) : j \neq i\} \in \Phi_i$. The proof of (6) follows from (1) and (4).

3. CONDITIONAL PROBABILITIES AND THE HAMMERSLEY–CLIFFORD THEOREM

While derivation of the results given in (2) and (6) change little from those using the usual definition of the negpotential function as

$Q(\mathbf{y}) \equiv \log\{g_D(\mathbf{y})/g_D(\mathbf{0})\}$, Besag's proof of the Hammersley–Clifford theorem (Hammersley and Clifford, 1971; Clifford, 1990) no longer holds for the definition (1). Our alternative proof of this important theorem will depend on two lemmas. The first is an extension of a result due to Cressie (1993, p. 416) and is stated here using the the notation of Section 2.

LEMMA 1. *Assume the MRF support condition and the existence of a joint density function $g_D(\mathbf{y})$ for $\mathbf{Y}$. Let $\mathbf{j}_m \equiv (j(1), j(2), ..., j(m))$ denote a generic index of $m \leqslant n$ distinct index values, and without loss of generality let $j(1) = i$ for some $i \in \{1, ..., n\}$. Also, let $\mathbf{j}_m^{-i} \equiv (j(2), ..., (jm))$ and let $T_m^{-i}(p)$ be the set of all distinct p-tuples formed from $\mathbf{j}_m^{-i}$, $p \leqslant m-1$. Then, $H_{i, ..., j(m)}$ in (3) may be written in terms of conditional probability density functions as*

$$
\begin{aligned}
&H_{i, j(2), ..., j(m)}(y(\mathbf{s}_i), y(\mathbf{s}_{j(2)}), ..., y(\mathbf{s}_{j(m)})) \\
&\quad = \sum_{t=0}^{m-2} \sum_{\mathbf{j}_{m-t}^{-i} \in T_M^{-i}(m-t)} \\
&\qquad \times \left\{ (-1)^{t-1} \log \left[\frac{g_i(y(\mathbf{s}_i) \mid \{y(\mathbf{s}_k) : k \in \mathbf{j}_{m-t}^{-i}\}, \{y^*(\mathbf{s}_h) : h \notin \mathbf{j}_{m-t}\})}{g_i(y^*(\mathbf{s}_i) \mid \{y(\mathbf{s}_k) : k \in \mathbf{j}_{m-t}^{-i}\}, \{y^*(\mathbf{s}_h) : h \notin \mathbf{j}_{m-t}\})} \right] \right\} \\
&\qquad + (-1)^{m-1} \log \left[\frac{g_i(y(\mathbf{s}_i) \mid \{y^*(\mathbf{s}_j) : j \neq i\})}{g_i(y^*(\mathbf{s}_i) \mid \{y^*(\mathbf{s}_j) : j \neq i\})} \right].
\end{aligned}
\tag{7}
$$

Proof. Proof of this lemma follows by initally noting that $Q(\mathbf{y}^*) = 0$, so that adding or subtracting this function from (3) does not change the resulting H-function. This, combined with manipulation of the summations in (3), shows that

$$
\begin{aligned}
&H_{i, j(2), ..., j(m)}(y(\mathbf{s}_i), ..., y(\mathbf{s}_{j(m)})) \\
&\quad = \sum_{t=0}^{m-2} \sum_{\mathbf{j}_{m-t}^{-i} \in T_M^{-i}(m-t)} \{(-1)^{t-1} [Q(\mathbf{y} : \{y(\mathbf{s}_h) = y^*(\mathbf{s}_h); h \notin \mathbf{j}_{m-t}\}) \\
&\qquad - Q(\mathbf{y}_i : \{y(\mathbf{s}_h) = y^*(\mathbf{s}_h); h \notin \mathbf{j}_{m-t}\})]\} \\
&\qquad + (-1)^{m-1} [Q(y(\mathbf{s}_i), \{y^*(\mathbf{s}_h) : h \neq i\}) - Q(\mathbf{y}^*)].
\end{aligned}
\tag{8}
$$

The relation (6) may then be applied to each expression contained in square brackets in (8), resulting in (7). ∎

Notice from this lemma that, if any of the arguments in a particular term $H_{i, j, ..., h}(y(\mathbf{s}_i), y(\mathbf{s}_j), ..., y(\mathbf{s}_h))$ are equal to the corresponding components of $\mathbf{y}^*$, then the term is equal to zero. The equations of Lemma 1, relating conditional density functions of the H-functions defined in (3) and used in the expansion (2), hold the key for modeling of the joint density function from knowledge of the conditionals alone. Clearly, if a valid joint exists and the conditionals of Lemma 1 are all positive, which is ensured by the MRF support condition, then the $\{H_{i,j,\ldots,h}\}$ are uniquely defined, once the value chosen for $\mathbf{y}^*$ in (1) is fixed.

Often, it will be useful to express (8) in product form. For example, H-functions of order 2 can be conveniently written as

$$H_{i,j}(y(\mathbf{s}_i), y(\mathbf{s}_j)) = \log\left[\frac{g_i(y(\mathbf{s}_i) \mid y(\mathbf{s}_j), \{y^*(\mathbf{s}_k) : k \neq i, j\})}{g_i(y^*(\mathbf{s}_i) \mid y(\mathbf{s}_j), \{y^*(\mathbf{s}_k) : k \neq i, j\})} \frac{g_i(y^*(\mathbf{s}_i) \mid \{y^*(\mathbf{s}_k) : k \neq i\})}{g_i(y(\mathbf{s}_i) \mid \{y^*(\mathbf{s}_k) : k \neq i\})}\right]. \tag{9}$$

The second result we shall use in the proof of the Hammersley–Clifford theorem is purely technical and is an easy consequence of the properties of conditional density functions. In this result, f will be used as generic notation for a density function, so that $f(x_i \mid x_j, x_k)$ denotes the conditional density function for X_i given x_j and x_k, while $f(x_i, x_j)$ denotes the marginal density function for X_i and X_j. Then we have the following lemma.

