IOWA STATE UNIVERSITY
Digital Repository

4-2007

Statistical Dependence in Markov Random Field Models

Mark S. Kaiser
Iowa State University, mskaiser@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/stat_las_preprints

Part of the Statistics and Probability Commons

Statistical Dependence in Markov Random Field Models

Abstract

Statistical models based on Markov random fields present a flexible means for modeling statistical dependencies in a variety of situations including, but not limited to, spatial problems with observations on a lattice. The simplest of such models, sometimes called ``auto-models'' are formulated from sets of conditional one-parameter exponential family densities or mass functions. Despite the attractive nature of these models for dealing with complex dependence structures, their application has been hindered by a lack of interpretability relative to the manner in which dependencies are represented. In particular, while the parameters that embody dependence are nicely isolated in these models, the meaning of numerical values of those parameters as representing dependence of varying strengths has been poorly understood. In addition, it is known that dependence parameters that are ``too large'' lead to un-interpretable or even degenerate behavior in data sets simulated from models having such parameters. The objectives of this article are to identify a concept of dependence that is generally applicable to Markov random field models based on one-parameter exponential families, and to demonstrate the relation between a quantification of this concept of dependence and the dependence parameters in models. It is then possible to both quantify the strength of statistical dependencies represented by particular numerical values of dependence parameters, and delineate ranges of those parameters that lead to separable interpretations of large-scale model components as marginal mean structure and small-scale components as additional statistical dependence.

Disciplines
Statistics and Probability

Statistical Dependence in Markov Random Field Models

Mark S. Kaiser

Department of Statistics

Iowa State University

April 2007

Abstract

Statistical models based on Markov random fields present a flexible means for modeling statistical dependencies in a variety of situations including, but not limited to, spatial problems with observations on a lattice. The simplest of such models, sometimes called "auto-models" are formulated from sets of conditional one-parameter exponential family densities or mass functions. Despite the attractive nature of these models for dealing with complex dependence structures, their application has been hindered by a lack of interpretability relative to the manner in which dependencies are represented. In particular, while the parameters that embody dependence are nicely isolated in these models, the meaning of numerical values of those parameters as representing dependence of varying strengths has been poorly understood. In addition, it is known that dependence parameters that are "too large" lead to uninterpretable or even degenerate behavior in data sets simulated from models having such parameters. The objectives of this article are to identify a concept of dependence that is generally applicable to Markov random field models based on one-parameter exponential families, and to demonstrate the relation between a quantification of this concept of dependence and the dependence parameters in models. It is then possible to both quantify the strength of statistical dependencies represented by particular numerical values of dependence parameters, and delineate ranges of those parameters that lead to separable interpretations of large-scale model components as marginal mean structure and small-scale components as additional statistical dependence.

KEYWORDS: spatial models, auto-logistic, Winsorized Poisson, directional dependence, standard bounds, conditionally specified models

1   Introduction

Markov random field models may be used to represent interactions among random variables that correspond to a finite set of locations in a general random field structure. Formulation of Markov random field models is accomplished through specification of a full conditional distribution for the variable at each location, given values for all of the other locations contained in some domain, often a spatial region. A Markov assumption is made that the full conditional distribution for a location depends functionally only on locations designated as its neighbors, and this is true for each of the entire set of locations. Although other forms are available for the conditional distributions (e.g., Arnold, Castillo and Sarabia, 1992; Kaiser and Cressie, 2000) the most commonly applied Markov random field models are based on one parameter exponential families, introduced as "auto-models" by Besag (1974). These include models based on binary conditionals, binomial conditionals, Poisson conditionals, and normal conditionals (considering variance as known or a nuisance parameter).

A puzzling aspect of Markov random field models is that, although the parameters that represent statistical dependencies in these models are easily identified, the way dependencies are reflected in model behavior is not well understood. In particular, it is seldom clear whether a given numerical value for a dependence parameter indicates dependence that is weak, moderate, or strong. Pickard (1977) considered an Ising model, which is a special case of a model with binary conditionals, and defined "critical values" for parameters that lead to correlations remaining non-zero as distance between sites grows large. In general, investigators who have worked with auto-models are aware that, in simulated data sets, values of dependence parameters that are "too large" can produce chaotic data behavior or random fields with constant value, even for values of the dependence parameters that should be theoretically possible. Unfortunately, what constitutes values that are "too large" remains a blurry aspect of the behavior of even simple Markov random field models.

The fundamental thesis of this article is that an appropriate conceptualization of dependence for Markov random field models is the difference between expectations conditional on neighboring values and expectations given independence with neighboring values. Using this basic notion of statistical dependence, we can make substantial progress in quantifying the strength of dependencies represented by given numerical parameter values, and in understanding conditions that are sufficient to guarantee non-degenerate model behavior.

The remainder of the article is organized as follows. Basic underlying theory important for Markov random field models with exponential family conditionals is given in Section 2. Section 3 proposes a quantification of dependencies in these models and develops a relation between that quantification and dependence parameters in the model. Sections 4, 5, and 6 deal with specific forms of dependence measures for simple models based on Gaussian, binary, and Winsorized Poisson conditionals, respectively, and illustrate some particulars of how these dependence measures compare with quantities based on other concepts of dependence, such as correlation. Section 7 briefly considers extensions of the fundamental ideas introduced to more complex models, and concluding remarks are given in Section 8.

2   Models Based on One-Parameter Exponential Families

A wide variety of situations can be formulated in terms of Markov random fields, but among the most natural are spatial settings, and the focus throughout this article will be on spatial problems in two-dimensional physical space. Everything contained here, however, generalizes to other random field settings in a straightforward manner, depending only on the definition of neighborhood structure. To formalize the setting, then, assume that there are available a set of spatial locations $\{\boldsymbol{s}_i : i = 1, \ldots, n\}$ where $\boldsymbol{s}_i \equiv (u_i, v_i)$ denotes a location at horizontal coordinate u_i and vertical co-

ordinate v_i. It is convenient to consider locations on a regular square lattice with $(u_i, v_i) \in \{1, 2, \ldots, L\} \times \{1, 2, \ldots, L\}$. Assume also that each location has a designated neighborhood $N_i \equiv \{\boldsymbol{s}_j : \boldsymbol{s}_j \text{ is a neighbor of } \boldsymbol{s}_i\}$. The simple configurations corresponding to what are called four-nearest and eight-nearest neighbors are well suited for use with regular lattices and will be used repeatedly what follows. Define random variables corresponding to locations as $\{Y(\boldsymbol{s}_i) : i = 1, \ldots, n\}$ and define values of these variables at neighboring locations of $\boldsymbol{s}_i$ as $\boldsymbol{y}(N_i) \equiv \{y(\boldsymbol{s}_j) : \boldsymbol{s}_j \in N_i\}$.

Given locations, neighbors, and neighboring values, a Markov random field model results from specifying, for $i = 1, \ldots, n$, the conditional probability mass or density functions $f(y(\boldsymbol{s}_i)|\{y(\boldsymbol{s}_j) : j \neq i\}) = f(y(\boldsymbol{s}_i)|\boldsymbol{y}(N_i))$. Under the provisos of known variance for normals and known number of independent and identically distributed binary trials for binomials, all of the distributions we consider here can be written as one-parameter exponential families of the form,

$$f\{y(\boldsymbol{s}_i)|\boldsymbol{y}(N_i)\} = \exp\left[A_i(\boldsymbol{y}(N_i))\,y(\boldsymbol{s}_i) - B_i(\boldsymbol{y}(N_i)) + C(y(\boldsymbol{s}_i))\right]. \tag{1}$$

In (1) $A_i(\cdot)$ is called the natural parameter function and, in parallel with traditional exponential family representations, $B_i(\cdot)$ is a function of $\boldsymbol{y}(N_i)$ only through $A_i(\boldsymbol{y}(N_i))$.

To render models based on probability density or mass or functions (1) useful, the natural parameter functions $A_i(\cdot)$ must be given more explicit form. Besag (1974) demonstrated that a necessary form of parameterization is,

$$A_i(\boldsymbol{y}(N_i)) = \alpha_i + \sum_{\boldsymbol{s}_j \in N_i} \eta_{i,j}\,y(\boldsymbol{s}_j), \tag{2}$$

where $\eta_{i,j} = \eta_{j,i}$ and one might need to specify other restrictions on the allowable values of the α_i and $\eta_{i,j}$ for a joint distribution corresponding to the specified conditionals to be identified. Kaiser and Cressie (2000) make this construction process explicit and give conditions both necessary and sufficient for a joint to be identified through the methodology originated by Besag (1974). Arnold *et al.* (1992) give more general conditions required for existence of a joint, although these conditions do not necessarily result in substantial guidance for construction of practical

models. In all of the models considered in this article the positivity condition of Besag (1974) will be assumed, as will pairwise-only dependence, noting that this latter condition is guaranteed for four-nearest neighbor structures by virtue of the Hammersley-Clifford theorem, but remains an assumption for other neighborhood arrangements.

