Brazilian Journal of Probability and Statistics
2009, Vol. 23, No. 2, 120–140
DOI: 10.1214/08-BJPS018

Pseudo-likelihood equations for Potts model on higher-order neighborhood systems: A quantitative approach for parameter estimation in image analysis

Alexandre L. M. Levada[a], Nelson D. A. Mascarenhas[b] and Alberto Tannús[a]

[a]*Universidade de São Paulo*
[b]*Universidade Federal de São Carlos*

Abstract. This paper presents analytical pseudo-likelihood (PL) equations for Potts Markov random field (MRF) model parameter estimation on higher-order neighborhood systems by expanding the derivative of the log-PL function based on the enumeration of all possible contextual configuration patterns given a neighborhood system. The proposed equations allow the modeling of less restrictive neighborhood systems in a large number of MRF applications in a computationally feasible way. To evaluate the proposed estimation method we propose a hypothesis testing approach, derived by approximating the asymptotic variance of MPL parameter estimators using the observed Fisher information. The definition of the asymptotic variance, together with the test size α and p-values, provide a complete framework for quantitative analysis. Experiments with synthetic images generated by Markov chain Monte Carlo simulation methods assess the accuracy of the proposed estimation method, indicating that higher-order neighborhood systems reduce the MPL estimator asymptotic variance and improve estimation performance.

1 Introduction

Since the beginning of statistical physics, scientists were interested in studying systems of particles arranged on a 2-D lattice (Ising (1925); Herisenberg (1928)). Those physical systems of particles were completely characterized by a global energy function (Hamiltonian), through the definition of the joint Boltzmann/Gibbs distribution. However, for a long time, the use of spatial models was restricted to theoretical analysis of those physical entities.

Later, with important advances on probability and statistics, as the Hammersley–Clifford theorem (Hammersley and Clifford (1971)) and the development of Markov chain Monte Carlo simulation (MCMC) (Metropolis et al. (1953); Geman and Geman (1984); Swendsen and Wang (1987); Wolff (1989)) together with relaxation algorithms for combinatorial optimization (Besag (1986); Marroquin, Mitter and Poggio (1987); Yu and Berthod (1995)), Markov random field (MRF)

Key words and phrases. Markov random fields, Potts model, maximum pseudo-likelihood estimation, Markov chain Monte Carlo simulation.

Received April 2008; accepted October 2008.

theory became a central topic with applications in fields including image processing, pattern recognition, computer vision and game theory. Those advances have led to a huge number of novel methodologies, especially in statistical applications regarding contextual modeling and spatial data analysis.

Among the existing MRF models, the Potts model is certainly the most studied and applied one. Basically, the Potts MRF model tries to represent the way individual elements (e.g., atoms, animals, image pixels, etc.) modify their behavior to conform to the behavior of other individuals in their vicinity. It is a model used to study collective effects based on consequences of local interactions. It has a major role in several research areas such as mathematics (Wu (1995); Adams (1994); Ge, Hu and Wang (1996); Jim and Zhang (2004)), physics (Montroll (1941); Enting and Guttmann (2003)), biology (Ouchi et al. (2003); Merks and Glazier (2005)), computer science (Berthod et al. (1996); Li (2001); Won and Gray (2004)) and even sociology (Liu, Luo and Shao (2001)).

However, one of the main difficulties relies exactly on MRF parameter estimation. Traditional methods, as maximum likelihood (ML), cannot be applied due to the existence of the partition function in the joint Gibbs distribution, which is computationally intractable. A solution proposed by Besag (1974) is to use the local conditional density functions (LCDF) to perform maximum pseudo-likelihood (MPL) estimation. Our motivations for employing this approach are:

- MPL estimation is a computationally feasible method.
- From a statistical perspective, MPL estimators have a series of desirable and interesting properties, such as consistency and asymptotic normality (Jensen and Künsh (1994); Winkler (2006)). Thus, it is possible to completely characterize their behavior in the limiting case.

A serious limitation of this approach has been the use of extremely restricted neighborhood systems. Actually, methods for Potts MRF model parameter estimation through MPL often consider only first-order neighborhood systems. A recent result in MRF literature (Frery, Correia and Freitas (2007)), based on a expansion of the log-pseudo-likelihood function on all possible spatial configuration patterns given a neighborhood system, shows a 67 term analytic expression for Potts model MPL estimation on second-order systems, employing a considerably larger number of terms than the result we propose here.

This paper presents an explicit derivation of Potts model pseudo-likelihood equations for higher-order neighborhood systems, more precisely, second and third orders, leading to feasible MRF parameter estimation. The proposed equations allow the representation of less restrictive contextual systems in a large number of MRF applications, such as image restoration and contextual classification. Furthermore, we also propose a hypothesis testing approach to validate the obtained results. Our objective is to propose an approximation to the asymptotic variance of maximum pseudo-likelihood estimators of Potts model parameters, using the observed Fisher information. The definition of test statistics, together with P-values, calculated using our approximation for the asymptotic variance, provide a complete framework for quantitative analysis in Potts MRF model parameter estimation in image processing applications.

The remainder of the paper is organized as follows. Section 2 introduces the Potts MRF model and presents the derived pseudo-likelihood equations for second and third-order neighborhood systems. Section 3 describes asymptotic properties of the MPL estimation and the proposed approximation for the asymptotic variance of the Potts model MPL estimator. The hypothesis testing framework for quantitative analysis of Potts MRF model parameter estimation, the experiments and results are shown on Section 4. Finally, Section 5 presents the conclusion and final remarks.

2 Pseudo-likelihood equations on higher-order neighborhood systems

The fundamental notion associated with Markov property is the conditional independence, since the knowledge of a local region isolates a single element from the entire field. Let Ω denote the integer lattice in the Euclidean $\mathbb{R}^2$ space, where points can be represented by pairs of integers (m, n). The 2-D indexing scheme can be easily transformed to a 1-D scheme by using a lexicographic notation. A MRF defined on Ω is a collection of random variables for which the probability of a given site value given the entire lattice is equal to the probability of the site value given a finite support region of the lattice, called neighborhood (Waks, Tretiak and Gregoriou (1990)).

Traditionally, two groups have developed extensions of 1-D Markov process for 2-D data. The first approach adopts most ideas and tools from statistical mechanics and expresses the Markov nature of a random field in a non-causal way. The other group's primary goal is to extend 1-D hidden Markov models (HMM) to 2-D causal MRF models (Won and Gray (2004)). The chief obstacle for this extension is the lack of a natural ordering for a 2-D grid. As a result, an artificial ordering must be assumed.

Neighborhood systems are characterized by its shape (causal or non-causal) and extension (order). A neighborhood structure is causal if all elements of the neighborhood region of support belong to a half of the plane (asymmetrical), that is, if the field can be reordered into a 1-D random vector that satisfies the Markov property (Zhang, Fieguth and Wang (2000)). Otherwise, in case of symmetrical region of supports regarding the central element, the neighborhood is called non-causal. Non-causal neighborhood systems are referred as zero-order, first-order (4 neighbors), second-order (8 neighbors) and so on. Let N_i^k be the non-causal neighborhood system of order k for the pixel x_i. The finite support regions for pixel x_i from first to fifth-order non-causal neighborhood systems are shown in Figure 1.

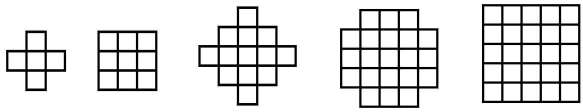

Figure 1 *Finite support regions representing several non-causal neighborhood systems.*

2.1 Bayesian estimation in stochastic image processing

In stochastic image processing, the Bayesian paradigm defines an elegant mathematical framework to reduce the solution space in several ill-posed problems by incorporating prior knowledge in the form of a priori probabilities, $P(x)$. Bayesian estimation consists in choosing the estimate which minimizes the expected cost (Bayes risk) taken with respect to the posterior probability distribution. The most frequently used cost function is the uniform function. It is known that in this case, the optimal Bayes estimate is obtained by the MAP (*Maximum a Posteriori*) criterion. The main advantages of the MAP criterion can be summarized as:

- Since maximizing the posterior probability $P(x|y)$ is equivalent to maximizing $P(y|x)P(x)$, which means, the likelihood function (provided by the observations) and the a priori probability, it is possible to systematically incorporate information about the image formation and prior knowledge.
- The MAP estimate can be understood as a regularization procedure. Often, in image processing applications, maximum likelihood solutions are not reasonable due to the presence of ill-posed behavior.