Lemma 2. *For a set of n random variables $\{X_i : i = 1, \ldots, n\}$ and any distinct $i, j, k \in \{1, \ldots, n\}$, define $\mathbf{w} \equiv \mathbf{w}(i, j, k) \equiv \{x_h : h \neq i, j, k\}$. Assume that $f(x_i \mid x_j, x_k, \mathbf{w})$ depends on both x_j and x_k but that $f(x_j \mid x_i, x_k, \mathbf{w})$ does not depend on x_k (which implies that $f(x_k \mid x_i, x_j, \mathbf{w})$ does not depend on x_j). Then*

$$f(x_i \mid x_j, x_k, \mathbf{w}) = f(x_i \mid x_j, \mathbf{w})\, f(x_i \mid x_k, \mathbf{w}) \frac{f(x_j, \mathbf{w})\, f(x_k, \mathbf{w})}{f(x_i, \mathbf{w})\, f(x_j, x_k, \mathbf{w})}, \tag{10}$$

where it is assumed that all densities are positive.

Proof. By assumption,

$$f(x_j \mid x_i, \mathbf{w})\, f(x_k \mid x_i, \mathbf{w}) = f(x_j \mid x_i, x_k, \mathbf{w})\, f(x_k \mid x_i, \mathbf{w}) = \frac{f(x_i, x_j, x_k, \mathbf{w})}{f(x_i, \mathbf{w})}.$$

Also,

$$f(x_j \mid x_i, \mathbf{w})\, f(x_k \mid x_i, \mathbf{w}) = \frac{f(x_i, x_j, \mathbf{w})\, f(x_i, x_k, \mathbf{w})}{\{f(x_i, \mathbf{w})\}^2}.$$

Then,

$$f(x_i, x_j, x_k, \mathbf{w}) = \frac{f(x_i, x_j, \mathbf{w})\, f(x_i, x_k, \mathbf{w})}{f(x_i, \mathbf{w})}$$

and

$$f(x_i \mid x_j, x_k, \mathbf{w}) = \frac{f(x_i, x_j, \mathbf{w})\, f(x_i, x_k, \mathbf{w})}{f(x_i, \mathbf{w})\, (x_j, x_k, \mathbf{w})},$$

from which the result follows immediately. ▮

The well-known theorem of Hammersley and Clifford depends on the concept of a "clique," defined as a single location or any set of locations such that each location is contained in the dependence set of every other location in the set. In the language of graphs, a subset of vertices (which, recall, correspond to locations in a MRF) form a clique if every pair of vertices in the subset are an edge of the graph (e.g., Speed, 1978). Given the notation of this paper, the result may be stated as follows:

THEOREM 2 (Hammersley and Clifford, 1971). *Consider random variables* $\mathbf{Y} = \{Y(\mathbf{s}_i) : i = 1, ..., n\}$, *with joint density* g_D *and associated dependence index sets* $\{D_i : i = 1, ..., n\}$. *Assume the MRF support condition* (4). *Then any function* $H_{i, j, ..., h}$ *given in Lemma* 1 *is equal to zero unless the locations* $\{\mathbf{s}_i, \mathbf{s}_j, ..., \mathbf{s}_h\}$ *form a clique.*

Proof. Consider any two locations $\mathbf{s}_i$ and $\mathbf{s}_j$ such that $j \notin D_i$. Then $g_i(y(\mathbf{s}_i) \mid \{y(\mathbf{s}_h) : h \neq i\})$ does not depend on $y(\mathbf{s}_j)$, and thus $H_{i,j}(\cdot) \equiv 0$ from Eq. (7) of Lemma 1. Similarly, for any three locations $\mathbf{s}_i$, $\mathbf{s}_j$, and $\mathbf{s}_k$, $H_{i,j,k}(\cdot) \equiv 0$, unless $j \in D_i$ and $k \in D_i$. If $\{\mathbf{s}_i, \mathbf{s}_j, \mathbf{s}_k\}$ do not form a clique, this leaves only the possibility that both $j \in D_i$ and $k \in D_i$ but $j \notin D_k$ (and hence also $k \notin D_j$). Application of Lemma 2 in Eq. (7) shows that $H_{i,j,k}(\cdot) \equiv 0$ for this case also. The same reasoning extends in a straightforward manner to higher-order interaction terms, completing the proof. ▮

4. CONSTRUCTING CONDITIONALLY SPECIFIED MODELS

To this point in our development, none of the results presented explicitly addresses the issue of model formulation using conditional distributions. Definition of the negpotential function (1) assumes the existence of a joint density function. Lemma 1 may be used to identify the H-functions of expansion (2) in any situation for which a joint exists and the MRF support condition holds; the conditional density functions that appear in Lemma 1 are those dictated by an assumed joint density. The version of the Hammersley–Clifford theorem stated here also requires the existence of a joint density. Thus, the modeling issue becomes how and when one can, given a set of conditional density functions and the corresponding dependence index sets, identify a valid joint density, that is, a joint density that possesses the specified conditionals. When this is the case, the resulting model is known as a MRF model.

In construction of a MRF model, we begin with a set of specified conditionals and attempt to identify a valid joint by applying the relations of Sections 2 and 3 in what is essentially the reverse order of their development. While the relations of Lemma 1 were developed from a joint density (and the MRF support condition), in MRF model construction we use those relations to define a set of H-functions from specified conditionals. We shall call such functions, H^c-functions, where the superscript c emphasizes that they are *constructed* from prescribed conditionals using (7). These H^c-functions are then used in the right-hand side of (2) to arrive at a *constructed* negpotential function Q^c. When can we be assured that this constructed negpotential function specifies, up to a normalizing constant, a joint distribution that possesses the specified conditionals? The following result, which strengthens and extends a proposition of Cressie and Lele (1992), answers this question.