A departure from other presentations of exponential family Markov random field models that will be used in this article is parameterization of the functions $A_i(\cdot)$ as, for $i = 1, \ldots, n$,

$$A_i(\boldsymbol{y}(N_i)) = \tau^{-1}(\kappa_i) + \sum_{\boldsymbol{s}_j \in N_i} \eta_{i,j} \left\{ y(\boldsymbol{s}_j) - \kappa_j \right\}, \tag{3}$$

where $\eta_{i,j} = \eta_{j,i}$, and τ^{-1} maps expected values into natural parameters for the exponential family in expression (1). Distributing the sum and collecting constants verifies that this parameterization satisfies the structure of expression (2). In (3), independence models result from taking all $\eta_{i,j} = 0$ so that $\tau^{-1}(\kappa_i)$ represents the natural parameter of a model without statistical dependence. Furukawa (2004) demonstrates that, under suitable restrictions on the magnitude of the $\eta_{i,j}$, the parameterization (3) renders κ_i nearly equal to the marginal mean of $Y(\boldsymbol{s}_i)$; $i = 1, \ldots, n$ for binary, binomial, and Winsorized Poisson models; further investigation of how these suitable restrictions on the $\eta_{i,j}$ can be identified is contained in the next section. Call (3) a *centered* parameterization for one-parameter exponential family Markov random field models. Models based on Gaussian conditionals are typically given a centered parameterization in terms of conditional expectations rather than natural parameters (e.g., Cressie, 1993), and Kaiser and Cressie (1997) also use what amounts to a centered parameterization for models with Winsorized Poisson conditionals. Caragea and Kaiser (2007) discuss the implications of centered parameterizations for considering model components as representing large-scale and small-scale spatial structure in the case of models with binary conditionals.

3 Quantifying Dependence

It is clear from expressions (2) and (3) that statistical dependencies in exponential family Markov random field models are captured through the values of the parameters $\{\eta_{i,j} : i = 1, \ldots, n; j \in N_i\}$. What is less clear is how to interpret numerical values of these parameters in terms of the type or strength of dependencies represented. Meaningful interpretation can be provided by considering the fundamental notion of statistical dependence that underlies Markov random field models. Mathematically, these models "capture" spatial dependence as the arithmetic difference in natural parameters under formulations that do and do not incorporate conditioning on neighbors. Thus, one key to lending interpretation to dependence parameters is the connection between differences of expectations and differences of natural parameters in exponential families of the type considered.

The basic notion of statistical dependence as the difference between expectation given conditioning values and expectation under independence has, under certain model structures, a connection with other well-known concepts of dependence such as covariance and correlation. Consider, for example, a pair of bivariate normal random variables, X and Y, say. The expected value of X conditional on $Y = y$ may be written as,

$$
\begin{aligned}
E(X|Y = y) &= E(X) + \frac{cov(X,Y)}{var(Y)}\{y - E(Y)\} \\
&= E(X) + \rho\, \frac{\sqrt{var(X)}}{\sqrt{var(Y)}}\{y - E(Y)\} \\
&= E(X) + F\{y - E(Y)\}, \tag{4}
\end{aligned}
$$

where ρ in the second line above is the correlation between X and Y. In (4), notice that (i) the difference between the conditional and marginal expectations of X may be written as a factor of the difference between the conditioning value of $Y = y$ and the marginal expectation of Y and (ii) the marginal expectation of X is also the expectation that would be given by this model if $cov(X,Y) = 0$, that is, under an independence model. In this case correlation between X and Y is proportional to

the factor F but, as will be demonstrated in Section 4, the factor F and correlation ρ represent different measures of the strength of dependence in spatial settings, even for models with Gaussian conditionals.

I propose that the notion of statistical dependence embodied in the factor F of expression (4) can be applied directly to Markov random field models. To develop this more formally, consider a simple version of (3) in which all neighborhoods are of the same size, $|N_i| = m; i = 1, \ldots, n$ and all dependence parameters are equal valued, $\eta_{i,j} = \eta; \; i = 1, \ldots, n; j \in N_i$. Then let $\gamma = m\eta$ and rewrite the natural parameter function of expression (3) as,

$$A_i(\boldsymbol{y}(N_i)) = \tau^{-1}(\kappa_i) + \gamma \frac{1}{m} \sum_{\boldsymbol{s}_j \in N_i} \{y(\boldsymbol{s}_j) - \kappa_j\}. \tag{5}$$

Denote the expected value of random variables $Y(\boldsymbol{s}_i)$ under an independence model as $E\{Y(\boldsymbol{s}_i)|\varnothing\}$ which, under parameterization (5), is $E\{Y(\boldsymbol{s}_i)|\varnothing\} = \kappa_i; i = 1, \ldots, n$. The dependence of $Y(\boldsymbol{s}_i)$ on its neighbors may now be cast as a bivariate dependence between $Y(\boldsymbol{s}_i)$ and the average deviation of neighboring values from their independence expectations, which is $(1/m)\sum_{\boldsymbol{s}_j \in N_i}\{y(\boldsymbol{s}_j) - \kappa_j\}$. The suggestion, then, is that the dependence between $Y(\boldsymbol{s}_i)$ and it's neighbors $\boldsymbol{Y}(N_i)$ is embodied in the quantity F_i where,

$$
\begin{aligned}
E\{Y(\boldsymbol{s}_i)|\boldsymbol{Y}(N_i) = \boldsymbol{y}(N_i)\} - E\{Y(\boldsymbol{s}_i)|\varnothing\} &= \tau(A_i) - \kappa_i \\
&= F_i \frac{1}{m} \sum_{\boldsymbol{s}_j \in N_i} \{y(\boldsymbol{s}_j) - \kappa_j\}. \tag{6}
\end{aligned}
$$

Now, expression (5) also gives that $(1/m)\sum_{\boldsymbol{s}_j \in N_i}\{y(\boldsymbol{s}_j) - \kappa_j\} = (1/\gamma)\{A_i - \tau^{-1}(\kappa_i)\}$ which, when substituted into (6) yields, for $i = 1, \ldots, n$,

$$F_i = \gamma \left[\frac{\tau(A_i) - \kappa_i}{A_i - \tau^{-1}(\kappa_i)} \right] = \gamma\, Q(A_i, \kappa_i). \tag{7}$$

Note that $Q(\cdot)$ is implicitly a function of $\{\kappa_j : j \in N_i\}$, $\boldsymbol{y}(N_i)$, and γ, because A_i is a function of these values. Note also that $Q(\cdot)$ will always be non-negative.

The quantities $F_i; i = 1, \ldots, n$ in (6) and (7) are *local* measures of dependence, and may vary among locations in a given spatial domain, depending not only on values of the parameters $\{\kappa_i; i = 1, \ldots, n\}$ and γ but also, for a given i, the neighboring

values $\boldsymbol{y}(N_i) = \{y(\boldsymbol{s}_j) : \boldsymbol{s}_j \in N_i\}$. In fact, for a constant $\kappa_i = \kappa$; $i = 1, \ldots, n$ and given γ, the values F_i will be a deterministic function of these neighboring values. In contrast to F_i, the dependence parameter γ in (5) and (7) has a constant value for all variables in the domain of interest, but is related to F_i through the function $Q(\cdot)$. Because all dependencies in the model are governed by γ, this parameter must function as a global measure of dependence but, as already mentioned, heretofore we have lacked any scale against which to judge values of γ. An appropriate scale will emerge from investigating the behavior of the function $Q(\cdot)$, but there are several complications that need discussion before this can be accomplished in a satisfactory manner.

One distinction between the value F_i appearing in (6) and (7) and the somewhat analogous F of expression (4) is that F_i is defined using the difference between conditional and independence model expectations, while (4) can be given meaning using either this difference or the difference between conditional and marginal expectations; in (4) the conditional expectation of X in a model with dependence, and the expected value of X under independence are identical. This does not necessarily remain true for Markov random field models based on conditional distributions other than Gaussian forms. As already mentioned, however, the parameterizations (3) and (5) result in the parameters κ_i being the independence model expectations and *nearly* the marginal expectations under dependence. This will be verified for a simple binary model using exact numerical computations in Section 5. However, this near correspondence between marginal expectations and κ_i breaks down for values of dependence parameters outside of certain ranges that can be determined by relating their endpoints to bounds on $|F_i| < 1$.

To motivate a restriction that $|F_i| < 1$; $i = 1, \ldots, n$ in (6) and (7), consider the intuition that in "nicely behaved" models, marginal expectations should have at least as much influence over conditional expectations as the conditioning values. This must be true, for example, for data realizations simulated from a model to exhibit the same pattern of large-scale structure, such as increasing trend in a given

spatial direction. If this restriction does not hold, then large-scale data structure is determined through an amalgamation of local neighborhood effects governed by dependence not marginal expectation. In expression (4), the absolute value of the factor F multiplying the difference $\{y - E(Y)\}$ is less than or equal to 1 if we either (i) consider standardized versions of X and Y, (ii) assume that the covariance matrix is diagonally dominant (which is sufficient for non-negative definiteness), or (iii) assume $var(X) = var(Y)$.