One of the most widely used prior models is the Potts MRF pairwise interaction model. Two fundamental characteristics of the Potts model considered here are: it is both isotropic and stationary. According to Hammersley and Clifford (1971), the Potts MRF model can be equivalently defined in two manners: by a joint Gibbs distribution (global model) or by a set of local conditional density functions (LCDFs). For a general kth order neighborhood system N_i^k, we define the former by the following expression:

$$P_\beta(X = x) = \frac{1}{Z_\beta} \exp\left\{ \sum_{r,s \in \Omega : s \in N_r^k} \beta[1 - \delta(x_r, x_s)] \right\}, \tag{2.1}$$

where $\delta(x_r, x_s)$ equals 0 if $x_r = x_s$ and 1 if $x_r \neq x_s$. However, for mathematical tractability and computational reasons we will adopt a local description of the probability model through the LCDFs of the Potts pairwise interaction model. The Potts model LCDF for a single observation is (Yamazaki and Gingras (1995)):

$$P_\beta(x_i = m_i | N_i^k) = \frac{\exp\{\beta U_i(m_i)\}}{\sum_{\ell=1}^{M} \exp\{\beta U_i(\ell)\}}, \tag{2.2}$$

where $U_i(\ell)$ is the number of neighbors of the ith element having label equal to ℓ, $\beta \in \mathbb{R}$ is the spatial dependency parameter, also known as inverse temperature, because from statistical mechanics $\beta = 1/k_B T$, where k_B is the Boltzmann constant, and $\ell \in G$, $G = \{1, 2, \ldots, M\}$, where M is the total number of labels.

2.2 Maximum pseudo-likelihood estimation

The main advantage of maximum pseudo-likelihood estimation is its computational simplicity. Fortunately, as the maximum likelihood (ML) estimator, the MPL estimator has also a series of desirable properties, such as consistency and asymptotic normality (Jensen and Künsh (1994)). The pseudo-likelihood function for the Potts MRF model is defined as

$$PL(\beta) = \prod_{s\in\Omega} p(x_s = m_s | N_s^k) = \prod_{s\in\Omega} \frac{\exp\{\beta U_s(m_s)\}}{\sum_{\ell=1}^{M} \exp\{\beta U_s(\ell)\}}. \tag{2.3}$$

Taking logarithms, differentiating on the parameter and setting the result to zero, leads to the following expression, that is the basis for the derivation of the proposed equations:

$$\begin{aligned} \frac{\partial}{\partial\beta} \log PL(\beta) &= \Psi(\beta) \\ &= \sum_{s\in\Omega} U_s(m_s) - \sum_{s\in\Omega} \left[\frac{\sum_{\ell=1}^{M} U_s(\ell) \exp\{\beta U_s(\ell)\}}{\sum_{\ell=1}^{M} \exp\{\beta U_s(\ell)\}} \right] = 0, \end{aligned} \tag{2.4}$$

where m_s denotes the observed value for the sth element of the field.

The objective of this paper is to generalize the estimation of Potts' β parameter for higher-order neighborhood systems in a computationally tractable way, by expanding equation (2.4) based on the number of occurrences of possible contextual configuration patterns, given a neighborhood system. Note that, as the first term of (2.4) is independent of β, we have to expand only the second term.

In first-order neighborhood systems, the enumeration of all possible configuration patterns is straightforward, since there are only five different cases, as shows Figure 2, from zero agreement (4 different labels) to total agreement (4 identical labels). These configurations can be represented by vectors, as presented in

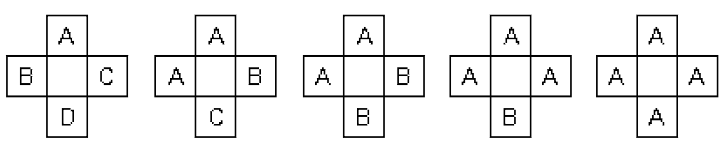

Figure 2 *Contextual configuration patterns for Potts MRF model in first-order neighborhood systems.*

relations (2.5), indicating the number of occurrences of each label around the central element. In the Potts model, location information is irrelevant, since it is an isotropic model:

$$\vec{v}_0 = [1, 1, 1, 1]; \qquad \vec{v}_1 = [2, 1, 1, 0]; \qquad \vec{v}_2 = [2, 2, 0, 0];$$
$$\vec{v}_3 = [3, 1, 0, 0]; \qquad \vec{v}_4 = [4, 0, 0, 0]. \tag{2.5}$$

2.2.1 *Calculating the number of configuration patterns.* Let N be the number of elements in the neighborhood (i.e., $N = 4, 8, 12, \ldots$). For each $L = 1, \ldots, N$, let

$$A_N(L) = \Big\{(a_1, \ldots, a_L) \text{ such that } a_i \in \{1, 2, \ldots, N\}, \; a_1 \le a_2 \le \cdots \le a_L, \sum_{i=1}^{N} a_i = N\Big\} \tag{2.6}$$

and $n_N(L)$ the number of elements of the set $A_N(L)$. Then, the number of possible configuration patterns λ, is given by $\lambda = n_N(1) + \cdots + n_N(L)$.

The solution vectors were found by exhaustive searching, isolating one variable and searching on the subspace spanned by the remainder variables. In order to reduce the computational burden due to high-dimensional vectors (for large neighborhood systems), we introduced a heuristic by restricting the search to the first quadrant of the subspace, since symmetrical vectors lead to identical solutions (i.e., $[7, 1, 0, 0, 0, 0, 0, 0] \equiv [1, 7, 0, 0, 0, 0, 0, 0]$). Table 1 presents the number of contextual configuration patterns, λ, and the elapse time for their generation for several neighborhood system orders, more precisely from first to fifth orders. Table 2 shows the solution vectors representing the possible configuration patterns on second-order neighborhood systems.

Given the complete set of contextual configuration patterns for a neighborhood system, it is possible to expand the second term of equation (2.4). We

Table 1 *Number of strategy configuration patterns for five different neighborhood systems*

Neighborhood system	Number of configuration patterns (λ)	Elapsed time (sec.)[1]
First order	5	0.01
Second order	22	0.21
Third order	77	4.47
Fourth order	637	29.31
Fifth order	1575	1517.24

[1] All experiments were executed in an Athlon X2 Dual Core 2.21Ghz processor with 2GB of RAM.

Table 2 *Solution vectors representing all contextual configuration patterns on second-order neighborhood systems*

[1, 1, 1, 1, 1, 1, 1, 1]	[2, 1, 1, 1, 1, 1, 1, 0]	[3, 1, 1, 1, 1, 1, 0, 0]	[2, 2, 1, 1, 1, 1, 0, 0]
[4, 1, 1, 1, 1, 0, 0, 0]	[3, 2, 1, 1, 1, 0, 0, 0]	[2, 2, 2, 1, 1, 0, 0, 0]	[5, 1, 1, 1, 0, 0, 0, 0]
[4, 2, 1, 1, 0, 0, 0, 0]	[3, 3, 1, 1, 0, 0, 0, 0]	[3, 2, 2, 1, 0, 0, 0, 0]	[2, 2, 2, 2, 0, 0, 0, 0]
[6, 1, 1, 0, 0, 0, 0, 0]	[5, 2, 1, 0, 0, 0, 0, 0]	[4, 3, 1, 0, 0, 0, 0, 0]	[4, 2, 2, 0, 0, 0, 0, 0]
[3, 3, 2, 0, 0, 0, 0, 0]	[4, 4, 0, 0, 0, 0, 0, 0]	[5, 3, 0, 0, 0, 0, 0, 0]	[6, 2, 0, 0, 0, 0, 0, 0]
	[7, 1, 0, 0, 0, 0, 0, 0]	[8, 0, 0, 0, 0, 0, 0, 0]	

can regard the numerator as a simple inner product of two vectors $\vec{U_s}$ and $\vec{w}_s$, where $\vec{U_s}$ represents the contextual configuration vector for the current pixel (i.e., $\vec{U_s} = [5, 2, 1, 0, 0, 0, 0, 0]$ in case of a second-order neighborhood system) and $\vec{w}_s$ is a vector such that $w_s[n] = \exp\{\beta U_s[n]\}$. Similarly, the denominator is the inner product of $\vec{w}_s$ with the identity column vector $\vec{r} = [1, 1, \ldots, 1]$. Thus, the second term of equation (2.4) can be expanded as a summation of λ terms, each one associated with a possible configuration pattern. However, as it involves the sum on all elements of the MRF, we define constants K_i, $i = 1, 2, \ldots, \lambda$, representing the number of occurrences of each possible configuration pattern along the entire field. The basic idea is that the set of K_i coefficients defines a contextual histogram, that is, instead of indicating the distribution of individual pixel gray levels, this set shows the distribution of spatial patterns, defined in terms of the neighborhood system order, along the random field. For instance, in image analysis applications, smooth images, with many homogeneous regions, tend to present little variation in these contextual patterns because of the high correlation between neighboring pixels, while noisy images tend to present more variability on these spatial configuration patterns.