THEOREM 3. *Assume that a set of univariate conditional probability density functions* $\{f_i(y(\mathbf{s}_i) \mid \{y(\mathbf{s}_h): h \neq i\}): i = 1, ..., n\}$, *and a value* $\mathbf{y}^*$ *for which the MRF support condition* (4) *is satisfied, have been specified. Then a joint distribution having those conditionals exists, and its density function* $f(\mathbf{y})$ *may be specified up to a normalizing constant through the application of formula* (7) *given in Lemma* 1 *and the expansion* (2), *if and only if the following conditions hold*:

(i) *The* H^c-*functions constructed from the specified conditionals using* (7) *are each invariant under permutation of their associated indices. That is, for* $1 \leqslant m \leqslant n$, *and any permutation* $\sigma(j(1), ..., j(m))$ *of indices* $j(1), ..., j(m)$,

$$H^c_{j(1), ..., j(m)}(\cdot) = H^c_{\sigma(j(1), ..., j(m))}(\cdot),$$

where the arguments of the right-hand side are similarly permuted.

(ii) *The negpotential function* Q^c, *constructed from the* H^c-*functions using* (2) *and defined on support* $\Omega = \{\mathbf{y}: \exp\{Q^c(\mathbf{y})\} > 0\}$, *satisfies*

$$\int_\Omega \exp\{Q^c(\mathbf{t})\}\, d\mu(\mathbf{t}) < \infty,$$

for the appropriate measure μ (*e.g., counting or Lebesgue*).

Proof. The proof consists of demonstrating the necessity of (i) and (ii) individually, and then the sufficiency of these two conditions together. First, suppose that a joint density function $f(\mathbf{y})$; $\mathbf{y} \in \Omega$, and $f(\mathbf{y}) \equiv 0$; $\mathbf{y} \notin \Omega$

corresponding to the set of specified conditional density functions, exists. All conditional densities used to construct H^c-functions from formula (7) of Lemma 1 are positive by the MRF support condition (4). Let $v \equiv (\mathbf{j}_m^{-i}, \{h : h \notin \mathbf{j}_m\})$. Then each term on the inside of the summation in Eq. (7) may be written as

$$\log\left[\frac{f_{i,v}(y(\mathbf{s}_i), \{y(\mathbf{s}_k) : k \in \mathbf{j}_m^{-i}\}, \{y^*(\mathbf{s}_h) : h \notin \mathbf{j}_m\})}{f_{i,v}(y^*(\mathbf{s}_i), \{y(\mathbf{s}_k) : k \in \mathbf{j}_m^{-i}\}, \{y^*(\mathbf{s}_h) : h \notin \mathbf{j}_m\})}\right],$$

and the final term of (7) may be written as

$$\log\left[\frac{f_{i,v}(y(\mathbf{s}_i), \{y^*(\mathbf{s}_j) : j \neq i\})}{f_{i,v}(y^*(\mathbf{s}_i), \{y^*(\mathbf{s}_j) : j \neq i\})}\right].$$

Since any joint density is invariant to permutation of the indices used to label values in the support, so too are these terms and each H^c-function that might be constructed from the set of specified conditional densities using Eq. (7). Condition (i) is thus necessary for the existence of a joint density.

For the negpotential function defined in (1) from the joint density $g_D(\mathbf{y})$,

$$\int_\Omega \exp\{Q(\mathbf{t})\}\, d\mu(\mathbf{t}) = \frac{1}{g_D(\mathbf{y}^*)}.$$

Since the definiton assumes that $\mathbf{y}^* \in \Omega$, this integral must be finite. Thus, the joint density $f(\mathbf{y})$, defined to be proportional to $\exp\{Q^c(\mathbf{y})\}$, must satisfy the same relation, and so condition (ii) is verified as necessary.

It remains to demonstrate that, in combination, conditions (i) and (ii) are sufficient for the result. If, in addition to the MRF support condition (4), condition (i) holds, then Q^c, constructed from the set of functions $\{H^c_{i,j,\ldots,h}\}$ using (2), is such that

$$Q^c(y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)) = Q^c(\sigma(y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n))),$$

for any permutation $\sigma(y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n))$. If $\exp\{Q^c(\cdot)\}$ is integrable by condition (ii), a joint density function may be constructed as

$$f(\mathbf{y}) \equiv \frac{\exp\{Q^c(\mathbf{y})\}}{\int_\Omega \exp\{Q^c(\mathbf{t})\}\, d\mu(\mathbf{t})}; \qquad \mathbf{y} \in \Omega.$$

Clearly, $f(\mathbf{y})$ integrates to one, and invariance of Q^c to permutation of location labels gives the same property to $f(\mathbf{y})$. Thus, $f(\mathbf{y})$; $\mathbf{y} \in \Omega$, is a joint density for $\mathbf{Y}$ that, upon applying the results of Sections 2 and 3, implies the specified conditional densities. ▮

The importance of Theorem 3 is that it indicates a method for construction of a joint distribution that possesses a set of specified conditional densities that define a MRF. Several observations about this theorem will help clarify some rather subtle issues that exist in the MRF approach to model specification.