Extending the intuition of the previous paragraph to Markov random field models, if the independence expectation κ_i has more influence over the conditional expectation $E\{Y(\boldsymbol{s}_i)|\boldsymbol{y}(N_i)\}$ than does the average of neighboring values, then the model will be "nicely behaved", and κ_i will also be close to the marginal expectation of $Y(\boldsymbol{s}_i)$. From (6), this will be true when $|F| \leq 1$. This issue will be revisited in what follows but, if the intuitive argument is accepted for the present, we may then define *standard bounds* for γ by taking $|F| = 1$ and considering $Q(A_i, \kappa_i)$ in (7) as a function of $\tau(A_i)$ for a given κ_i. For a specified conditional exponential family the possible values of $\tau(A_i)$ are known since these are conditional expectations. Denote this space of allowable values for $\tau(A_i)$ as Θ. Noting that the values of $\{\kappa_i : i = 1, \ldots, n\}$ must also lie in Θ allows the further definition of *uniform standard bounds* for γ by considering standard bounds to be a function of κ_i. Then define standard bounds for γ as,

$$|\gamma| \leq \left[\sup_{\tau(A_i) \in \Theta} \{Q(A_i,\, \kappa_i)\} \right]^{-1}, \tag{8}$$

and uniform standard bounds for γ over all possible values for κ_i as,

$$|\gamma| \leq \left[\sup_{\kappa_i \in \Theta} \sup_{\tau(A_i) \in \Theta} \{Q(A_i,\, \kappa_i)\} \right]^{-1}. \tag{9}$$

The inequalities in (8) and (9) have been called standard bounds because they are not necessarily required for models to be valid, that is, to have joint distributions. Following the previous argument based on intuitive reasoning I conjecture, however, that these bounds correspond to allowable values for γ that lead to the κ_i

being within a small neighborhood (mathematical, not spatial here) of the marginal expectations $E\{Y(\boldsymbol{s}_i)\}$; $i = 1, \ldots, n$. In this case then, the differences between conditional expectations and independence model expectations on the left hand side of expression(6) are nearly the same as differences between conditional and marginal expectations, and spatial structures generated by the models are dominated by marginal expectations or large-scale model components rather than dependencies or small-scale model components.

The values of the dependence parameter γ and the local dependencies F_i; $i = 1, \ldots, n$ are related through the function $Q(A_i, \kappa_i)$ of expression (7). The exact form of this function will differ depending on the types of conditional distributions chosen for a model and closer examination will be delayed until the subsequent sections of the article, but several behaviors of $Q(\cdot)$ are generally available. First, as a function of $\tau(A_i)$ for a given κ_i, $Q(A_i, \kappa_i)$ in (6) has a singularity at $\tau(A_i) = \kappa_i$. As $\tau(A_i)$ approaches κ_i, however,

$$\lim_{\tau(A_i) \to \kappa_i} Q(A_i, \kappa_i) = \lim_{\tau(A_i) \to \kappa_i} \left[\frac{d\,\tau(A_i)}{d\,A_i} \right] = \left[\frac{d\,\tau^{-1}(\kappa_i)}{d\kappa_i} \right]^{-1} = var\{Y(\boldsymbol{s}_i)|\emptyset\}, \quad (10)$$

the variance of $Y(\boldsymbol{s}_i)$ under an independence model. The last equality above follows from the correspondence of $\tau(\cdot)$ with $b'(\cdot)$ in the usual exponential family expression $f(y) = \exp\left[\theta y - b(\theta) + c(y)\right]$. Assuming $\tau(\cdot)$ is continuous, Q will be otherwise continuous and non-negative. It may be constant, monotone or convex, as will be shown for Gaussian, Winsorized Poisson, and binary models, respectively. If Q is convex for $\tau(A_i) \in \Theta$ and its maximum does not occur at $\tau(A_i) = \kappa_i$, differentiation shows that its maximum will occur at a value A_i^* such that,

$$Q(A_i^*, \kappa_i) = \left. \frac{d\,\tau(A_i)}{d\,A_i} \right|_{A_i = A_i^*} = var\{Y(\boldsymbol{s}_i)|A_i^*\}, \quad (11)$$

where $var\{Y(\boldsymbol{s}_i)|A_i^*\}$ denotes the conditional variance of $Y(\boldsymbol{s}_i)$ given a natural parameter function with value A_i^* and, as previously, the last step follows from the correspondence between Markov random field and ordinary versions of one-parameter exponential families.

In the development leading to expressions (8) through (11), $\tau(A_i)$ has been considered over the entire parameter space Θ, even for a fixed value of κ_i. For models formulated on the basis of discrete conditional distributions this is not strictly proper. For a given κ_i and set $\{\kappa_j : \boldsymbol{s}_j \in N_i\}$, (5) indicates that A_i may assume only a discrete and, in the cases of binary and Winsorized Poisson models, finite, number of values. How many such values there are depends on the size of the neighborhood chosen, and considering all $\tau(A_i) \in \Theta$ then allows expressions (8) through (11) to apply to models with any neighborhood size. Little is lost in adopting this convention, which will become clear in the more detailed considerations of binary and Winsorized Poisson models in Sections 6 and 7.

4   Models with Gaussian Conditionals

In this section the specific forms of quantities introduced in Section 3 will be examined for a simple model with Gaussian conditionals and, in particular, the proposed measures of dependence F_i in (6) will be related to several possible correlations that fit the structure of expression (4). Writing the Gaussian or normal density as a one-parameter exponential family, the natural parameter function of expression (5) is, with $\kappa_i = \kappa$ for $i = 1, \ldots, n$,

$$A_i(\boldsymbol{y}(N_i)) = \frac{\kappa}{\sigma^2} + \gamma \frac{1}{m} \sum_{\boldsymbol{s}_j \in N_i} \{y(\boldsymbol{s}_j) - \kappa\}. \tag{12}$$

In this model, $E\{Y(\boldsymbol{s}_i)|\boldsymbol{y}(N_i)\} = \sigma^2 A_i(\boldsymbol{y}(N_i))$ and $var\{Y(\boldsymbol{s}_i)|\boldsymbol{y}(N_i)\} = \sigma^2$. Notice that (12) is written in direct exponential family form rather than in terms of the conditional expectations, as is often done (e.g., Cressie, 1993), although tradition is followed in taking the conditional variances σ^2 to be constant. For the current purpose, which is investigating the way statistical dependence is represented by models rather than estimation, assume that σ^2 is known. For this Gaussian model, $\Theta \equiv (-\infty, \infty)$ and, from (7), $Q(\tau(A_i), \kappa) = \sigma^2$ for all $\tau(A_i) \in \Theta$, or $F_i = F = \gamma \sigma^2$. Considering positive statistical dependence, the standard upper bound and

the uniform standard upper bound on γ are both $\gamma < 1/\sigma^2$.

In this well-known model the joint distribution is Gaussian with constant mean κ and covariance matrix $\Sigma = (I - (\sigma^2 \gamma/m)H)^{-1} M$, where I is the $n \times n$ identity matrix, M is diagonal with non-zero elements all equal to σ^2, and H is $n \times n$ with $(i, j)^{th}$ element equal to 1 if locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are neighbors and equal to 0 if they are not neighbors. A square regular lattice with four-nearest neighbors structure wrapped on a torus provides a convenient theoretical tool to investigate the relations between γ, F, and elements of the joint covariance matrix for this model. For this neighborhood structure, the value of κ is irrelevant to the joint covariance matrix. For a given γ, marginal variances on the diagonal of Σ are equal, as are covariances for any two locations separated by the same distance on the lattice; in particular, the covariance of $Y(\boldsymbol{s}_i)$ and the variables at each of its four neighbors are equal, although the covariances among the neighbors are not. This model might be described as being stationary and isotropic. The value of the conditional variance σ^2 scales all variances and covariances an equal amount and so, while important for the values of these quantities, σ^2 will not change correlations.

A numerical investigation was conducted by computing the joint covariance matrix for a square 30×30 lattice wrapped on a torus, with $\sigma^2 = 1$ and for various values of γ. The values of Σ converge elementwise as lattice size increases and, although the rate of convergence apparently depends on the magnitude of γ, I found a 30×30 lattice large enough to have reached convergence to 6 decimal places in all elements of the covariance matrix and for all of the values of γ used. Two correlations were compared to the values of F obtained (here, $F = \gamma$), these being the pairwise marginal correlation between a location and any one of its neighbors, and the correlation between a location and the sum of its neighbors; the variance of this sum is available from having access to the full covariance matrix. Both of these correlations are the same for all locations on the lattice, although they are not equal to each other. The joint distribution of the random variable at a location and any one of its neighbors will be bivariate normal, as too will the joint distribution of the

variable and the sum of its neighbors. Thus, both of these correlations must satisfy the relation of expression (4) exactly.

The standard upper bound for γ with $\sigma^2 = 1$, is $\gamma < 1$, which is exactly the same as what is needed for positive definiteness of the covariance matrix based on eigenvalues of the neighborhood matrix H (e.g., Cressie, 1993, p. 471) when computed for four neighbors on a torus. That is, for this model based on conditional Gaussian distributions, the standard bound on γ is the same as the uniform standard bound on γ is the same as what is known to be necessary for a positive definite covariance matrix in the joint distribution. In the numerical investigation then, 25 values of γ on the interval $(0, 1)$ were used. Figure 1 presents a plot of the two correlations examined against F. The curve labeled "pairwise" depicts the correlation between a location and any one of its neighbors while that labeled "neighborhood" depicts the correlation between a location and the sum of its neighbors. The striking feature of Figure 1 is that neither of these correlations exceeds 0.5 until F exceeds 0.8 (interpolating, neighborhood correlation is 0.5 when F is about 0.85), although both approach 1 as F approaches 1.