It is worthwhile noting that symmetrical configuration patterns offer the same contribution to the pseudo-likelihood equation, since the inner product between two vectors does not depend on the order of the elements. In other words, what effectively contributes for the pseudo-likelihood equation is only the configuration of the neighboring pixels regardless the central pixel value. The complete analytical expression for a second-order system is given by equation (2.7). Note that the resulting equation is transcendental, which means that it has no closed-form solution. Note also that in the case of a reduced number of labels, the equation is further simplified, since many K_i coefficients, $i = 1, 2, \ldots, \lambda$, are zero, simply because many contextual configuration patterns are impossible. For instance, in the case of only two possible labels (0–1 spins, the Ising model), equation (2.7) only has five terms instead of the 22 original ones. Thus, a reduction on the number of possible individual labels M represents a constraint in the total number of contextual configuration patterns, decreasing the computational cost of Potts MRF model parameter estimation. Note that in a physical interpretation, we are using the proposed equations to estimate a quantity called *inverse temperature* in a system of particles arranged on a 2-D lattice, using only pairwise interactions:

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\log PL(\beta)\\
&\quad=\sum_{s\in\Omega}U_s(m_s)-\frac{8e^{8\hat{\beta}}}{e^{8\hat{\beta}}+M-1}K_1-\frac{7e^{7\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{\hat{\beta}}+M-2}K_2\\
&\qquad-\frac{6e^{6\hat{\beta}}+2e^{2\hat{\beta}}}{e^{6\hat{\beta}}+e^{2\hat{\beta}}+M-2}K_3-\frac{6e^{6\hat{\beta}}+2e^{\hat{\beta}}}{e^{6\hat{\beta}}+2e^{\hat{\beta}}+M-3}K_4\\
&\qquad-\frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+M-2}K_5-\frac{5e^{5\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-3}K_6\\
&\qquad-\frac{5e^{5\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+3e^{\hat{\beta}}+M-4}K_7-\frac{8e^{4\hat{\beta}}}{2e^{4\hat{\beta}}+M-2}K_8\\
&\qquad-\frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-3}K_9-\frac{4e^{4\hat{\beta}}+4e^{2\hat{\beta}}}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{10}\\
&\qquad-\frac{4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{11}-\frac{4e^{4\hat{\beta}}+4e^{\hat{\beta}}}{e^{4\hat{\beta}}+4e^{\hat{\beta}}+M-5}K_{12}\\
&\qquad-\frac{6e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{13}-\frac{6e^{3\hat{\beta}}+2e^{\hat{\beta}}}{2e^{3\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{14}\\
&\qquad-\frac{3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{15}-\frac{3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{16}\\
&\qquad-\frac{3e^{3\hat{\beta}}+5e^{\hat{\beta}}}{e^{3\hat{\beta}}+5e^{\hat{\beta}}+M-6}K_{17}-\frac{8e^{2\hat{\beta}}}{4e^{2\hat{\beta}}+M-4}K_{18}\\
&\qquad-\frac{6e^{2\hat{\beta}}+2e^{\hat{\beta}}}{3e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{19}-\frac{4e^{2\hat{\beta}}+4e^{\hat{\beta}}}{2e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{20}\\
&\qquad-\frac{2e^{2\hat{\beta}}+6e^{\hat{\beta}}}{e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-7}K_{21}-\frac{8e^{\hat{\beta}}}{8e^{\hat{\beta}}+M-8}K_{22}=0.
\end{aligned}
\tag{2.7}
$$

Similarly, the pseudo-likelihood equation for third-order neighborhood systems is obtained by expanding equation (2.4) on the seventy-seven configuration patterns obtained by solving the equations generated by equation (2.6) for $N=12$. The complete expression for the pseudo-likelihood equation for the Potts model using third-order neighborhood systems is given by equation (3.4).

In all the experiments along this paper, the MPL estimator is obtained by finding the zero of the derived pseudo-likelihood equation using a numerical method.

We chose Brent's method (Brent (1973)), a numerical algorithm that does not require the computation (or even the existence) of derivatives or analytical gradients. In this case, the computation of derivatives of the objective function would be prohibitive, given the large extension of the expressions. The advantages of this method can be summarized by: it uses a combination of bisection, secant and inverse quadratic interpolation methods, leading to a very robust approach and also it has superlinear convergence rate.

3 On the asymptotic variance of Potts MRF MPL estimator

Unbiasedness is not granted by either ML or MPL estimation. Actually, there is no method that guarantees the existence of unbiased estimators for a fixed N-size sample. Often, in the exponential family, MLE coincide with UMVU (*Uniform Minimum Variance Unbiased*) estimators because they are functions of complete sufficient statistics (if MLE is unique, then it is a function of sufficient statistics). Also, there are several characteristics that make ML estimation a reference method (Lehmann (1983); Bickel (1991); Casella and Berger (2002)). Making the sample size grow infinitely ($N \to \infty$), MLE becomes asymptotically unbiased and efficient. Unfortunately, there is no result showing that the same occurs in MPL estimation. In this section, we propose an approximation for the asymptotic variance of Potts MRF model MPL estimator in terms of expressions for observed Fisher information using both first and second derivatives.

3.1 Observed Fisher information

Often, in practice, it is not possible to calculate the expected Fisher information, $I(\beta)$. In such cases, we can adopt the observed Fisher information, $I_{\text{obs}}(\beta)$ instead. Furthermore, it has been shown (Efron and Hinkley (1978)) that the use of the observed information number is superior to the expected information number, as it appears in the Cramér–Rao lower bound. The observed Fisher information, in terms of the pseudo-likelihood function, is defined by

$$I_{\text{obs}}(\beta) = \left[\frac{\partial}{\partial\beta} \log PL(X;\beta)\right]^2, \tag{3.1}$$

and can be estimated by the following, justified by the law of large numbers:

$$\hat{I}^1_{\text{obs}}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\partial\beta} \log p(x_i;\beta)\right]^2\bigg|_{\beta=\hat{\beta}}, \tag{3.2}$$

since $I(\beta) = E[\hat{I}^1_{\text{obs}}(\beta)]$, making $\hat{I}^1_{\text{obs}}(\beta) \approx I(\beta)$. Similarly, $I_{\text{obs}}(\beta)$ can be estimated using the second derivative of the likelihood function

$$\hat{I}^2_{\text{obs}}(\beta) = -\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial^2}{\partial\beta^2} \log p(x_i;\beta)\right]\bigg|_{\beta=\hat{\beta}}, \tag{3.3}$$