1. The permutation invariance condition of (i) is the primary tool available to verify the validity of a conditionally specified model. Lemma 1 provides a way to construct H^c-functions from conditional densities, which are then used to construct Q^c from the relation (2). If the H^c-functions satisfy (i), then, subject to the integrability condition of (ii), a valid model has been formulated. Condition (i) refers only to the *labels* used to identify random variables, and should not be confused with exchangeability. The condition assumes great importance because the expansion (2), which is used to construct Q^c, contains only sums over ordered indices. We must have assurance that the particular ordering chosen is immaterial to specification of the joint density, and the ensuing likelihood. To clarify this concept, consider the simple case of two random variables $Y(\mathbf{s}_i)$ and $Y(\mathbf{s}_j)$. If these variables are ordered as (i, j) the negpotential constructed will be, in reduced notation, $Q^c_{i,j} = H_i + H_j + H_{i,j}$. If, however, the variables are ordered as (j, i) the negpotential constructed will be $Q^c_{j,i} = H_i + H_j + H_{j,i}$. Assuming that both of these functions are integrable (condition (ii)), the resulting joint specifications will be the same if and only if $H_{i,j} = H_{j,i}$. The theorem indicates that this same requirment extends to the general case for collections of n random variables.

2. The MRF support condition (4) is implied by, and hence is a weaker condition than, the frequently assumed positivity condition, although its verification requires knowledge of Ω and $\{\Phi_i : i = 1, ..., n\}$. In practice, it is often possible to obtain these from knowledge of constraints on the random variables of the underlying process, followed by a validity check from the constructed joint density. A simple example will be given in Section 5.1 in which the positivity condition is not met, but Theorem 3 may be applied to identify a joint distribution. At the same time, the MRF support condition is stronger than the "incidence set" condition of Arnold and Press (1989), which is necessary for existence of a valid joint distribution. Thus, while the MRF support condition is precisely the appropriate condition for our construction of MRF models (see Theorems 1 and 3), we do not claim that our construction is the most general. A simple example in Section 5.2 will illustrate a situation in which the MRF support condition (4) is not met, so that Theorem 3 may not be used to identify a joint distribution, but the incidence set condition is met and a valid joint does exist.

3. Although not explicitly mentioned in Theorem 3, the Hammersley–Clifford Theorem (Theorem 2) is important in the MRF approach to

model formulation because it indicates precisely how many H^c-functions are to be used in construction of Q^c. In complex settings, such as image-analysis applications involving hundreds or even thousands of random variables, identification of clique structure and the Hammersley–Clifford theorem make the task of constructing Q^c from the expansion (2) a manageable one.

4. Condition (ii) merely states that $\exp\{Q^c(\cdot)\}$ is integrable on Ω. In practice, evaluation of this integral is nearly always problematic. Theorem 3 provides a method for specification of a joint density only up to a normalizing constant, which generally involves the values of unknown parameters. Thus, the development of inferential methods, such as Markov Chain Monte Carlo methods (e.g., Geyer and Thompson 1992), for distributions known only up to a constant has been of great importance. This is an area of intense current research, although it is not the focus of this article.

5. ILLUSTRATIONS OF MODEL CONSTRUCTION

We present here three examples that illustrate both the use and limitations of the MRF approach. The first two examples are rather specialized, focusing on the role of the MRF support condition (4). These examples make use of a discrete, bivariate setting. The third example is more realistic, illustrating the potential of the MRF approach for formulation of nonstandard statistical models.

5.1. *Theorem* 3 *Is Applicable*

Suppose we have two random variables,

$$Y(\mathbf{s}_1) \in \Omega_1 = \{0, 1\}; \qquad Y(\mathbf{s}_2) \in \Omega_2 = \{1, 2, 3\},$$

and consider the conditional specifications,

$$\Pr[\,Y(\mathbf{s}_1) = 0 \mid Y(\mathbf{s}_2) = 1\,] = 1.00 \qquad \Pr[\,Y(\mathbf{s}_2) = 1 \mid Y(\mathbf{s}_1) = 0\,] = 0.70$$

$$\Pr[\,Y(\mathbf{s}_1 = 0 \mid Y(\mathbf{s}_2) = 2\,] = 0.25 \qquad \Pr[\,Y(\mathbf{s}_2) = 2 \mid Y(\mathbf{s}_1) = 0\,] = 0.20$$

$$\Pr[\,Y(\mathbf{s}_1 = 0 \mid Y(\mathbf{s}_2) = 3\,] = 0.20 \qquad \Pr[\,Y(\mathbf{s}_2) = 1 \mid Y(\mathbf{s}_1) = 1\,] = 0.00$$

$$\Pr[\,Y(\mathbf{s}_2) = 2 \mid Y(\mathbf{s}_1) = 1\,] = 0.60,$$

with all other conditional probabilities obtained by appropriate addition and then subtraction from 1. We now attempt to identify the joint distribution of $Y(\mathbf{s}_1)$ and $Y(\mathbf{s}_2)$ by means of Theorem 3. First, let $y^*(\mathbf{s}_1) = 0$ and

$y^*(\mathbf{s}_2)=2$. From the conditional specifications, we have that $\Omega=\{(0, 1), (0, 2), (0, 3), (1, 2), (1, 3)\}$, and the MRF support condition (4) is easily verified. On the other hand, the positivity condition is not met, since $(1, 1) \notin \Omega$. Using Lemma 1 to construct H^c-functions results in

$$
\begin{aligned}
H_1^c(0) &= 0, & H_{1,2}^c(0, 1) &= 0,\\
H_1^c(1) &= \log(3.0), & H_{1,2}^c(0, 2) &= 0,\\
H_2^c(1) &= \log(3.5), & H_{1,2}^c(0, 3) &= 0,\\
H_2^c(2) &= 0, & H_{1,2}^c(1, 2) &= 0,\\
H_2^c(3) &= \log(0.50), & H_{1,2}^c(1, 3) &= \log(4.0/3.0),
\end{aligned}
$$

and it is readily verified that $H_{1,2}^c(y(\mathbf{s}_1), y(\mathbf{s}_2)) = H_{2,1}^c(y(\mathbf{s}_2), y(\mathbf{s}_1))$ for all $(y(\mathbf{s}_1), y(\mathbf{s}_2))^T \in \Omega$. Construction of the negpotential function using expansion (2) then yields