The point to be taken from this illustration is not that either the meaning of correlation or the meaning of differences between marginal and conditional expectations have changed. The point is that they are no longer equivalent, as they are in the bivariate case that corresponds to expression (4). Mathematically, the reasons for this depend on whether one considers the pairwise correlation between a location and any one of its neighbors, or the neighborhood correlation between a location and the sum of its neighbors. In the pairwise case, the marginal variances of $Y(\boldsymbol{s}_i)$ and the variable at one of its neighbors, $Y(\boldsymbol{s}_j^*)$ say, are equal (although they change as F changes) and the joint distribution of these variables must be bivariate normal. But these variables do not fit the structure of a situation in which expressions (4) and (6) should be equivalent because in this case expression (4) gives the expected value of $Y(\boldsymbol{s}_i)$ conditional on only the one neighbor chosen, $Y(\boldsymbol{s}_j^*)$, while expression (6) gives the expected value of $Y(\boldsymbol{s}_i)$ conditional on the set of all its neighbors. For

neighborhood correlation, however, the conditional expectation of (6) is identical to the conditional expectation of (4) with $Y = \sum_{\boldsymbol{s}_j \in N_i} Y(\boldsymbol{s}_j)$, but now the variances corresponding to $var(X)$ and $var(Y)$ of expression (4) are no longer equal, and they change in different ways as F changes, so ρ and F no longer have the same meanings. Because any variables that satisfy expression (4) must also be such that $var(X|Y) = var(X)(1 - \rho^2)$, the marginal variances corresponding to the values of Figure 1 balloon as $F \to 1$ in order for the conditional variance to remain constant. For example, in Figure 1 with conditional variance $\sigma^2 = 1$, marginal variances were 2.467 at $F = 0.9960$ and 280.138 at $F = 0.9999$. These marginal variances could be controlled by allowing conditional variance to vary, but this would have no effect on the correlations of Figure 1 which are invariant to changes in conditional variance.

The measure of dependence F seems preferable to either pairwise or neighborhood correlation for spatial problems, even with Gaussian models, because it has more ready interpretation. In the case of an independence model with constant expectation κ, the conditional expectation $E\{Y(\boldsymbol{s}_i)|\boldsymbol{y}(N_i)\} = \tau(A_i)$ and the neighborhood average may be denoted as $\bar{y}_{N_i} = (1/m)\sum_{\boldsymbol{s}_j \in N_i} y(\boldsymbol{s}_j)$. Expression (6) may then be rewritten as,

$$\left(\frac{\tau(A_i) - \kappa}{\kappa} \right) = F \left(\frac{\bar{y}_{N_i} - \kappa}{\kappa} \right). \tag{13}$$

This indicates that F represents the fraction of the average proportional deviation from marginal expectation in the neighborhood that equals the proportional deviation of the conditional expectation from the marginal expectation at location $\boldsymbol{s}_i$. For example, if $F = 0.5$ and the average neighborhood value is 20% greater than the marginal expectation, then the conditional expectation is 10% greater than the marginal expectation. If the average neighborhood value is 20% lower than the marginal expectation, then the conditional expectation is 10% lower than the marginal expectation. In non-Gaussian models F in (13) would be replaced by F_i and for models with non-constant mean κ in (13) would be replaced with κ_i, but the essential interpretation would remain unchanged.

5   Models with Binary Conditionals

Continuing to assume $\kappa_i = \kappa$ for $i = 1, \ldots, n$ and constant neighborhood size m as in the previous section, (5) gives the natural parameter function for a model with binary conditional distributions as,

$$A_i(\boldsymbol{y}(N_i)) = \log\left\{\frac{\kappa}{(1-\kappa)}\right\} + \gamma\frac{1}{m}\sum_{\boldsymbol{s}_i \in N_i}\{y(\boldsymbol{s}_j) - \kappa\}. \tag{14}$$

Unlike the Gaussian model of the previous section, for this binary conditionals model it is not true that all local dependencies given by the F_i of (6) and (7) have the same value, nor is it true that standard bounds for γ are constant across values of κ. In this case, $\tau(\kappa) = \log(\kappa) - \log(1-\kappa)$ and $\tau(A_i) = \exp(A_i)/\{1 + \exp(A_i)\}$ so that, from (7),

$$Q(A_i, \kappa) = \frac{\exp(A_i) - \kappa\{1 + \exp(A_i)\}}{\{1 + \exp(A_i)\}\{A_i - \log(\kappa) + \log(1-\kappa)\}}. \tag{15}$$

Aside from the singularity at $\tau(A_i) = \kappa$, $Q(\cdot)$ in (15) is continuous and convex when considered as a function of $\tau(A_i)$ for a given κ. The use of expressions (8) and (11) then give standard bounds for $|\gamma|$ as $[\tau(A_i^*)\{1 - \tau(A_i^*)\}]^{-1}$, where $\tau(A_i^*)$ is the value in Θ that produces equality in expression (11). For most values of κ this A_i^* must be found numerically, although doing so is relatively easy since, for fixed κ, $Q(\cdot)$ is a function in one dimension. Figure 2 presents a graph of standard bounds across values of κ for this simple binary model, the uniform standard bound occurring at $\gamma = 4.0$ which corresponds to $\kappa = 0.5$. The bounds are symmetric about $\kappa = 0.5$, with values at $\kappa = 0.75$ and $\kappa = 0.25$ both being $|\gamma| \leq 4.2917$ and values at $\kappa = 0.90$ and $\kappa = 0.10$ both being $|\gamma| \leq 5.0664$.

The standard bounds of Figure 2 are not needed for models with binary conditionals to have joint distributions. The claim was made previously in Section 3, however, that standard bounds are needed to ensure that κ is not only the expected value of $Y(\boldsymbol{s}_i)$; $i = 1, \ldots, n$ under an independence formulation ($\gamma = 0$), but also nearly the marginal expected value in models including dependence. This can be demonstrated through exact calculations for a reduced situation involving a spatial

transect in one dimension. For a binary conditionals model having natural parameter function (14), locations $\boldsymbol{s}_i \equiv \{1, 2, \ldots, D\}$ and neighborhoods $N_i \equiv \{\boldsymbol{s}_j = \boldsymbol{s}_i \pm 1\}$, the common marginal expectation of the $Y(\boldsymbol{s}_i)$ may be computed directly through the negpotential function (e.g., Besag, 1974; Kaiser and Cressie, 2000) which in this case becomes,

$$H(\boldsymbol{y}) = \sum_{i=1}^{n} \left[\log \left\{ \frac{\kappa}{(1-\kappa)} \right\} - \eta\kappa \right] y(\boldsymbol{s}_i) + \frac{\gamma}{m} \sum_{i,j \in \mathcal{P}} y(\boldsymbol{s}_j) y(\boldsymbol{s}_i), \qquad (16)$$

where $\mathcal{P} \equiv \{i, j : 1 \leq i < j \leq n; \boldsymbol{s}_j \in N_i\}$ and for this model $m = 2$. I have denoted the negpotential as the function $H(\cdot)$ rather than the usual $Q(\cdot)$ used in most references only because Q has already been used for the function in (7). With Ω denoting the support of the joint probability mass function, the common marginal expected value for $Y(\boldsymbol{s}_i)$; $i = 1, \ldots, n$ can be computed as,

$$E\{Y(\boldsymbol{s}_i)\} = \frac{\sum_{\boldsymbol{y} \in \Omega} y(\boldsymbol{s}_i) H(\boldsymbol{y})}{\sum_{\boldsymbol{y} \in \Omega} H(\boldsymbol{y})}. \qquad (17)$$

Exact computation of (17) is feasible on a transect with two neighbors because the joint support Ω remains of manageable size for transects long enough for marginal expectations to have converged to at least 4 decimal places; recall that the joint distribution for Markov random field models depends to some degree on lattice size. Marginal expectations were computed using (17) for a binary conditionals model having natural parameter function (14) and 20 locations on a torus (here, circle). Expectations were calculated for $\kappa \in \{0.1, 0.2, \ldots, 0.9\}$, each with $\gamma \in \{0.2, 0.4, \ldots, 6.0\}$. The resulting 270 marginal expectations are plotted against values of γ in Figure 3, with a line representing each value of κ considered. Plotting symbols for the lines in Figure 3 occur in pairs for $\kappa \in \{0.1, 0.9\}$, $\kappa \in \{0.2, 0.8\}$, $\kappa \in \{0.3, 0.7\}$, $\kappa \in \{0.4, 0.6\}$, and a single line for $\kappa = 0.5$, because of the symmetries exhibited. This figure illustrates that, for values of γ less than the uniform standard bound of $\gamma = 4.0$, marginal expectations are reasonably close to the corresponding values of κ, although not exactly constant except for the case $\kappa = 0.5$. But, as γ exceeds this bound, the relation between κ and marginal expectation

becomes less easily discernible, even though the models remain well defined (i.e., have joint distributions). The "decay" in correspondence between κ and marginal expectations occurs more rapidly as γ exceeds the uniform standard bound of 4.0 for κ closer to 0.5, in concert with the differences between standard bounds and the uniform standard bound of Figure 2. Thus, Figure 3 illustrates, at least for a binary conditionals model, the relation between values of γ restricted by standard bounds and interpretability of κ as nearly marginal expectation. It is worthy of note that all of the calculations going into Figure 3 were produced using a centered version of a binary conditionals model. The interpretability allowed by the combination of centering and adhering to standard bounds for γ are not available in the typical version of this model used in applications (e.g., Gumpertz, Graham, and Ristaino, 1997; Wu, and Huffer, 1997).