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\log PL(\beta)\\
&\quad=\sum_{s\in\Omega}U_s(m_s)-\frac{12e^{12\hat{\beta}}}{e^{12\hat{\beta}}+M-1}K_1-\frac{11e^{11\hat{\beta}}+e^{\hat{\beta}}}{e^{11\hat{\beta}}+e^{\hat{\beta}}+M-2}K_2\\
&\qquad-\frac{10e^{10\hat{\beta}}+2e^{2\hat{\beta}}}{e^{10\hat{\beta}}+e^{2\hat{\beta}}+M-2}K_3-\frac{9e^{9\hat{\beta}}+3e^{3\hat{\beta}}}{e^{9\hat{\beta}}+e^{3\hat{\beta}}+M-2}K_4\\
&\qquad-\frac{8e^{8\hat{\beta}}+4e^{4\hat{\beta}}}{e^{8\hat{\beta}}+e^{4\hat{\beta}}+M-2}K_5-\frac{7e^{7\hat{\beta}}+5e^{5\hat{\beta}}}{e^{7\hat{\beta}}+e^{5\hat{\beta}}+M-2}K_6\\
&\qquad-\frac{12e^{6\hat{\beta}}}{2e^{6\hat{\beta}}+M-2}K_7-\frac{12e^{4\hat{\beta}}}{3e^{4\hat{\beta}}+M-3}K_8\\
&\qquad-\frac{5e^{5\hat{\beta}}+4e^{4\hat{\beta}}+3e^{3\hat{\beta}}}{e^{5\hat{\beta}}+e^{4\hat{\beta}}+e^{3\hat{\beta}}+M-3}K_9-\frac{10e^{5\hat{\beta}}+2e^{2\hat{\beta}}}{2e^{5\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{10}\\
&\qquad-\frac{6e^{6\hat{\beta}}+6e^{3\hat{\beta}}}{e^{6\hat{\beta}}+2e^{3\hat{\beta}}+M-3}K_{11}\\
&\qquad-\frac{6e^{6\hat{\beta}}+4e^{4\hat{\beta}}+2e^{2\hat{\beta}}}{e^{6\hat{\beta}}+e^{4\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{12}\frac{6e^{6\hat{\beta}}+5e^{5\hat{\beta}}+e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{5\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{13}\\
&\qquad-\frac{7e^{7\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{e^{7\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{14}\\
&\qquad-\frac{7e^{7\hat{\beta}}+4e^{4\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{4\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{15}-\frac{8e^{8\hat{\beta}}+4e^{2\hat{\beta}}}{e^{8\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{16}\\
&\qquad-\frac{8e^{8\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{8\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{17}\\
&\qquad-\frac{9e^{9\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{9\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{18}\frac{10e^{10\hat{\beta}}+2e^{\hat{\beta}}}{e^{10\hat{\beta}}+2e^{\hat{\beta}}+M-3}K_{19}\\
&\qquad-\frac{12e^{3\hat{\beta}}}{4e^{3\hat{\beta}}+M-4}K_{20}-\frac{4e^{4\hat{\beta}}+6e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{e^{4\hat{\beta}}+2e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-4}K_{21}\\
&\qquad-\frac{8e^{4\hat{\beta}}+4e^{2\hat{\beta}}}{2e^{4\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{22}\frac{8e^{4\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{2e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{23}\\
&\qquad-\frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}+4e^{2\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+2e^{2\hat{\beta}}+M-4}K_{24}
\end{aligned}
\tag{3.4}
$$

$$
\begin{aligned}
&- \frac{5e^{5\hat{\beta}} + 6e^{3\hat{\beta}} + e^{\hat{\beta}}}{e^{5\hat{\beta}} + 2e^{3\hat{\beta}} + e^{\hat{\beta}} + M - 4} K_{25} - \frac{5e^{5\hat{\beta}} + 4e^{4\hat{\beta}} + 2e^{2\hat{\beta}} + e^{\hat{\beta}}}{e^{5\hat{\beta}} + e^{4\hat{\beta}} + e^{2\hat{\beta} + e^{\hat{\beta}} + M - 4}} K_{26} \\
&- \frac{10e^{5\hat{\beta}} + 2e^{\hat{\beta}}}{2e^{5\hat{\beta}} + 2e^{\hat{\beta}} + M - 4} K_{27} - \frac{6e^{6\hat{\beta}} + 6e^{2\hat{\beta}}}{e^{6\hat{\beta}} + 3e^{2\hat{\beta}} + M - 4} K_{28} \\
&- \frac{6e^{6\hat{\beta}} + 3e^{3\hat{\beta}} + 2e^{2\hat{\beta}} + e^{\hat{\beta}}}{e^{6\hat{\beta}} + e^{3\hat{\beta}} + e^{2\hat{\beta} + e^{\hat{\beta}} + M - 4}} K_{29} - \frac{6e^{6\hat{\beta}} + 4e^{4\hat{\beta}} + 2e^{\hat{\beta}}}{e^{6\hat{\beta}} + e^{4\hat{\beta}} + 2e^{\hat{\beta}} + M - 4} K_{30} \\
&- \frac{7e^{7\hat{\beta}} + 4e^{2\hat{\beta}} + e^{\hat{\beta}}}{e^{7\hat{\beta}} + 2e^{2\hat{\beta}} + e^{\hat{\beta}} + M - 4} K_{31} - \frac{7e^{7\hat{\beta}} + 3e^{3\hat{\beta}} + 2e^{\hat{\beta}}}{e^{7\hat{\beta}} + e^{3\hat{\beta}} + 2e^{\hat{\beta}} + M - 4} K_{32} \\
&- \frac{8e^{8\hat{\beta}} + 2e^{2\hat{\beta}} + 2e^{\hat{\beta}}}{e^{8\hat{\beta}} + e^{2\hat{\beta}} + 2e^{\hat{\beta}} + M - 4} K_{33} - \frac{9e^{9\hat{\beta}} + 3e^{\hat{\beta}}}{e^{9\hat{\beta}} + 3e^{\hat{\beta}} + M - 4} K_{34} \\
&- \frac{6e^{3\hat{\beta}} + 6e^{2\hat{\beta}}}{2e^{3\hat{\beta}} + 3e^{2\hat{\beta}} + M - 5} K_{35} - \frac{9e^{3\hat{\beta}} + 2e^{2\hat{\beta}} + e^{\hat{\beta}}}{3e^{3\hat{\beta}} + e^{2\hat{\beta}} + e^{\hat{\beta}} + M - 5} K_{36} \\
&- \frac{4e^{4\hat{\beta}} + 8e^{2\hat{\beta}}}{e^{4\hat{\beta}} + 4e^{2\hat{\beta}} + M - 5} K_{37} - \frac{4e^{4\hat{\beta}} + 3e^{3\hat{\beta}} + 4e^{2\hat{\beta}} + e^{\hat{\beta}}}{e^{4\hat{\beta}} + e^{3\hat{\beta}} + 2e^{2\hat{\beta} + e^{\hat{\beta}} + M - 5}} K_{38} \\
&- \frac{4e^{4\hat{\beta}} + 6e^{3\hat{\beta}} + 2e^{\hat{\beta}}}{e^{4\hat{\beta}} + 2e^{3\hat{\beta}} + 2e^{\hat{\beta}} + M - 5} K_{39} - \frac{8e^{4\hat{\beta}} + 2e^{2\hat{\beta}} + 2e^{\hat{\beta}}}{2e^{4\hat{\beta}} + e^{2\hat{\beta}} + 2e^{\hat{\beta}} + M - 5} K_{40} \\
&- \frac{5e^{5\hat{\beta}} + 6e^{2\hat{\beta}} + e^{\hat{\beta}}}{e^{5\hat{\beta}} + 3e^{2\hat{\beta}} + e^{\hat{\beta}} + M - 5} K_{41} - \frac{5e^{5\hat{\beta}} + 3e^{3\hat{\beta}} + 2e^{2\hat{\beta}} + 2e^{\hat{\beta}}}{e^{5\hat{\beta}} + e^{3\hat{\beta}} + e^{2\hat{\beta} + 2e^{\hat{\beta}} + M - 5}} K_{42} \\
&- \frac{5e^{5\hat{\beta}} + 4e^{4\hat{\beta}} + 3e^{\hat{\beta}}}{e^{5\hat{\beta}} + e^{4\hat{\beta}} + 3e^{\hat{\beta}} + M - 5} K_{43} - \frac{6e^{6\hat{\beta}} + 4e^{4\hat{\beta}} + 2e^{\hat{\beta}}}{e^{6\hat{\beta}} + 2e^{2\hat{\beta}} + 2e^{\hat{\beta}} + M - 5} K_{44} \\
&- \frac{6e^{6\hat{\beta}} + 3e^{3\hat{\beta}} + 3e^{\hat{\beta}}}{e^{6\hat{\beta}} + e^{3\hat{\beta}} + 3e^{\hat{\beta}} + M - 5} K_{45} \frac{7e^{7\hat{\beta}} + 2e^{2\hat{\beta}} + 3e^{\hat{\beta}}}{e^{7\hat{\beta}} + e^{2\hat{\beta}} + 3e^{\hat{\beta}} + M - 5} K_{46} \\
&- \frac{8e^{8\hat{\beta}} + 4e^{\hat{\beta}}}{e^{8\hat{\beta}} + 4e^{\hat{\beta}} + M - 5} K_{47} - \frac{12e^{2\hat{\beta}}}{6e^{2\hat{\beta}} + M - 6} K_{48} \\
&- \frac{3e^{3\hat{\beta}} + 8e^{2\hat{\beta}} + e^{\hat{\beta}}}{e^{3\hat{\beta}} + 4e^{2\hat{\beta}} + e^{\hat{\beta}} + M - 6} K_{49} - \frac{6e^{3\hat{\beta}} + 4e^{2\hat{\beta}} + 2e^{\hat{\beta}}}{2e^{3\hat{\beta}} + 2e^{2\hat{\beta}} + 2e^{\hat{\beta}} + M - 6} K_{50} \\
&- \frac{9e^{3\hat{\beta}} + 3e^{\hat{\beta}}}{3e^{3\hat{\beta}} + 3e^{\hat{\beta}} + M - 6} K_{51} - \frac{4e^{4\hat{\beta}} + 6e^{2\hat{\beta}} + 2e^{\hat{\beta}}}{e^{4\hat{\beta}} + 3e^{2\hat{\beta}} + 2e^{\hat{\beta}} + M - 6} K_{52}
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}+3e^{\hat{\beta}}+M-6}}K_{53}-\frac{8e^{4\hat{\beta}}+4e^{\hat{\beta}}}{2e^{4\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{54}\\
&-\frac{5e^{5\hat{\beta}}+4e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-6}K_{55}-\frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}+4e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{56}\\
&-\frac{6e^{6\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{57}-\frac{7e^{7\hat{\beta}}+5e^{\hat{\beta}}}{e^{7\hat{\beta}}+5e^{\hat{\beta}}+M-6}K_{58}\\
&-\frac{10e^{2\hat{\beta}}+2e^{\hat{\beta}}}{5e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-7}K_{59}-\frac{3e^{3\hat{\beta}}+6e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{3\hat{\beta}}+3e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-7}K_{60}\\
&-\frac{6e^{3\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-7}K_{61}-\frac{4e^{4\hat{\beta}}+4e^{2\hat{\beta}}+4e^{\hat{\beta}}}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-7}K_{62}\\
&-\frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+5e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+5e^{\hat{\beta}}+M-7}K_{63}-\frac{5e^{5\hat{\beta}}+2e^{2\hat{\beta}}+5e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+5e^{\hat{\beta}}+M-7}K_{64}\\
&-\frac{6e^{6\hat{\beta}}+6e^{\hat{\beta}}}{e^{6\hat{\beta}}+6e^{\hat{\beta}}+M-7}K_{65}-\frac{8e^{2\hat{\beta}}+4e^{\hat{\beta}}}{4e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-8}K_{66}\\
&-\frac{3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+5e^{\hat{\beta}}}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+5e^{\hat{\beta}}+M-8}K_{67}-\frac{6e^{3\hat{\beta}}+6e^{\hat{\beta}}}{2e^{3\hat{\beta}}+6e^{\hat{\beta}}+M-8}K_{68}\\
&-\frac{4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+6e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-8}K_{69}-\frac{5e^{5\hat{\beta}}+7e^{\hat{\beta}}}{e^{5\hat{\beta}}+7e^{\hat{\beta}}+M-8}K_{70}\\
&-\frac{6e^{2\hat{\beta}}+6e^{\hat{\beta}}}{3e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-9}K_{71}-\frac{3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+7e^{\hat{\beta}}}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+7e^{\hat{\beta}}+M-9}K_{72}\\
&-\frac{4e^{4\hat{\beta}}+8e^{\hat{\beta}}}{e^{4\hat{\beta}}+8e^{\hat{\beta}}+M-9}K_{73}-\frac{4e^{2\hat{\beta}}+8e^{\hat{\beta}}}{2e^{2\hat{\beta}}+8e^{\hat{\beta}}+M-10}K_{74}\\
&-\frac{3e^{3\hat{\beta}}+9e^{\hat{\beta}}}{e^{3\hat{\beta}}+9e^{\hat{\beta}}+M-10}K_{75}-\frac{2e^{2\hat{\beta}}+10e^{\hat{\beta}}}{e^{2\hat{\beta}}+10e^{\hat{\beta}}+M-11}K_{76}\\
&-\frac{12e^{\hat{\beta}}}{12e^{\hat{\beta}}+M-12}K_{77}=0.
\end{aligned}
$$