$$
\begin{aligned}
Q^c(0, 1) &= H_1^c(0) + H_2^c(1) + H_{1,2}^c(0, 1) = \log(3.5),\\
Q^c(0, 2) &= H_1^c(0) + H_2^c(2) + H_{1,2}^c(0, 2) = 0.0,\\
Q^c(0, 3) &= H_1^c(0) + H_2^c(3) + H_{1,2}^c(0, 3) = \log(0.5),\\
Q^c(1, 2) &= H_1^c(1) + H_2^c(2) + H_{1,2}^c(1, 2) = \log(3.0),\\
Q^c(1, 3) &= H_1^c(1) + H_2^c(3) + H_{1,2}^c(1, 3) = \log(3.0) + \log(0.5) + \log(4.0/3.0),
\end{aligned}
$$

so that $\sum_{\Omega} \exp\{Q^c(y(\mathbf{s}_1), y(\mathbf{s}_2))\} = 10.0$, and the constructed joint probabilities are

$$
\begin{aligned}
\Pr[Y(\mathbf{s}_1)=0,\ Y(\mathbf{s}_2)=1] &= 0.35,\\
\Pr[Y(\mathbf{s}_1)=0,\ Y(\mathbf{s}_2)=2] &= 0.10,\\
\Pr[Y(\mathbf{s}_1)=0,\ Y(\mathbf{s}_2)=3] &= 0.05,\\
\Pr[Y(\mathbf{s}_1)=1,\ Y(\mathbf{s}_2)=2] &= 0.30,\\
\Pr[Y(\mathbf{s}_1)=1,\ Y(\mathbf{s}_2)=3] &= 0.20.
\end{aligned}
$$

It may be easily verified that this joint distribution possesses the conditional probabilities given at the start of this subsection.

5.2. *Theorem* 3 *Is Not Applicable*

We now present a simple example in which we are not able to apply Theorem 3. For the same two random variables and support sets of

Section 5.1, namely $\Omega_1 = \{0, 1\}$ and $\Omega_2 = \{1, 2, 3\}$, consider the conditional specifications

$$\Pr[Y(\mathbf{s}_1)=0 \mid Y(\mathbf{s}_2)=1]=1.00 \qquad \Pr[Y(\mathbf{s}_2)=1 \mid Y(\mathbf{s}_1)=0]=0.70$$

$$\Pr[Y(\mathbf{s}_1)=0 \mid Y(\mathbf{s}_2)=2]=0.43 \qquad \Pr[Y(\mathbf{s}_2)=2 \mid Y(\mathbf{s}_1)=0]=0.30$$

$$\Pr[Y(\mathbf{s}_1)=0 \mid Y(\mathbf{s}_2)=3]=0.00 \qquad \Pr[Y(\mathbf{s}_2)=1 \mid Y(\mathbf{s}_1)=1]=0.00$$

$$\Pr[Y(\mathbf{s}_2)=2 \mid Y(\mathbf{s}_1)=1]=0.40.$$

For these conditional specifications, $\Omega = \{(0, 1), (0, 2), (1, 2), (1, 3)\}$ and there is no value $y^*(\mathbf{s}_1) \in \Omega_1$ that satisfies the MRF support condition (4), although either $y^*(\mathbf{s}_2)=1$ or $y^*(\mathbf{s}_2)=2$ could be used. If one attempts to apply Theorem 3 with $y^*(\mathbf{s}_1)=0$ and $y^*(\mathbf{s}_2)=2$ in this situation, the resulting H^c-functions become

$$H_1^c(0)=0, \qquad H_{1,2}^c(0,1)=0,$$

$$H_1^c(1)=\log(0.57/0.43), \qquad H_{1,2}^c(0,2)=0,$$

$$H_2^c(1)=\log(0.70/0.30), \qquad H_{1,2}^c(1,2)=0,$$

$$H_2^c(2)=0, \qquad H_{1,2}^c(1,3)=\infty,$$

$$H_2^c(3)=-\infty.$$

The non-existence of $H_2^c(3)$ and $H_{1,2}^c(1,3)$ indicates the failure of Theorem 3 in this case. One might be tempted to define the problem away, by taking $\infty - \infty = 0$ in construction of Q^c from the form of expansion (2). Doing so results in a joint distribution, but it is not the *correct* joint distribution, as may be seen from simple calculations.

This does not mean that a joint distribution with the specified conditional probabilities of this example fails to exist. While MRF support condition (4) is not met, the incidence set condition of Arnold and Press (1989) does hold. This condition states that if $N_1 = \{(y(\mathbf{s}_1), y(\mathbf{s}_2)) : f(y(\mathbf{s}_1) \mid y(\mathbf{s}_2)) > 0\}$ and $N_2 = \{(y(\mathbf{s}_1), y(\mathbf{s}_2)) : f(y(\mathbf{s}_2) \mid y(\mathbf{s}_1)) > 0\}$, then $N_1 = N_2$. Here, $N_1 = N_2 = \{(0, 1), (0, 2), (1, 2), (1, 3)\}$, and the condition holds. In addition, the compatibility condition of Arnold and Press (1989) is easily verified, so that a joint distribution with the given conditionals does exist. In this simple case, it may be seen that such a joint is,

$$\Pr[Y(\mathbf{s}_1)=0, \quad Y(\mathbf{s}_2)=1]=0.35,$$

$$\Pr[Y(\mathbf{s}_1)=0, \quad Y(\mathbf{s}_2)=2]=0.15,$$

$$\Pr[Y(\mathbf{s}_1)=1, \quad Y(\mathbf{s}_2)=2]=0.20,$$

$$\Pr[Y(\mathbf{s}_1)=1, \quad Y(\mathbf{s}_2)=3]=0.30.$$

As mentioned in Section 2, it is also an easy exercise to verify that the condition of Lemma 0.1 in Besag (1994) is met for this example, so that the correct joint could be simulated using the Gibbs sampler.

