Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

ESTIMATING THE VARIANCE OF THE MAXIMUM PSEUDO-LIKELIHOOD ESTIMATOR

Lynne Seymour
Department of Statistics
University of Georgia
Athens, GA 30602-1952
E-mail: seymour@stat.uga.edu

Abstract

The use of the pseudo-likelihood estimator for Gibbs-Markov random field models has a distinct advantage over more conventional approaches mainly due to its computational efficiency. Indeed, the maximum pseudo-likelihood estimator (MPLE) is often used as the Monte Carlo parameter in Markov chain Monte Carlo (MCMC) simulations. The MPLE itself has some very nice estimation properties, though its variance is still undiscovered. In this paper, the moving-block bootstrap is employed to estimate the variance of the MPLE in the Ising model.

KEY WORDS: parameter estimation, Gibbs random fields, Markov random fields, parametric bootstrap, moving-block bootstrap, subsampling

1 Introduction

Gibbs-Markov random fields (GMRFs) are statistical models used to study the spatial relationships among data taken on a grid. Although these models were developed around the turn of the century in physics as models for particle interactions (Gibbs, 1902), it has only been within the past fifteen years or so that they have been seriously considered by statisticians and other scientists as models for spatially related data. The form of the distribution used in practice is fairly simple:

$$P(X_\Lambda = x_\Lambda) = \frac{\exp\{\boldsymbol{\theta} \cdot \boldsymbol{Y}(x_\Lambda)\}}{Z}, \tag{1}$$

where x_Λ represents observations taken on a grid Λ, $\boldsymbol{\theta}$ represents the parameters of the distribution, $\boldsymbol{Y}(\cdot)$ is a function strictly of the data, and $\mathcal{Z}$ is a normalizing constant.

This model has many applications throughout the scientific community. The pioneering paper of Besag (1974) used simple GMRFs to study the occurrence of new plant growth in a defunct mine, and to study grain and straw yield of wheat plots. In each of these models, the area of interest was partitioned into a grid, and the presence of new growth and grain/straw yields, respectively, were the data taken at each grid site. Classical development of this model has been in image analysis, in which the grid consists of the pixels on the computer screen, and the observed data are the colors at each pixel. The model then aids in, for example, image restoration (Geman and Geman, 1984; Tjelmeland and Besag, 1998), detecting boundaries (Geman, Geman, Graffigne, and Dong, 1990), or recognizing and simulating textures (Geman and Graffigne, 1986; Seymour, 1993). More recently, Smith et. al (2000) and Seymour (2000) use such a model to study social networks; more specifically, the clients of a social service agency are considered the "grid", and the data indicate whether the clients changed case managers.

A first step in making inferences with any statistical model is parameter estimation, which can be especially challenging for GMRFs due to the size of the grid and/or the dependence of the data. Maximum likelihood estimation (MLE) is a standard technique for most models, but the nature of the GMRF model makes the MLE computationally intractable for even reasonable grid sizes. A commonly-used way of circumventing this problem is to approximate the MLE using Markov chain Monte Carlo (MCMC) methods (Geyer and Thompson, 1992), but this can become very cumbersome in some applications and can be unstable for cases in which the dependence in the data is strong (Seymour and Ji, 1996). Another standard technique is method-of-moments estimation, which requires that the dependence between data at neighboring grid points to be weak (Sherman and Carlstein, 1994).

A non-standard parameter estimation technique that is very easy to implement is the maximum pseudo-likelihood estimate (MPLE). Instead of maximizing the likelihood based on the model in (1), a pseudo-likelihood is maximized. The pseudo-likelihood which Besag (1975) first proposed simply multiplies the conditional distributions at grid sites given the values at neighboring sites:

$$\mathcal{PL}(\boldsymbol{\theta}; x_\Lambda) = \prod_{i \in \Lambda^*} P(X_{\{i\}} = x_{\{i\}} | X_{\mathcal{N}_i} = x_{\mathcal{N}_i}), \tag{2}$$

where $\Lambda^* \subset \Lambda$ is the set of all sites in Λ with a complete set of neighbors in Λ, and $\mathcal{N}_i \subset \Lambda$ is the set of points which are neighbors of the site $i \in \Lambda^*$.

These single-site conditional distributions are easily calculated, making the MPLE extremely appealing as an estimation technique.