Having described the overall effects of standard bounds for the dependence parameter γ, it is interesting to examine the manner in which the dependence measures F_i of expressions (6) and (7) vary across locations for fixed values of the parameters κ and γ in (14). As mentioned previously, for fixed values of these parameters, the natural parameter functions $A_i; i = 1, \ldots, n$ and hence also conditional expectations assume values in only a discrete finite set determined by the possible sums for neighboring values. That is, for fixed κ and γ, the F_i are exact functions of $\sum_{\boldsymbol{s}_j \in N_i} y(\boldsymbol{s}_j)$ or equivalently, the average deviation of neighboring values from κ, namely $(1/m) \sum_{\boldsymbol{s}_j \in N_i} \{y(\boldsymbol{s}_j) - \kappa\}$. For a four-nearest neighborhood structure there will be 5 such values, while for an eight-nearest neighborhood structure there will be 9 possible values. Figure 4 presents the effect of these possible neighboring values on the measure of local dependence F_i for situations in which $\kappa = 0.5$. In this figure, lines represent values of $\gamma = 4.0, 3.0, 2.0$, and 1.0 from top to bottom; for this kappa, $\gamma = 4.0$ is the standard bound and is also the uniform standard bound across all values of κ (see Figure 2). Plotting symbols in Figure 4 are 'x' for a model with eight-nearest neighbors, and 'o' for a model with four-nearest neighbors. It can be seen from Figure 4 that dependence is not constant across neighboring values and

the degree of non-constancy diminishes as γ decreases in value.

Figure 5 presents the same information as Figure 4 for a fixed value of $\kappa = 0.75$. In this case, the standard bound is $\gamma = 4.2918$ (c.f., Figure 2) and the top line of Figure 5 illustrates this case. The remaining lines are as in Figure 4, for $\gamma = 4.0$ (the uniform standard bound), and $\gamma = 3.0, 2.0$ and 1.0. Here, dependence is not symmetric about κ, which is intuitive since the possible values of neighborhood deviation are not centered at zero. The greatest value of $F_i = 0.9995$ occurs in Figure 5 for a value of -0.375 in average deviation of neighboring values. The value that would correspond to the standard bound on γ computed from (11) with A_i^* allowed to vary continuously over Θ is $F_i = 1.0$ at average deviation -0.38, verifying the previous claim that not much is sacrificed in computing standard bounds by allowing the natural parameter function to vary continuously over its full parameter space rather than restricted to the discrete set of values possible for a fixed κ. A plot for $\kappa = 0.25$ is not shown but would be a mirror image of Figure 5, with a maximum of $F_i = 0.9995$ occurring at a value of 0.375 for average neighborhood deviation. Figure 6, for $\kappa = 0.90$ illustrates that as κ becomes more extreme the discrepancy between standard bound and uniform standard bound increases.

6   Models with Winsorized Poisson Conditionals

In this section I consider models with Winsorized Poisson conditionals as introduced by Kaiser and Cressie (1997). With $\kappa_i = \kappa; i = 1, \ldots, n$, (5) gives the natural parameter function for this model as,

$$A_i(\boldsymbol{y}(N_i)) = \log(\kappa) + \gamma \frac{1}{m} \sum_{\boldsymbol{s}_j \in N_i} \{y(\boldsymbol{s}_j) - \kappa\}, \tag{18}$$

where $0 < \kappa < R$ for a chosen Winsorization value R, and $-\infty < A_i < \log(R)$. Similar to the binary conditionals model of the previous section, local dependencies given by $F_i; i = 1, \ldots, n$ vary as a function of neighboring values. With $\tau(A_i) =$

$\exp(A_i)$ and $\tau^{-1}(\kappa) = \log(\kappa)$, expression (7) becomes,

$$Q(A_i, \kappa) = \frac{\exp(A_i) - \kappa}{A_i - \log(\kappa)}. \tag{19}$$

At first glance, one might suppose that, for a fixed κ, the function Q in (19) would behave in a manner similar to $exp(x)/x$; $-\infty < x < \log(R)$, which describes a curve much like a gamma or hyperbolic function with center at 0, a point where the function does not exist. But, the actual behavior of (19) is that of $(x-y)/\{log(x) - log(y)\}$ for $0 < x < R$ and fixed y, which can be shown to be monotone increasing in x for any $0 < y < R$, although doing so is not necessarily a trivial exercise. The result is that (19) is a monotone increasing function of $-\infty < A_i < \log(R)$ or, equivalently, $0 < \exp(A_i) < R$. Standard bounds for γ given a fixed κ are then, from (8),

$$|\gamma| < \left[\lim_{\tau(A_i) \to R} Q(A_i, \kappa) \right]^{-1} = \frac{\log(R) - \log(\kappa)}{R - \kappa}. \tag{20}$$

The function $(R-\kappa)/\{\log(R)-\log(\kappa)\}$ exhibits the same behavior as $(x-y)/\{log(x) - log(y)\}$ for $0 < y < R$ and x fixed, which is again monotone increasing, and then uniform standard bounds are, from (9),

$$|\gamma| < \left[\lim_{\kappa \to R} \lim_{\tau(A_i) \to R} Q(A_i, \kappa) \right]^{-1} = \left[\lim_{\kappa \to R} \frac{R - \kappa}{\log(R) - \log(\kappa)} \right]^{-1} = \frac{1}{R}. \tag{21}$$

Figure 7 presents a graph of standard bounds for $|\gamma|$ across a number of values of κ for the Winsorization value $R = 20$.

The exact common marginal expected value of the $Y(\boldsymbol{s}_i)$ is difficult to compute for the Winsorized Poisson model due to the size of the joint support Ω, but Monte Carlo simulation verifies that exceeding the standard bounds on γ results in the same type of unruly behavior shown for the binary case in Figure 3 (Furukawa, 2004). With the standard bounds enforced, κ can be interpreted as nearly the marginal expectation.

As was the case for the binary conditionals model, for a fixed κ the natural parameter function of expression (18) can assume only a discrete and finite number of values depending on the sum of neighboring values $\{y(\boldsymbol{s}_j) : j \in N_i\}; i = 1, \ldots, n.$

For this Poisson model, however, the number of such possible values is rather large, being R^m for number of neighbors m and Winsorization value R; with $R = 20$ there are $160,000$ possible values in a four-nearest neighbor structure. Thus, the convention of allowing the natural parameter to vary continuously over its entire parameter space in calculation of standard bounds, even for a fixed κ, will have negligible effect. The large number of possible neighboring sums also allows the effect of the average deviation among neighbors, $(1/m) \sum \{y(\boldsymbol{s}_j) - \kappa\}$, on dependencies F_i to be presented as a continuous curve (plotting points results in the same graphs on any scale that allows standard size paper to be used). A graph of local dependence measures F_i against average deviation of neighboring values is presented in Figure 8 for the case $\kappa = 5$ and $R = 20$, and in Figure 9 for $\kappa = 10$ and $R = 20$. As for the binary conditionals model of Section 5, values of F_i in these figures follow the same curve for four-nearest and eight-nearest neighbor situations. In contrast to the situation for the binary conditionals model, Figure 8 illustrates a rather dramatic difference between the strengths of dependencies that can be represented with γ allowed to achieve its standard bound as opposed to only its uniform standard bound (compare Figure 8 with Figure 5 and even Figure 6). Figure 9 shows that this difference diminishes as κ becomes nearer R, which is also indicated by expressions (20) and (21). The difference between standard and uniform standard bounds for Winsorized Poisson models is of practical import, as random variables following such a model exhibit Poisson-like behavior for values of κ small relative to the Winsorization value R (Kaiser and Cressie, 1997).

For this simple Winsorized Poisson model, dependence at a location $\boldsymbol{s}_i$ as represented by F_i increases as the departure of the neighboring values from κ increases. The range of such dependencies available depends on the values of both κ and γ, as illustrated by Figures 8 and 9. This property is shared by the binary conditionals model of Section 5, but the Gaussian conditionals model of Section 4 has constant F_i values for a fixed set of parameters.

7 Models with Non-Constant Parameters

To this point in the article, the dependence parameters given originally in expression (3) as $\eta_{i,j}$ have been assumed to be constant, beginning with $\gamma = m\eta_{i,j} = m\eta$ for all i, j in (5). In Sections 5, 6, and 7, models were considered to have all $\kappa_i = \kappa$ which, for dependence parameters within their standard bounds, results in nearly constant marginal mean. Models with the number of parameters reduced in these ways have been used to demonstrate relations among various quantities considered and have simplified the derivation of standard and uniform standard bounds for γ. But nothing in the original definition of the dependence measures F_i of expression (6) depends on having constant κ_i and the definition is, in fact, totally independent of any dependence parameters. Most of what has been presented in this article can be extended to more complex models in a straightforward manner, although some complications do arise. In this section, extensions to models with non-constant parameters are briefly summarized.