3.2 Asymptotic variance of MPL estimators

Asymptotic evaluations uncover the most fundamental properties of a mathematical procedure, providing a powerful and general tool for statistical analysis. From statistical inference theory it is known that both MLE and MPLE share two important properties: consistency and asymptotic normality. However, differently from the ML estimator, there is no result showing the asymptotic efficiency of MPL estimators. In this section, we obtain an expression for the asymptotic variance of Potts MPL parameter estimators as the ratio of expected Fisher information calculated using first and second derivatives of the pseudo-likelihood function. The starting point is to use the expression for the asymptotic covariance matrix of MPL estimators defined in (Liang and Yu (2003)):

$$C(\vec{\beta}) = H(\vec{\beta})^{-1} K(\vec{\beta}) H(\vec{\beta})^{-1}, \tag{3.5}$$

where the matrices $H(\vec{\beta})$ and $K(\vec{\beta})$ are defined as

$$H(\vec{\beta}) = E_\beta[\nabla^2 F(X;\vec{\beta})] \quad \text{and} \tag{3.6}$$

$$K(\vec{\beta}) = \mathrm{Var}_\beta[\nabla F(X;\vec{\beta})], \tag{3.7}$$

with $F(X;\vec{\beta})$ denoting the logarithm of the pseudo-likelihood function. Since β is a scalar in the Potts MRF model, we can simplify expressions (3.6) and (3.7) and define the asymptotic variance as

$$C(\beta) = V_N(\beta) = \frac{\mathrm{Var}_\beta[\frac{\partial}{\partial\beta}\log L(X;\beta)]}{E_\beta^2[\frac{\partial^2}{\partial\beta^2}\log L(X;\beta)]}. \tag{3.8}$$

By applying the definition of variance and using the observed Fisher information to approximate the expected values, equation (3.8) is further simplified to

$$V_N(\beta) = \frac{E_\beta[(\frac{\partial}{\partial\beta}\log L(X;\beta))^2]}{E_\beta^2[\frac{\partial^2}{\partial\beta^2}\log L(X;\beta)]}, \tag{3.9}$$

since

$$\begin{aligned} E_\beta\left[\frac{\partial}{\partial\beta}\log L(X;\beta)\right] &\approx \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\beta}\log p(x_i|N_i^k,\beta)\bigg|_{\beta=\hat{\beta}} \\ &= \frac{1}{N}\frac{\partial}{\partial\beta}\log\prod_{i=1}^{N} p(x_i|N_i^k,\beta)\bigg|_{\beta=\hat{\beta}} \\ &= \frac{1}{N}\frac{\partial}{\partial\beta}\log L(X;\beta)\bigg|_{\beta=\hat{\beta}} = 0. \end{aligned} \tag{3.10}$$

3.3 Approximating the asymptotic variance

In this section we derive expressions for observed Fisher information using both first and second derivatives of the logarithm of the pseudo-likelihood function in order to compute an approximation to the asymptotic variance, given by equation (3.9).