5.3. *A Beta Conditionals Model*

Consider a situation involving dependent random variables that assume values on the interval (0, 1). Such situations might arise, for example, in which each of a number of materials are graded by one or more individuals, such as wine or coffee tasting. While one might assume that individuals are independent, the same assumption would be less reasonable for gradings of different materials by the same individual. Another hypothetical setting might involve the examination of the proportion of individuals from a set of small adjacent geographic areas with roughly equal populations (e.g., neighborhood blocks) that vote for a candidate or a proposition in a local election. Other appropriate situations are not difficult to visualize. In such cases, we may wish to model observed proportions directly, and a beta distribution is then an obvious choice. Specification of a model in which the full set of conditional distributions are beta may be developed as follows.

Define $\mathbf{Y} \equiv \{ Y(\mathbf{s}_i) : i = 1, ..., n \}$ for appropriately chosen "locations" $\mathbf{s}_i$; $i = 1, ..., n$, and assume that a neighborhood structure has been defined through a set of dependence sets $\{ N_i : i = 1, ..., n \}$ and the associated dependence index sets $\{ D_i : i = 1, ..., n \}$. Define $\mathbf{Y}(N_i) \equiv \{ Y(\mathbf{s}_j) : \mathbf{s}_j \in N_i \}$. Beta conditional probability density functions may be assigned to each element of $\mathbf{Y}$ through the exponential family structure,

$$f_i(y(\mathbf{s}_i) \mid \mathbf{y}(N_i)) = \exp \left\{ \sum_{k=1}^{2} A_{i,k}(\mathbf{y}(N_i))\, T_k(y(\mathbf{s}_i)) - B_i(\mathbf{y}(N_i)) + C_i(y(\mathbf{s}_i)) \right\}. \tag{11}$$

Here, the $A_{i,k}(\cdot)$ are natural parameter functions and the T_k are sufficient statistics. For (11) to define beta density functions requires that $-1 < A_{i,1}(\cdot) < \infty$, $-1 < A_{i,2}(\cdot) < \infty$, $T_1(y(\mathbf{s}_i)) = \log\{ y(\mathbf{s}_i) \}$, $T_2(y(\mathbf{s}_i)) = \log\{ 1 - y(\mathbf{s}_i) \}$, $B_i(\cdot) = \log\{ \Gamma(A_{i,1}(\cdot) + 1) \} + \log\{ \Gamma(A_{i,2}(\cdot) + 1) \} - \log\{ \Gamma(A_{i,1}(\cdot)\, A_{i,2}(\cdot) + 2) \}$, and $C_i(y(\mathbf{s}_i)) = 0$.

Assume the positivity condition, so that Ω is the n-fold cartesian product of (0, 1). Also assume the pairwise-only dependence assumption of Besag (1974), so that all cliques contain only individual random variables or pairs of random variables. Then, from (7) of Lemma 1, the general form of H^c-functions needed are

$$H_i^c(y(\mathbf{s}_i)) = \log \left[\frac{f_i(y(\mathbf{s}_i) \mid \{ y^*(\mathbf{s}_j) : j \neq i \})}{f_i(y^*(\mathbf{s}_i) \mid \{ y^*(\mathbf{s}_j) : j \neq i \})} \right] \tag{12}$$

and

$$H^c_{i,j}(y(\mathbf{s}_i), y(\mathbf{s}_j))$$
$$= \log\left[\frac{f_i(y(\mathbf{s}_i) \mid y(\mathbf{s}_j), \{y^*(\mathbf{s}_k): k \neq i, j\})}{f_i(y^*(\mathbf{s}_i) \mid y(\mathbf{s}_j), \{y^*(\mathbf{s}_k): k \neq i, j\})} \frac{f_i(y^*(\mathbf{s}_i) \mid \{y^*(\mathbf{s}_k): k \neq i\})}{f_i(y(\mathbf{s}_i) \mid \{y^*(\mathbf{s}_k): k \neq i\})}\right]. \tag{13}$$

Substitution of (11) into (12) and (13) yields

$$H^c_i(y(\mathbf{s}_i)) = \sum_{k=1}^{2} [A_{i,k}(\mathbf{y}^*(N_i))\{T_k(y(\mathbf{s}_i)) - T_k(y^*(\mathbf{s}_i))\}] + C_i(y(\mathbf{s}_i)) - C_i(y^*(\mathbf{s}_i)) \tag{14}$$

and

$$H^c_{i,j}(y(\mathbf{s}_i), y(\mathbf{s}_j))$$
$$= \sum_{k=1}^{2} [\{A_{i,k}(y(\mathbf{s}_j), \mathbf{y}^*(N_i \backslash \mathbf{s}_j)) - A_{i,k}(\mathbf{y}^*(N_i))\}\{T_k(y(\mathbf{s}_i)) - T_k(y^*(\mathbf{s}_i))\}]. \tag{15}$$

Model formulation is completed by specifying a form for the functions $\{A_{i,1}(\cdot), A_{i,2}(\cdot): i = 1, ..., n\}$ such that the conditions of Theorem 3 are satisfied. One such specification for beta conditionals is

$$\begin{aligned} A_{i,1}(\mathbf{y}(N_i)) &= \alpha_{i,1} - \sum_{j \in D_i} \eta_{i,j} \log\{1 - y(\mathbf{s}_j)\} \\ A_{i,2}(\mathbf{y}(N_i)) &= \alpha_{i,2} - \sum_{j \in D_i} \eta_{i,j} \log\{y(\mathbf{s}_j)\}. \end{aligned} \tag{16}$$