7.1 Non-Constant Marginal Mean Structures

That under suitable restrictions the parameters κ_i of (3) and (5) correspond nearly to marginal expectations of $Y(\boldsymbol{s}_i)$ for $i = 1, \ldots, n$ is useful for incorporating the influence of covariates on large-scale model structure. One way this may be accomplished is along the lines of generalized linear models by taking, for some simple function $h(\cdot)$, $h(\kappa_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$; $i = 1, \ldots, n$, for a covariate vector $\boldsymbol{x}_i \equiv (x_{i,1}, \ldots, x_{i,p})^T$ and parameters $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_p)^T$. If $h(\cdot)$ is chosen as the canonical link for the one-parameter exponential family conditionals considered here, then the natural parameter function of expression (5) becomes,

$$A_i(\boldsymbol{y}(N_i)) = \boldsymbol{x}_i^T \boldsymbol{\beta} + \gamma \frac{1}{m} \sum_{\boldsymbol{s}_j \in N_i} \left\{ y(\boldsymbol{s}_j) - \tau(\boldsymbol{x}_i^T \boldsymbol{\beta}) \right\}, \tag{22}$$

and the relation between F_i and γ in expression (7) takes the form,

$$F_i = \gamma \left[\frac{\tau(A_i) - \tau(\boldsymbol{x}_i^T \boldsymbol{\beta})}{A_i - \boldsymbol{x}_i^T \boldsymbol{\beta}} \right] = \gamma \tilde{Q}(A_i, \boldsymbol{x}_i^T \boldsymbol{\beta}). \tag{23}$$

Similar expressions would result with other choices of $h(\cdot)$, with fairly obvious modifications to (22) and (23). Everything else from Section 3 would continue to hold as already presented, with the realization that standard bounds of expression (8) would be conditional on the different values of κ_i resulting from different values of the covariate vector. While this does not change expression (8) it introduces a complication; a single standard bound was available for the simple models of Sections 4, 5, and 6 because there was only one value of κ with which to contend. A conservative approach for a known covariates $\boldsymbol{x}_i$; $i = 1, \ldots, n$ would be to choose $\kappa_i = \max_i\{h^{-1}(\boldsymbol{s}_i)\}$ for use in computing a single standard bound for a given data set.

7.2   Non-Constant Dependence Parameters

Allowing dependence parameters to vary is slightly more involved than allowing different κ_i values. First note that the parameterization of (3), having $\{\eta_{i,j} : i = 1, \ldots, n;\ j \in N_i\}$ is more general than is practical, even subject to the necessary condition that $\eta_{i,j} = \eta_{j,i}$ for all i, j. The number of free parameters needs to be reduced for application to any actual problem to allow for estimation and inference. A strength of Markov random field models is that many structures are possible that allow dependence to vary while still maintaining a reasonably small number of free parameters (e.g., Kaiser, Daniels, Furukawa and Dixon, 2002; Zhu, Huang and Wu, 2005). For problems with locations on a regular lattice, a common approach is to allow dependence to vary among horizontal, vertical, and perhaps diagonal directions. Gumpertz *et al.* (1997) consider all of these possibilities in an application with eight nearest neighbors.

Suppose that large-scale structure is taken to be constant, but it is desired to divide a complete neighborhood into G directional groups as $N_i = \{N_i^1 \bigcup N_i^2 \bigcup \ldots, \bigcup N_i^G\}$ with the directional neighborhoods N_i^g; $g = 1, \ldots, G$ disjoint. In this case, a separate dependence parameter can be assigned to each directional neighborhood and

the natural parameter functions written as, for $i = 1, \ldots, n$,

$$A_i(\boldsymbol{y}(N_i)) = \tau^{-1}(\kappa) + \sum_{g=1}^{G} \gamma_g \frac{1}{m_g} \sum_{\boldsymbol{s}_j \in N_i^g} \{y(\boldsymbol{s}_j) - \kappa\}, \tag{24}$$

where $m_1, \ldots, m_G$ are the sizes of the directional neighborhoods and $m = \sum_g m_g$ is the total neighborhood size. Typically, the m_g; $g = 1, \ldots, G$ will all be the same value, but this is not necessary as long as conditions sufficient for existence of a joint are met.

To define directional versions of the local dependence measures F_i; $i = 1, \ldots, n$, we can apply expressions (6) and (7) to each directional neighborhood individually, ignoring all other directions. A set of G relations for each i emerges from the application of (6) as,

$$\tau(A_i) - \kappa = F_{i,g} \frac{1}{m_g} \sum_{\boldsymbol{s}_j \in N_i^g} \{y(\boldsymbol{s}_j) - \kappa\}; \quad g = 1, \ldots, G; \ i = 1, \ldots, n, \tag{25}$$

and, similarly, the application of expression (7) results in,

$$F_{i,g} = \gamma_g \left[\frac{\tau(A_i) - \kappa}{A_i - \tau^{-1}(\kappa)} \right] = \gamma_g Q(A_i, \kappa); \quad g = 1, \ldots, G; \ i = 1, \ldots, n. \tag{26}$$

In (25) and (26), the natural parameter function A_i appears in each equation for $g = 1, \ldots, G$, as does the function $Q(A_i, \kappa)$. This implies that exactly the same standard and uniform standard bounds developed for a single dependence parameter γ in (8) and (9) applies to *each* dependence parameter γ_g; $g = 1, \ldots, G$. To ensure that κ remains near the marginal expectation we need the additional restriction that the sum these parameters also adheres to the standard bound. If, for example, we are considering positive dependence parameters and γ_{sb} is the upper standard bound from (8) for a single parameter, then standard bounds for the set would be $\gamma_g < \gamma_{sb}$; $g = 1, \ldots, G$ subject to $\sum_g \gamma_g < \gamma_{sb}$.

The relation between $F_{i,g}$ and average deviation in directional neighborhoods $(1/m_g) \sum_{\boldsymbol{s}_j \in N_i^g} \{y(\boldsymbol{s}_j) - \kappa\}$ is, for each g, also identical to those shown previously (e.g., Figure 4, Figure 8). The relation between deviation of conditional from independence model expectations $E\{Y(\boldsymbol{s}_i)|\boldsymbol{y}(N_i)\} - E\{Y(\boldsymbol{s}_i)|\emptyset\} = \tau(A_i) - \kappa$ and

the simultaneous collection of directional neighborhood deviations is, from the G equations in expression (25) for each $i = 1, \ldots, n$,

$$
\begin{aligned}
\tau(A_i) - \kappa &= \sum_{g=1}^{G} \frac{m_g}{m} \{\tau(A_i) - \kappa\} = \sum_{g=1}^{G} \frac{m_g}{m} F_{i,g} \frac{1}{m_g} \sum_{\boldsymbol{s}_j \in N_i^g} \{y(\boldsymbol{s}_j) - \kappa\} \\
&= \frac{1}{m} \sum_{g=1}^{G} F_{i,g} \sum_{\boldsymbol{s}_j \in N_i^g} \{y(\boldsymbol{s}_j) - \kappa\},
\end{aligned}
\tag{27}
$$

and this is the directional equivalent of expression (6).

8   Concluding Remarks

The primary objective of this article has been to identify a quantifiable concept of statistical dependence that is applicable in general to Markov random field models based on one-parameter exponential families. The concept of dependence that has emerged is the factor that relates a difference between expectations conditional on and independent of neighboring values to the average deviation of conditioning values from their expectations under independence. This concept of dependence is familiar to statisticians in other contexts for which it is equivalent to covariance or correlation, but that equivalence does not hold for spatial problems. I have argued that dependence as "difference in expectations" is more interpretable for Markov random field models than is correlation, even in the case of Gaussian models.

Dependence as difference in expectations may be quantified through a set of local measures denoted as values F_i; $i = 1, \ldots, n$ in this article, and any F_i is allowed to assume a value between 0 and 1. For fixed model parameters, these local measures depend in a functional way on sums or averages of neighboring values. The relation between dependence measures F_i and dependence parameter γ in simple Markov random field models leads to restrictions on the range of values for γ, called standard bounds, that produce large-scale model structures nearly equal to marginal expectations. A global measure of dependence for an entire data set then suggests itself as $|\gamma|/\gamma_{sb}$, where γ_{sb} is the standard bound. Because standard bounds are,

in general, dependent on other model parameters, one could replace the standard bound γ_{sb} with a uniform standard bound. For some models this would have no effect (i.e., Gaussian), for some models this would have a moderate effect (i.e., binary) and for other models this could have a fairly substantial effect (i.e., Winsorized Poisson).

Extensions of results from simple Markov random field models to more complex models are entirely possible, although many more model structures are possible than those examined here, and further investigation will prove profitable. More difficult will be extending the ideas of dependence developed for models based on one-parameter exponential family conditional distributions to models based on multi-parameter exponential family conditionals. In fact, even the centering of parameterizations has not been successfully accomplished in multi-parameter models to date. There remain these, and other, open problems in interpretation and use of Markov random field models. Nevertheless, having a fundamental notion of the manner in which statistical dependencies are reflected by models, having a means of quantifying the strength of those dependencies, and having a way to find the related implications for values of specific model parameters are several steps toward a fuller understanding of statistical applications of Markov random field models.

References

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

Besag, J.E. (1974), Spatial interaction and the statistical analysis of lattice systems

(with discussion), *Journal of the Royal Statistical Society*, B **36**, 192-236.

Caragea, P. and Kaiser, M.S. (2007), Covariates and time in the autologistic model, Preprint 2006-19, Department of Statistics, Iowas State University, Ames, Iowa.

Cressie, N.A.C. (1993), *Statistics for Spatial Data*, rev. ed., Wiley: New York.

Furukawa, K. (2004), Development of Markov Random Field Models Based on Exponential Family Conditional Distributions, Unpublished PhD. dissertation, Iowa State University, Ames, Iowa.