3.3.1 *Observed Fisher information using the first derivative.* From the LCDF of the Potts MRF model

$$\frac{\partial}{\partial\beta}\log p_\beta(x_i|N_i^k) = \left\{U_i(m_i) - \frac{\sum_{\ell=1}^{M} U_i(\ell)e^{\beta U_i(\ell)}}{\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}}\right\}. \tag{3.11}$$

Thus, according to equation (3.2) we have the following approximation for the numerator of the asymptotic variance:

$$E_\beta\left[\left(\frac{\partial}{\partial\beta}\log L(X;\beta)\right)^2\right] \approx \frac{1}{N}\sum_{i=1}^{N}\left\{\left[U_i(m_i) - \frac{\sum_{\ell=1}^{M} U_i(\ell)e^{\beta U_i(\ell)}}{\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}}\right]^2\right\}, \tag{3.12}$$

which, after some few algebraic manipulations, becomes

$$\begin{aligned} E_\beta\left[\left(\frac{\partial}{\partial\beta}\log L(X;\beta)\right)^2\right] &\approx \hat{I}^1_{\text{obs}}(\hat{\beta}_{\text{MPL}}) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{[\sum_{\ell=1}^{M}(U_i(m_i) - U_i(\ell))e^{\beta U_i(\ell)}]^2}{[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}]^2}\right\}. \end{aligned} \tag{3.13}$$

3.3.2 *Observed Fisher information using the second derivative.* Deriving equation (3.11) again in β leads to another approximation to the observed Fisher information

$$\begin{aligned} -E_\beta\left[\frac{\partial^2}{\partial\beta^2}\log L(X;\beta)\right] \approx \frac{1}{N}\sum_{i=1}^{N}\Bigg\{&\frac{[\sum_{\ell=1}^{M} U_i(\ell)^2 e^{\beta U_i(l)}][\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}]}{[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}]^2} \\ &- \frac{[\sum_{\ell=1}^{M} U_i(\ell)e^{\beta U_i(\ell)}]^2}{[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}]^2}\Bigg\}. \end{aligned} \tag{3.14}$$

In order to compare both expressions for Fisher information, we rewrite the above expression. After algebraic manipulations we show that equation (3.14) reduces to (see the Appendix for details)

$$\begin{aligned} &-E_\beta\left[\frac{\partial^2}{\partial\beta^2}\log L(X;\beta)\right] \\ &\quad\approx \hat{I}^2_{\text{obs}}(\hat{\beta}_{\text{MPL}}) \\ &\quad= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\sum_{\ell=1}^{M-1}[\sum_{k=\ell+1}^{M}(U_i(\ell) - U_i(k))^2 e^{\beta(U_i(\ell)+U_i(k))}]}{[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}]^2}\right\}. \end{aligned} \tag{3.15}$$

The proposed approximation allows the calculation of the asymptotic variance of the maximum pseudo-likelihood estimator of the Potts MRF model. From previous sections we have seen that the sequence of MPL estimators T_n asymptotically follow a normal distribution. Therefore, with the proposed equations, it is possible to completely characterize the asymptotic behavior of the MPL estimators of the Potts MRF model. In many image processing applications, asymptotic evaluations are sensible since typical random field sizes are 128×128, 256×256 and 512×512. The final expression for estimating the asymptotic variance in terms of Fisher information is

$$\widehat{\mathrm{Var}}_N(\hat{\beta}_{\mathrm{MPL}}) = \frac{\hat{I}^1_{\mathrm{obs}}(\hat{\beta}_{\mathrm{MPL}})}{[\hat{I}^2_{\mathrm{obs}}(\hat{\beta}_{\mathrm{MPL}})]^2}, \tag{3.16}$$

where $\hat{I}^1_{\mathrm{obs}}(\hat{\beta}_{\mathrm{MPL}})$ and $\hat{I}^2_{\mathrm{obs}}(\hat{\beta}_{\mathrm{MPL}})$ denotes the observed Fisher information using the first and the second derivatives of the logarithm of the pseudo-likelihood function, respectively.

4 Experiments and results in image analysis

In order to demonstrate the application of the asymptotic variance in testing and evaluating the proposed pseudo-likelihood equations for Potts MRF model parameter estimation, we present the results obtained in experiments using Markov chain Monte Carlo simulation methods (Dubes and Jain (1989); Landau and Binder (2000); Chib (2004); Winkler (2006)) by comparing the values of $\hat{\beta}_{\mathrm{MPL}}$, asymptotic variances, test statistics and P-values regarding second and third-order neighborhood systems using synthetic images, representing several Potts model outcomes.

Briefly speaking, sampling is the process of generating a realization of a random field, given a model whose parameters have been specified or estimated (Dubes and Jain (1989)). Markov chain Monte Carlo algorithms can be used to generate samples from the posterior probability distributions by simulating a proper Markov chain over all possible states.

The objective of the proposed evaluation methodology is to validate the following hypothesis:

H: the proposed pseudo-likelihood equations provide results that are statistically equivalent to the real parameter values, that is,

$$H : \beta = \hat{\beta}_{\mathrm{MPL}}. \tag{4.1}$$

The asymptotic distribution of the sequence of Potts model MPL estimators $T_n(\beta)$ is normal (Jensen and Künsh (1994)):

$$T_n(\beta) \approx N(\mu_n(\beta), \sigma^2_n(\beta)). \tag{4.2}$$

Using the consistency property of MPL estimators and adopting our approximation for the asymptotic variance, we completely characterize the asymptotic distribution of the estimator

$$\beta_n \approx N(\hat{\beta}_{\mathrm{MPL}}, \widehat{\mathrm{Var}}_n(\hat{\beta}_{\mathrm{MPL}})), \tag{4.3}$$

and then define the following test statistic:

$$Z_n = \frac{\beta_n - \mu_n(\beta)}{\sigma_n^2(\beta)} = \frac{\beta_n - \hat{\beta}_{\mathrm{MPL}}}{\widehat{\mathrm{Var}}_n(\hat{\beta}_{\mathrm{MPL}})} \approx N(0, 1) \tag{4.4}$$

creating the decision rule: reject H if $|Z_n| > c$.

Considering a test size α (in all experiments in this work we set $\alpha = 0.1$), that is, the maximum probability of incorrectly rejecting H is α, we have $c = 1.64$.

However, we want to quantify the evidence against or in favor of the hypothesis H. We propose a complete analysis in terms of the test statistic, the test size and the P-values, calculated by $P(|Z_n| > z_{\mathrm{obs}})$ in this case (two-sided test). The higher the P-values, the more evidence in favor of H. In case of a small P-value, we should doubt of the hypothesis being tested. In other words, in order to reject H, we should have a test size α significantly higher than the P-value. Suppose we have observed a very high value for the test statistic (z_{obs}). In this case $P(|Z_n| > z_{\mathrm{obs}})$, would be approximately zero, indicating that the hypothesis should be rejected. This approach provides a statistically meaningful way to report the results of a hypothesis testing procedure, providing a complete framework for quantitative analysis.

For the experiments, we adopted both single spin–flip MCMC methods, Gibbs Sampler (Geman and Geman (1984)) and Metropolis (Metropolis et al. (1953)), and cluster-flipping MCMC methods, more precisely the Swendsen–Wang (SW) algorithm (Swendsen and Wang (1987); Landau and Binder (2000)), to generate several Potts model outcomes using different known β parameter values. Simulated images for second and third-order neighborhood systems were generated and are shown in Figures 3, 4, 5 and 6. Our objective is to study and assess the behavior of the proposed MPL estimation method (a local approach) using both local and global simulation algorithms. The MPL estimators, obtained by the derived pseudo-likelihood equations were compared with the real parameter values. This information, together with the test statistics and P-values, obtained using the

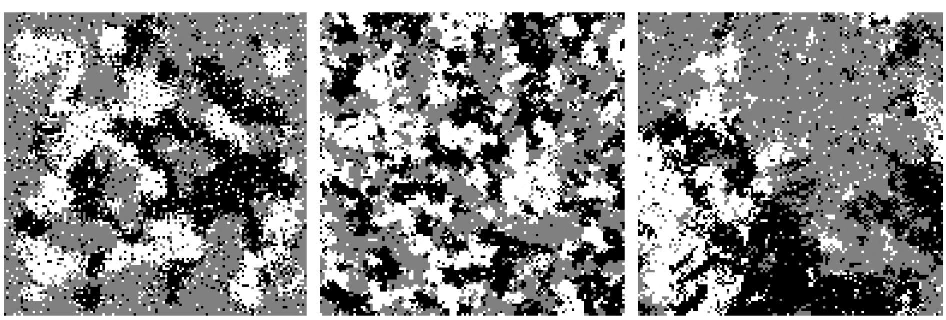

Figure 3 *Synthetic images generated by MCMC simulation algorithms using second-order neighborhood systems for $M = 3$: Gibbs Sampler ($\beta = 0.45$), Metropolis ($\beta = 0.5$) and Swendsen–Wang ($\beta = 0.4$), respectively.*