Here, to ensure that $-1 < A_{i,k}(\cdot) < \infty$; $k = 1, 2$, we require $\alpha_{i,k} > -1$ and $\eta_{i,j} \geqslant 0$ for $k = 1, 2$ and $i, j = 1, ..., n$. Substitution of (16) into (15) yields

$$\begin{aligned} H^c_{i,j}(y(\mathbf{s}_i), y(\mathbf{s}_j)) = -\eta_{i,j}(&[\log\{y(\mathbf{s}_j)\} - \log\{y^*(\mathbf{s}_j)\}] \\ &\times [\log\{1 - y(\mathbf{s}_i)\} - \log\{1 - y^*(\mathbf{s}_i)\}]) \\ &-\eta_{i,j}([\log\{1 - y(\mathbf{s}_j)\} - \log\{1 - y^*(\mathbf{s}_j)\}] \\ &\times [\log\{y(\mathbf{s}_i)\} - \log\{y^*(\mathbf{s}_i)\}]), \end{aligned}$$

which, if $\eta_{i,j} = \eta_{j,i}$, is symmetric in i and j, verifying condition (i) of Theorem 3. The negpotential function constructed from these H^c-functions is, up to an additive constant,

$$
\begin{aligned}
Q^c(\mathbf{y}) = \sum_{i=1}^{n} [\alpha_{i,1} \log\{y(\mathbf{s}_i)\} + \alpha_{i,2} \log\{1 - y(\mathbf{s}_i)\}] \\
- \sum\sum_{1 \leqslant i < j \leqslant n} \eta_{i,j}[\log\{y(\mathbf{s}_i)\} \log\{1 - y(\mathbf{s}_j)\} \\
+ \log\{1 - y(\mathbf{s}_i)\} \log\{y(\mathbf{s}_j)\}],
\end{aligned} \tag{17}
$$

where $\eta_{i,j} \equiv 0$ if $j \notin D_i$. That Q^c is integrable, under the restrictions $\alpha_{i,k} > -1$ and $\eta_{i,j} \geqslant 0$, follows from the general results of Arnold *et al.* (1992).

This model results in positive dependence among the components of $\mathbf{Y}$. To see this, let $\kappa_{i,1} = -\sum_{j \in D_i} \eta_{i,j} \log(1 - y(\mathbf{s}_j))$ and $\kappa_{i,2} = -\sum_{j \in D_i} \eta_{i,j} \log(y(\mathbf{s}_j))$. The conditional expectation of $Y(\mathbf{s}_i)$ may then be written as

$$
E_i \equiv E(Y(\mathbf{s}_i) \mid \mathbf{y}(D_i)) = \frac{\alpha_{i,1} + \kappa_{i,1} + 1}{\alpha_{i,1} + \kappa_{i,1} + \alpha_{i,2} + \kappa_{i,2} + 2}.
$$

Consider $\kappa_{i,1}$ and $\kappa_{i,2}$ as functions of only one component of $\mathbf{y}(N_i)$, say $y(\mathbf{s}_v)$. Then $\kappa_{i,1} \geqslant 0$ is increasing in $y(\mathbf{s}_v)$ and $\kappa_{i,2} \geqslant 0$ is decreasing in $y(\mathbf{s}_v)$, while E_i is increasing in $\kappa_{i,1}$ and decreasing in $\kappa_{i,2}$. Thus, E_i is monotone increasing in the value $y(\mathbf{s}_v)$, and dependence of $Y(\mathbf{s_i})$ on $Y(\mathbf{s_v})$, in terms of regression, is positive. Since this is true for any component of $\mathbf{y}(N_i)$, and we have assumed pairwise-only dependence, the parameterization (16) may only be used to model situations involving positive dependence.

An example data set was simulated from model (16) and is presented in Fig. 1. The data graphed in Fig. 1 were generated in a bivariate setting with $\mathbf{s}_i \equiv (u, v)$, where u indexes independent pairs, $u = 1, ..., 50$, and v indexes variables within pairs, $v = 1, 2$. Dependence index sets were defined as $D_i \equiv \{j: \mathbf{s}_i - \mathbf{s}_j = (0, \pm 1)\}$ so that each random variable $Y(\mathbf{s}_i)$ has a dependence set consisting of the other variable in its pair. Parameter values used to generate data were $\alpha_{i,1} = 2$, $\alpha_{i,2} = 1$ and $\eta_{i,j} = 4(I_{j \in D_i})$ (with I_A the indicator function of an event A). These data resulted in a sample mean of 0.66 and sample variance of 0.038 for observed values $\{y(1, 1), y(2, 1), ..., y(50, 1)\}$ and a sample mean of 0.64 and sample variance of 0.034 for observed values $\{y(1, 2), y(2, 2), ..., y(50, 2)\}$. The sample correlation was $r = 0.66$. For comparison, an independence model would have generated data from distributions with expectations of 0.60 and variances of 0.040 for all variables.

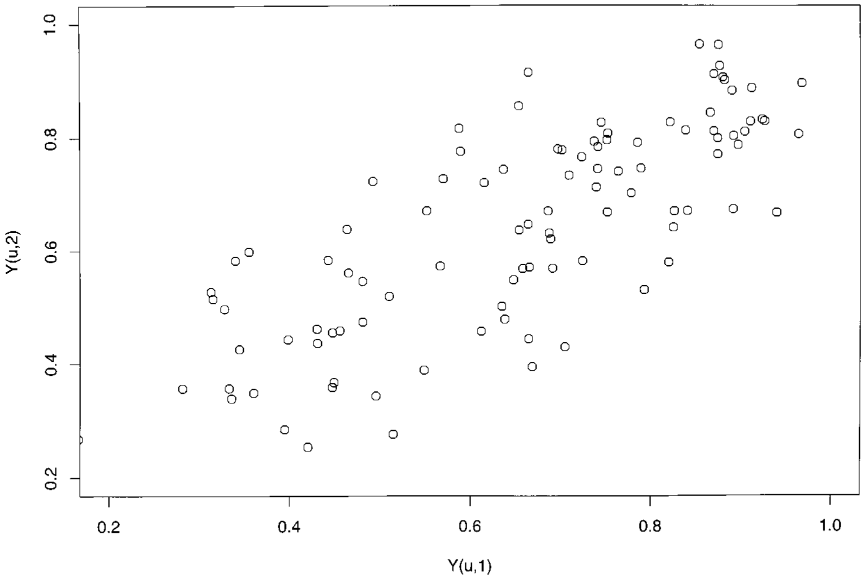

FIG. 1. Scatterplot of 50 observations simulated from a bivariate model with beta conditional distributions.