Gumpertz, M.L., Graham, J.M., and Ristaino, J.B. (1997), Autologistic model of spatial pattern of Phytophtora epidemic in bell pepper: effects of soil variables on disease presence, *Journal of Agricultural, Biological and Environmental Statistics* **2**, 131-156.

Kaiser, M.S. and Cressie, N. (1997), Modeling Poisson random variables with positive spatial dependence, *Statistics and Probability Letters* **35**, 423-432.

Kaiser, M.S., Daniels, M.J., Furukawa, K. and Dixon, P. (2002), Analysis of particulate matter air pollution using Markov random field models of spatial dependence, *Environmetrics* **13**, 615-628.

Kaiser, M.S. and Cressie, N. (2000), The construction of multivariate distributions from Markov random fields, *Journal of Multivariate Analysis* **73**, 199-220.

Pickard, D.K. (1977), Asymptotic inference for an Ising lattice, II, *Advances in Applied Probability* **9**, 476-501.

Wu, H. and Huffer, F.W. (1997), Modeling the distribution of plant species using the autologistic regression model, *Environmental and Ecological Statistics* **4**, 49-64.

Zhu, J., Huang, H-C. and Wu, J. (2005), Modeling spatial-temporal binary data using Markov random field models, *Journal of Agricultural, Biological, and Environmental Statistics*, **10**, 212-225.
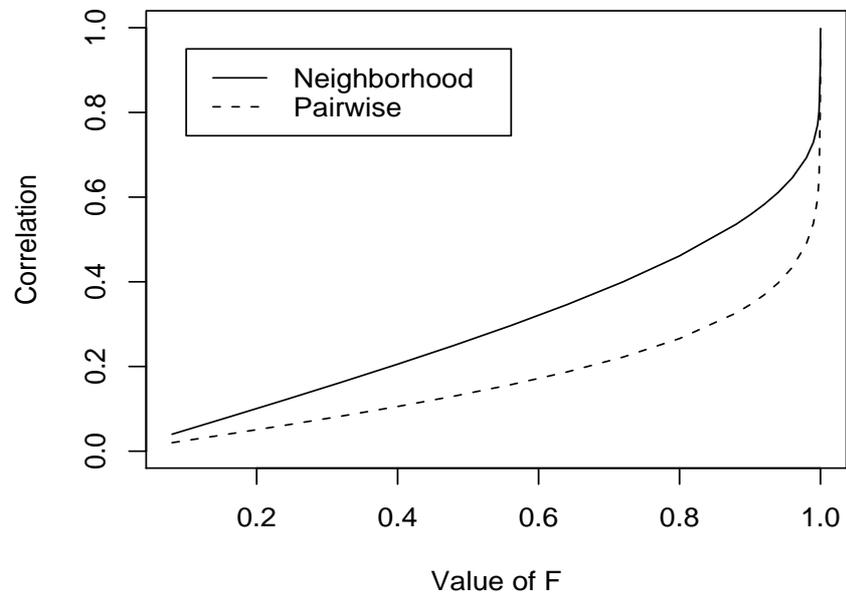
Figures

Figure 1: Plot of two correlation measures against Markov random field F.
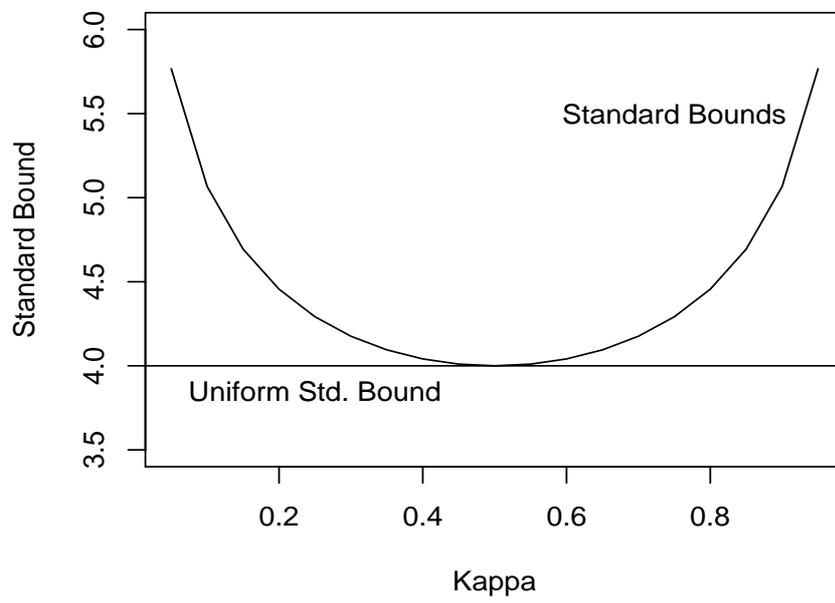
Figure 2: Standard and uniform standard bounds on γ for a simple binary conditionals model.
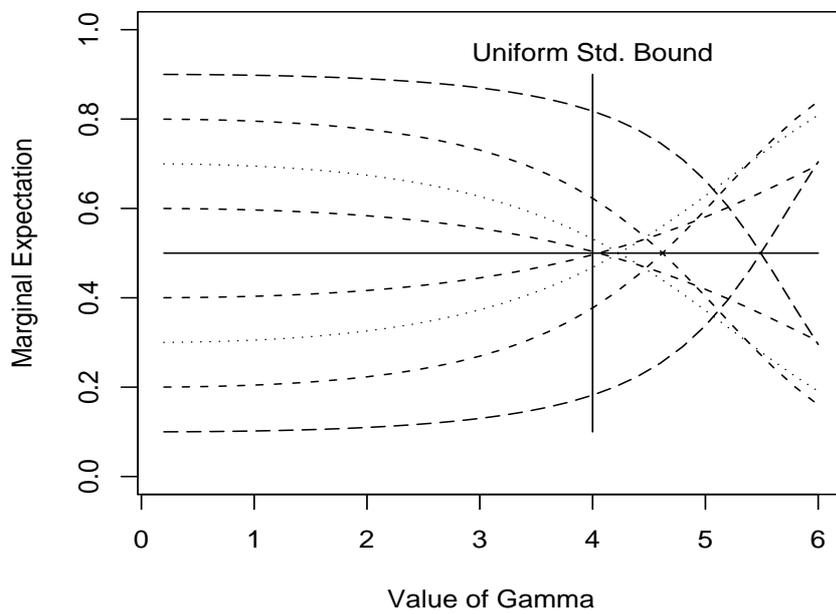
Figure 3: Marginal expected values versus values of γ for simple models with binary conditionals. Lines represent, from top to bottom, $\kappa = 0.90, 0.80, \ldots, 0.10$, and the solid vertical line demarcates the uniform standard bound of $\gamma = 4.0$.
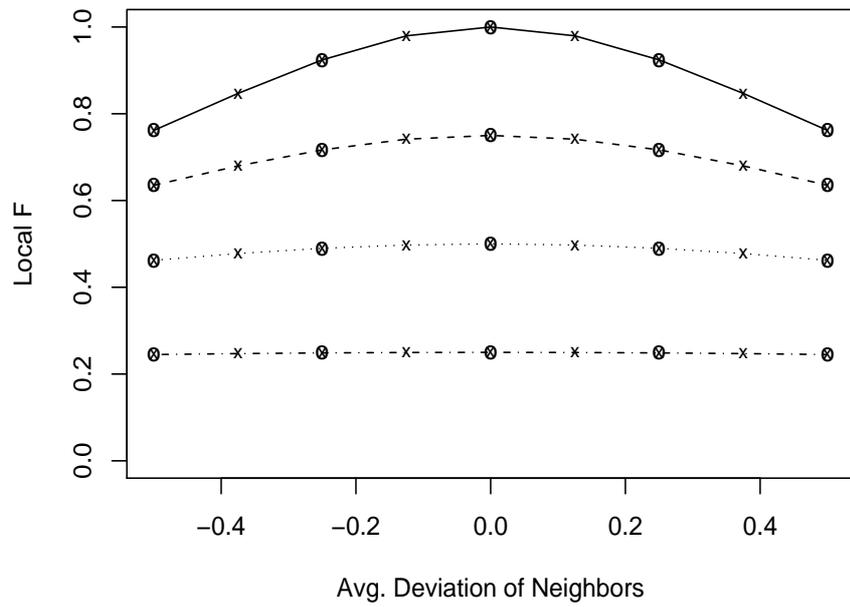
Figure 4: Relation of local dependence measures F_i to the average deviation of neighboring values from $\kappa = 0.5$. The top line is for $\gamma = 4.0$, the standard and uniform standard bound. Other lines correspond to, in descending order, $\gamma = 3.0$, $\gamma = 2.0$, and $\gamma = 1.0$. Symbols given as 'x' denote values for an eight-nearest neighbors model while those given as 'o' denote a four-nearest neighbors model.
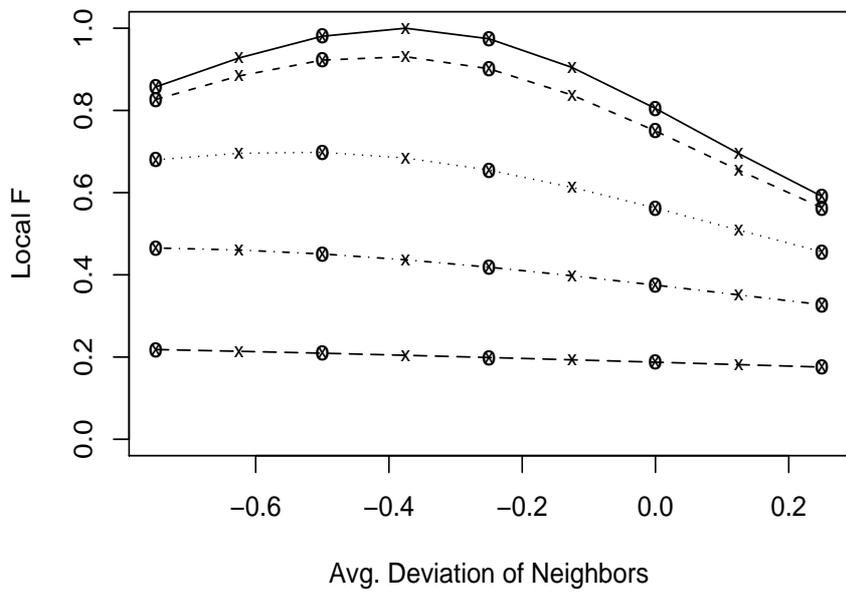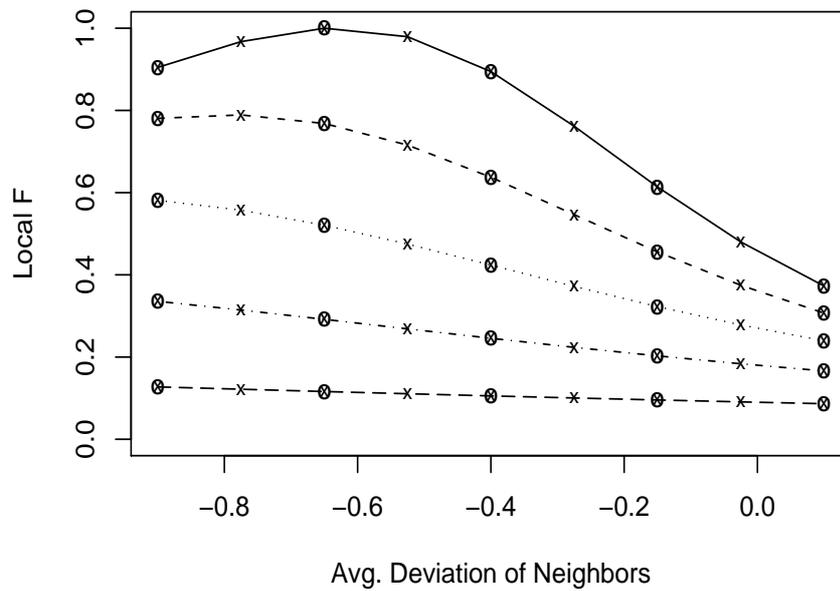
Figure 5: Relation of local dependence measures F_i to the average deviation of neighboring values from $\kappa = 0.75$. The top line is for $\gamma = 4.2918$, the standard bound for this κ. Other lines correspond to, in descending order, $\gamma = 4.0$, $\gamma = 3.0$, $\gamma = 2.0$, and $\gamma = 1.0$. Symbols given as 'x' denote values for an eight-nearest neighbors model while those given as 'o' denote a four-nearest neighbors model.
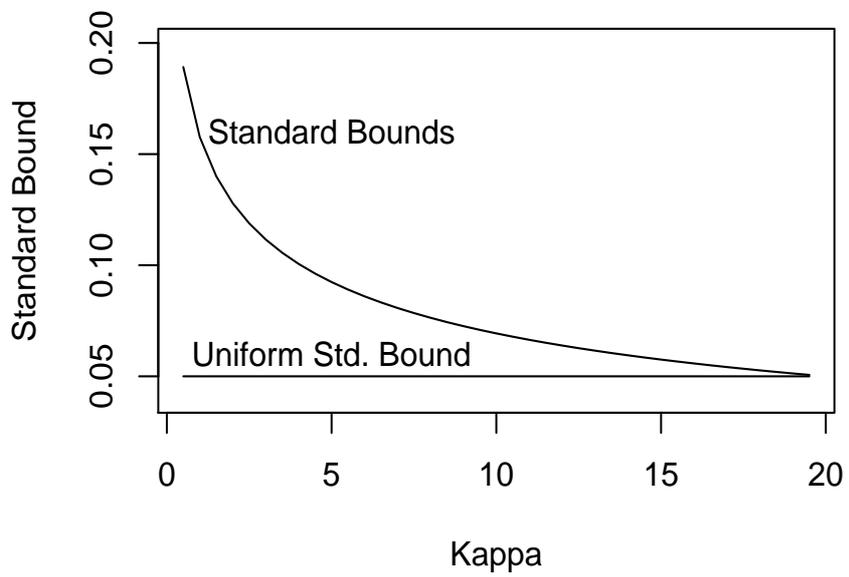
Figure 6: Relation of local dependence measures F_i to the average deviation of neighboring values from $\kappa = 0.9$. The top line is for $\gamma = 5.0705$, the standard bound for this κ. Other lines correspond to, in descending order, $\gamma = 4.0$, $\gamma = 3.0$, $\gamma = 2.0$, and $\gamma = 1.0$. Symbols given as 'x' denote values for an eight-nearest neighbors model while those given as 'o' denote a four-nearest neighbors model.