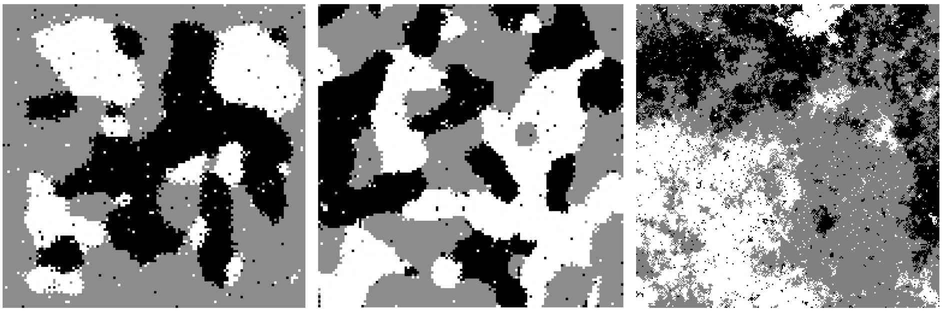

Figure 4 *Synthetic images generated by MCMC simulation algorithms using third-order neighborhood systems for $M = 3$: Gibbs Sampler ($\beta = 0.45$), Metropolis ($\beta = 0.5$) and Swendsen–Wang ($\beta = 0.4$), respectively.*

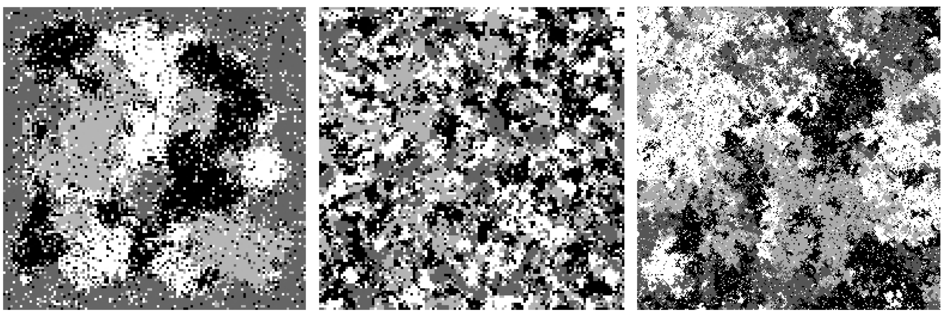

Figure 5 *Synthetic images generated by MCMC simulation algorithms using second-order neighborhood systems for $M = 4$: Gibbs Sampler ($\beta = 0.45$), Metropolis ($\beta = 0.5$) and Swendsen–Wang ($\beta = 0.4$), respectively.*

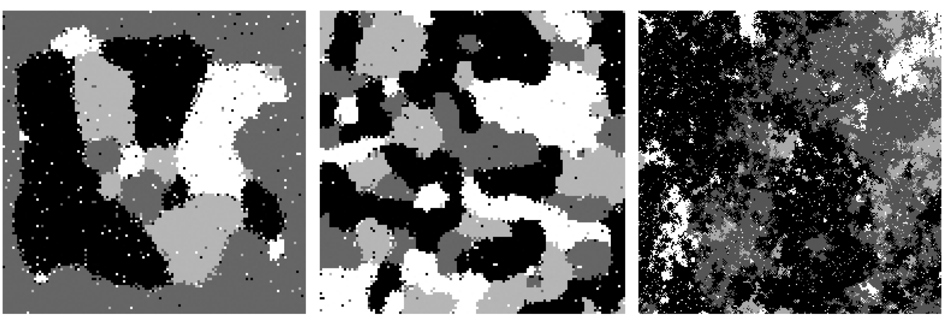

Figure 6 *Synthetic images generated by MCMC simulation algorithms using second-order neighborhood systems for $M = 4$: Gibbs Sampler ($\beta = 0.45$), Metropolis ($\beta = 0.5$) and Swendsen–Wang ($\beta = 0.4$), respectively.*

proposed approximation for the asymptotic variance, provide a statistical procedure to validate and assess the accuracy of he proposed estimation method. The

Table 3 *MPL estimators, observed Fisher information, asymptotic variances, test statistics and P-values for synthetic MCMC simulated images using second-order neighborhood systems*

	Swendsen–Wang		Gibbs Sampler		Metropolis	
M	3	4	3	4	3	4
β	0.4	0.4	0.45	0.45	0.5	0.5
$\hat{\beta}_{\text{MPL}}$	0.4460	0.4878	0.3849	0.4064	0.4814	0.4889
$\lvert\beta - \hat{\beta}_{\text{MPL}}\rvert$	0.0460	0.0878	0.0651	0.0436	0.0186	0.0111
$\hat{I}^1_{\text{obs}}$	0.4694	0.6825	0.8450	1.3106	0.3908	0.8258
$\hat{I}^2_{\text{obs}}$	3.0080	3.3181	3.8248	4.5387	2.2935	2.6436
$\widehat{\text{Var}}_n(\hat{\beta}_{\text{MPL}})$	0.0519	0.0620	0.0578	0.0636	0.0743	0.1182
Z_n	0.2458	0.3571	0.2707	0.1729	0.0682	0.0322
p-values	0.8104	0.7264	0.7872	0.8650	0.9520	0.9760

Table 4 *MPL estimators, observed Fisher information, asymptotic variances, test statistics and P-values for synthetic MCMC simulated images using third-order neighborhood systems*

	Swendsen–Wang		Gibbs Sampler		Metropolis	
M	3	4	3	4	3	4
β	0.4	0.4	0.45	0.45	0.5	0.5
$\hat{\beta}_{\text{MPL}}$	0.3602	0.3772	0.4185	0.4309	0.4896	0.4988
$\lvert\beta - \hat{\beta}_{\text{MPL}}\rvert$	0.0398	0.0228	0.0315	0.0191	0.0104	0.0012
$\hat{I}^1_{\text{obs}}$	0.2738	0.5372	0.1104	0.1433	0.0981	0.1269
$\hat{I}^2_{\text{obs}}$	3.5691	4.6800	1.8703	2.3416	1.4165	1.4547
$\widehat{\text{Var}}_n(\hat{\beta}_{\text{MPL}})$	0.0215	0.0245	0.0316	0.0261	0.0489	0.0600
Z_n	0.2510	0.1456	0.1772	0.1182	0.0470	0.0049
p-values	0.8036	0.8886	0.8572	0.9044	0.9602	0.9940

obtained results for second and third-order neighborhood systems are shown in Tables 3 and 4, respectively.

The obtained results clearly show that the asymptotic variance is reduced in third-order neighborhood systems, increasing the P-values, suggesting that the use of higher-order systems improves Potts model MPL estimation, since it enhances the accuracy of the method. Note that in all cases, the test statistic is far below the threshold $c = 1.64$ and the P-values are far above the test size $\alpha = 0.1$. Considering the observed data used in the experiments, we conclude that the differences between the real parameters and the proposed MPL estimators are not significant. Therefore, based on statistical evidences, it is strongly recommended that we accept the hypothesis H, assessing the accuracy of proposed methodology.