6. CONCLUSION

The purpose of this article has been to present the basic theory needed for the construction of MRF models from the full set of conditional distributions. We have redefined the negpotential function, weakened the support condition required for existence and construction of the joint distribution, reproved the Hammersley–Clifford theorem using the new negpotential function and support condition, and given necessary and sufficient conditions for which our construction formula yields a valid joint distribution.

The simple examples presented in Sections 5.1 and 5.2, while mainly of pedagogical interest, indicate that the conditions required for construction of a joint distribution from the conditional distributions are not mere technicalities. But, given that these conditions are satisfied, Theorem 3 provides a clear recipe by which a multivariate model can be constructed. The MRF approach to statistical model formulation leads to new classes of multivariate models, such as the beta-conditionals model given in Section 5.3.

Building a joint distribution through specification of the conditional distributions, called here the MRF approach to multivariate modeling, is appealing because dependencies are specified site-by-site through conditional quantities (e.g., conditional mean, conditional variance) that have an intuitive interpretation. There remain unresolved statistical issues in the use of multivariate models obtained from the MRF approach, particularly parameterization issues and inferential issues. Nonetheless, using the theory of this paper, we are able to formulate useful statistical models for the analysis of data that exhibit non-Gaussianity, nonlinearity, and dependence.

ACKNOWLEDGMENTS

This research was supported by the U.S. Environmental Protection Agency under Assistance Agreement CR822919-01-0, the Office of Naval Research (Cressie) under Grant N00014-93-1-0001, and the National Science Foundation under grant DMS 9816630 (Kaiser). The authors thank the editors and referees for their constructive comments on the original submission.

REFERENCES

B. C. Arnold and S. J. Press, Compatible conditional distributions, *J. Am. Statist. Assoc.* **84** (1989), 152–156.

B. C. Arnold and D. Strauss, Bivariate distributions with conditionals in prescribed exponential families, *J. Roy. Statist. Soc.* **53** (1991), 365–375.

B. C. Arnold, E. Castillo, and J. Sarabia, "Conditionally Specified Distributions," Lecture Notes in Statistics, Vol. 73, Springer-Verlag, Berlin, 1992.

M. S. Bartlett, "An Introduction to Stochastic Processes," 2nd ed., Cambridge Univ. Press, Cambridge, UK, 1966.

J. E. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion), *J. Roy. Statist. Soc. B* **36** (1974), 192–236.

J. E. Besag, Discussion of Tierney L. (1994), Markov chains for exploring posterior distributions, *Ann. Statist.* **22** (1994), 1734–1741.

D. Brook, On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neihbor systems, *Biometrika* **51** (1964), 481–483.

P. Clifford, Markov random fields in statistics, *in* "Disorder in Physical Systems," pp. 20–32, Oxford Univ. Press, New York, 1990.

L. Cohen, Probability distributions with given multivariate marginals, *J. Math. Phys.* **25** (1984), 2402–2403.

D. R. Cox and N. Wermuth, Linear dependencies represented by chain graphs (with discussion), *Statist. Sci.* **8** (1993), 204–283.

E. Cramer, Conditional iterative proportional fitting for Gaussian distributions, *J. Multivariate Anal.* **65** (1998), 261–276.

N. A. C. Cressie, "Statistics for Spatial Data," revised ed., Wiley, New York, 1993.

N. Cressie and S. Lele, New models for Markov random fields, *J. Appl. Probab.* **29** (1992), 877–884.

C. M. Cuadras, Probability distributions with given multivariate marginals and given dependence structure, *J. Multivariate Anal.* **42** (1992), 51–66.

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.

A. Gelman and X.-L. Meng, A note on bivariate distributions that are conditionally normal, *Am. Statist.* **45** (1991), 125–126.

C. J. Geyer and E. A. Thompson, Constrained Monte Carlo maximum likelihood for dependent data (with discussion), *J. Roy. Statist. Soc. B* **54** (1992), 657–699.

J. M. Hammersley and P. Clifford, Markov fields on finite graphs and lattices, Unpublished manuscript, Oxford University, (1971).

S. L. Lauritzen, "Graphical Models," Oxford Univ. Press, Oxford, UK, 1996.

H. Li, M. Scarsini, and M. Shaked, Linkages: A tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals, *J. Multivariate Anal.* **56** (1996), 20–41.

A. W. Marshall and I. Olkin, Families of multivariate distributions, *J. Am. Statist. Assoc.* **83** (1988), 834–841.

J. Moussouris, Gibbs and Markov random systems with constraints, *J. Statist. Phys.* **10** (1974), 11–33.

L. Rüschendorf, Construction of multivariate distributions with given marginals, *Ann. Inst. Statist. Math.* **37** (1985), 225–233.

T. P. Speed, Relations between models for spatial data, contingency tables and Markov fields on graphs, *Suppl. Adv. Appl. Probab.* **10** (1978), 111–122.

J. Whittaker, "Graphical Models in Applied Multivariate Statistics," Wiley, Chichester, 1990.

P. Whittle, Stochastic processes in several dimensions, *Bull. Int. Statist. Inst.* **40** (1963), 970–994.