Figure 7: Standard and uniform standard bounds for the dependence parameter γ from a Winsorized Poisson model with Winsorization value $R = 20$.
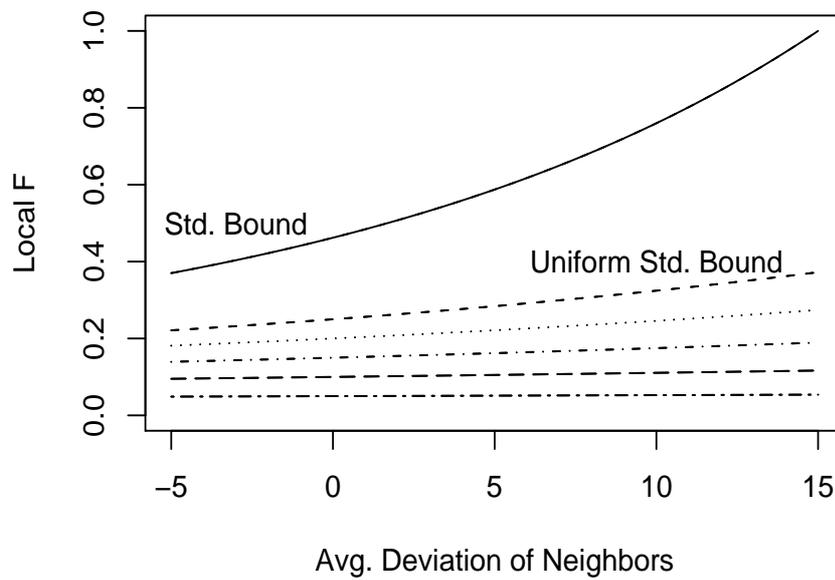
Figure 8: Relation of local dependence measures F_i to the average deviation of neighboring values from $\kappa = 5$ in a Winsorized Poisson model with $R = 20$. The top curve is for the standard bound $\gamma = 0.09242$, the next lower curve for the uniform standard bound $\gamma = 0.05$, and other curves for $\gamma = 0.04, 0.03, 0.02, 0.01$ in decreasing order.
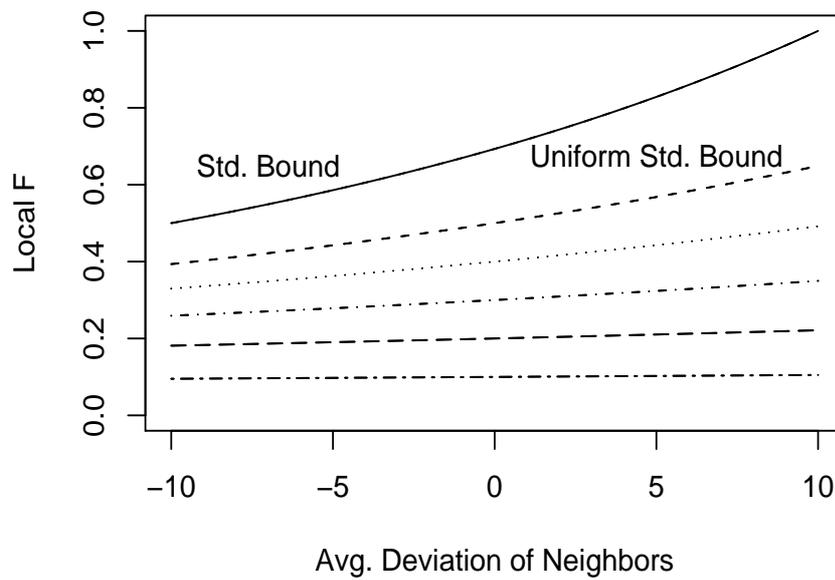
Figure 9: Relation of local dependence measures F_i to the average deviation of neighboring values from $\kappa = 10$ in a Winsorized Poisson model with $R = 20$. The top curve is for the standard bound $\gamma = 0.09242$, the next lower curve for the uniform standard bound $\gamma = 0.05$, and other curves for $\gamma = 0.04, 0.03, 0.02, 0.01$ in decreasing order.