5 Conclusion

The definition of asymptotic variance is an important tool for statistical analysis in MRF parameter estimation. In this paper, we have addressed the problem of characterizing the asymptotic normal distribution of maximum pseudo-likelihood estimators of Potts MRF model parameter. First, we derived analytical pseudo-likelihood equations for Potts model maximum pseudo-likelihood estimation on higher-order neighborhood systems. The major contribution is that the proposed equations allow the modeling of less restrictive neighborhood systems in a large number of MRF applications in a computationally feasible way. We also proposed a hypothesis testing approach for quantitative data analysis, deriving an approximation for the asymptotic variance of MPL Potts model parameter estimators using the observed Fisher information. Our motivation was the possibility of complete characterization of the asymptotic distribution of the Potts MPL estimator in the limiting case, allowing interval estimation and also hypothesis testing concerning a given statement about the model parameter. The main conclusion of this work is that higher-order neighborhood systems can improve Potts MRF model parameter estimation accuracy by reducing the asymptotic variance. Future works may include a detailed analysis of the observed Fisher information for single observations, that is, single contextual configuration patterns and also the investigation of conditions for information equality regarding the β parameter in the Potts model.

Appendix

Simplification of the expression for observed Fisher information using second derivative of the pseudo-likelihood function. From the numerator of equation (3.14), after some algebraic manipulations, we can write

$$\begin{aligned}\left[\sum_{\ell=1}^{M} U_i(\ell)^2 e^{\beta U_i(\ell)}\right]\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right] &= \sum_{\ell=1}^{M}\left[U_i(\ell)e^{\beta U_i(\ell)}\right]^2 \\ &+ \sum_{\ell=1}^{M-1}\left\{\sum_{k=\ell+1}^{M}\left[\left(U_i(\ell)^2+U_i(k)^2\right)e^{\beta(U_i(\ell)+U_i(k))}\right]\right\}.\end{aligned} \tag{5.1}$$

Expanding the square in the second term of the numerator of equation (3.14) gives

$$\begin{aligned}\left[\sum_{\ell=1}^{M} U_i(\ell)e^{\beta U_i(\ell)}\right]^2 \\ = \sum_{\ell=1}^{M}\left[U_i(\ell)e^{\beta U_i(\ell)}\right]^2 + 2\sum_{\ell=1}^{M-1}\left\{\sum_{k=\ell+1}^{M}\left[U_i(\ell)U_i(k)e^{\beta(U_i(\ell)+U_i(k))}\right]\right\}.\end{aligned} \tag{5.2}$$

Finally, the numerator of (3.14) is reduced to

$$\sum_{\ell=1}^{M-1}\left\{\sum_{k=\ell+1}^{M}\left[\left(U_i^2(\ell)-2U_i(\ell)U_i(k)+U_k^2(k)\right)e^{\beta(U_i(\ell)+U_i(k))}\right]\right\}, \tag{5.3}$$

which leads to the expression presented in equation (3.15).

Acknowledgments

We would like to thank FAPESP for the financial support through Alexandre L. M. Levada student scholarship (grant no. 06/01711-4) and also the reviewers for the valuable contributions and discussions.

References

Adams, C. C. (1994). *The Knot Book*. W. H. Freeman, New York. MR1266837

Berthod, M., Kato, Z., Yu, S. and Zerubia, J. (1996). Bayesian image classification using Markov random fields. *Image and Vision Computing* **14** 285–295.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B* **36** 192–236. MR0373208

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society. Series B* **48** 259–302. MR0876840

Bickel, P. J. (1991). *Mathematical Statistics*. Holden Day, New York.

Brent, R. (1973). *Algorithms for Minimization Without Derivatives*. Prentice Hall, New York. MR0339493

Casella, G. and Berger, R. L. (2002). *Statistical Inference*, 2nd ed. Duxbury, New York.

Chib, S. (2004). Markov chain Monte Carlo technology. In *Handbook of Computational Statistics* (J. E. Gentle, W. Härdle and Y. Mori, eds.) 72–98. Springer, Berlin. MR2090139

Dubes, R. and Jain, A. (1989). Random field models in image analysis. *Journal of Applied Statistics* **16** 131–164.

Efron, B. F. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65** 457–487. MR0521817

Enting, I. G. and Guttmann, A. J. (2003). Susceptibility amplitudes for the three- and four-state Potts models. *Physica A: Statistical Mechanics and Its Applications* **321** 90–107. MR1993604

Frery, A. C., Correia, A. H. and Freitas, C. C. (2007). Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information. *IEEE Transactions on Geoscicence and Remote Sensing* **45** 3098–3109.

Ge, M. L., Hu, L. and Wang, Y. (1992). Knot theory, partition function and fractals. *Journal of Knot Theory and Ramifications* **5** 37–54. MR1373809

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.

Hammersley, J. M. and Clifford, P. (1971). Markov field on finite graphs and lattices. Unpublished.

Heisenberg, W. (1928). Zur Theorie des Ferromagnetismus. *Zeitschr. f. Physik* **49** 619–636.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschr. f. Physik* **39** 253–258.

Jensen, J. L. and Künsh, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics* **46** 475–486. MR1309718

Jim, X. and Zhang, F. (2004). Jones polynomials and their zeros for a family of links. *Physica A: Statistical and Theoretical Physics* **333** 183–196. MR2100217

Landau, D. P. and Binder, K. (2000). A guide to Monte Carlo simulations in statistical physics. Cambridge Univ. Press, Cambridge. MR1781083

Lehmann, E. L. (1983). *Theory of Point Estimation*. Wiley, New York. MR0702834

Li, S. Z. (2001). *Markov Random Field Modeling in Image Analysis*, 2nd ed. Springer, New York.

Liang, G. and Yu, B. (2003). Maximum pseudo likelihood estimation in network tomography. *IEEE Transactions on Signal Processing* **51** 2043–2053.

Liu, Z., Luo, J. and Shao, C. (1996). Potts model for exaggeration of a simple rumor transmitted by recreant rumormongers. *Physical Review E* **64** 046134.

Marroquin, J., Mitter, S. and Poggio, T. (1987). Probabilistic solution of ill-posed problems in computer vision. *Journal of American Statistical Society* **82** 76–89.

Merks, R. M. H. and Glazier, J. A. (2005). A cell-centered approach to developmental biology. *Physica A: Statistical Mechanics and Its Applications* **352** 113–130.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equation of state calculations by fast computer machines. *Journal of Physical Chemistry* **21** 1987–2092.

Montroll, E. (1941). Statistical mechanics of nearest neighbor systems. *Journal of Chemical Physics* **9** 706–721.

Ouchi, N. B., Glazier, J. A., Rieu, J. P., Upadhyaya, A. and Sawada, Y. (2003). Improving the realism of the cellular Potts model in simulations of biological cells. *Physica A: Statistical Mechanics and Its Applications* **329** 451–458. MR2026851

Swendsen, R. and Wang, J. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters* **58** 86–88.

Waks, A., Tretiak, O. J. and Gregoriou, G. K. (1990). Restoration of noisy regions modeled by noncausal Markov random fields of unknown parameters. *Proceedings of the International Conference on Pattern Recognition (ICPR)* **2** 170–175.

Winkler, G. (2006). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, 2nd ed. Springer, New York. MR1950762

Wolff, U. (1989). Collective Monte Carlo updating for spin systems. *Physical Review Letters* **62** 361–364.

Won, C. S. and Gray, R. M. (2004). *Stochastic Image Processing*. Kluwer Academics/Plenum Publishers, New York.

Wu, F. Y. (1992). Jones polynomial as a Potts model partition function. *Journal of Knot Theory and Ramifications* **1** 47–57. MR1155093

Yamazaki, T. and Gingras, D. (1995). Image classification using spectral and spatial information based on MRF Models. *IEEE Transactions on Image Processing* **4** 1333–1339.

Yu, S. and Berthod, M. (1995). A game strategy approach for image labeling. *Computer Vision and Image Understanding* **61** 32–37.

Zhang, J., Fieguth, P. and Wang, D. (2000). Random field models. In *Handbook of Image and Video Processing* (A. Bovik, ed.) 301–312. Academic Press, Orlando, FL.

A. L. M. Levada
A. Tannús
Instituto de Física de São Carlos
Universidade de São Paulo
São Carlos/SP, 13560–970
Brazil
E-mail: alexandreluis@ursa.ifsc.usp.br
goiano@ifsc.usp.br

N. D. A. Mascarenhas
Departamento de Computação
Universidade Federal de São Carlos
Rodovia Washington Luis, km 235
São Carlos/SP, 13565–905
Brazil
E-mail: nelson@dc.ufscar.br
URL: http://www.dc.ufscar.br/~nelson