The MPLE is often used to initialize the MCMC method of parameter estimation, which is preferred over the MPLE because it gives an approximation of the likelihood function. The MPLE is used in this way because it is quickly computed, and because it is close to the parameter the MCMC method is trying to estimate (which cuts down on MCMC iterations).

2 Modelling Background

Let Λ_n be an $n \times n$ lattice in $\mathbb{Z}^2$, and let X_i be a random variable associated with the site $i \in \mathbb{Z}^2$. Then $X_{\Lambda_n} = \{X_i, i \in \Lambda_n\}$ is called a random field on Λ_n. The state space S is the collection of all possible values of X_i, $i \in \mathbb{Z}^2$, and $\Omega_n = S^{\Lambda_n}$ is the collection of all possible realizations of the random field X_{Λ_n}. For a site $i \in \mathbb{Z}^2$, a collection $\mathcal{N}_i$ of sites having the properties $i \notin \mathcal{N}_i$ and $i \in \mathcal{N}_j \Leftrightarrow j \in \mathcal{N}_i$ is called the neighborhood of the site i. The collection of all neighborhoods $\mathfrak{N} = \{\mathcal{N}_i, i \in \mathbb{Z}^2\}$ is called the neighborhood system.

A random field is called a Markov random field (MRF) with respect to the neighborhood system $\mathfrak{N}$ if its probability distribution P on $\Omega_{\mathbb{Z}^2}$ satisfies

$$P\left(X_i = x_i | X_{\mathbb{Z}^2\backslash\{i\}} = x_{\mathbb{Z}^2\backslash\{i\}}\right) = P\left(X_i = x_i | X_{\mathcal{N}_i} = x_{\mathcal{N}_i}\right) \tag{3}$$

for each $i \in \mathbb{Z}^2$ and $x \in \Omega_{\mathbb{Z}^2}$. where $\mathbb{Z}^2\backslash\{i\}$ is the set of all sites in $\mathbb{Z}^2$ except site i. These single-site conditional probabilities are called the local characteristics of the MRF.

The pair potential $\mathfrak{U} = \left\{u(x_i, x_j) : i, j \in \mathbb{Z}^2, i \neq j\right\}$ is a collection of deterministic functions which quantify how values at pairs of sites interact. If each of the functions in $\mathfrak{U}$ is zero for every pair of sites farther than a fixed finite range R from each other, then the collection is called a pair-potential of range R. The energy associated with $x \in \Omega_{\mathbb{Z}^2}$ on Λ_n, denoted $H_{\Lambda_n}(x)$, is a functional of $u(\cdot,\cdot)$ which summarizes all of the pair interactions of the random field on Λ_n. For example, an energy function may take the form

$$H_{\Lambda_n}(x) = \sum_{i \in \Lambda_n} \left(\sum_{j \in N_i, j \in \Lambda_n} u(x_i, x_j) + \sum_{j \in N_i, j \notin \Lambda_n} u(x_i, x_j) \right),$$

which gives an additive summary of the pair interactions on Λ_n.

A random field is called a Gibbs random field (GRF) induced by the pair-potential $\mathfrak{U}$ if its probability distribution P on $\Omega_{\mathbb{Z}^2}$ satisfies

$$P\left(X_{\Lambda_n} = x_{\Lambda_n} | X_{\Lambda_n^c} = x_{\Lambda_n^c}\right) = \frac{\exp\left[-H_{\Lambda_n}(x_{\Lambda_n}; x_{\Lambda_n^c})\right]}{\mathcal{Z}_{\Lambda_n}} \tag{4}$$

where $\mathcal{Z}_{\Lambda_n}$ is a normalizing factor which is a sum over all possible $x \in \Omega_{\Lambda_n}$. One of the more difficult problems with using GRFs is that, in some cases, these conditional distributions do not uniquely determine the distribution of the field. Such a condition is called phase transition, and implies spatial long-range dependence in the field.

A random field is a MRF if and only if it is a GRF induced by finite-range potentials (Besag, 1974; Geman, 1991). The model given in (1) is an exponential family random field which satisfies both (3) and (4); hence the name Gibbs-Markov random field.

3 The MPLE and its Properties

The likelihood function is computationally intractable for GMRFs due to the normalizing factor in (4): For a 100×100 binary grid, which is trivial in most applications, this factor is a sum of 2^{100^2} terms. Hence alternatives to the likelihood are very desirable for these models. Computationally intensive alternatives include using a Markov chain Monte Carlo approximation to the likelihood function (Geyer and Thompson, 1992), and estimating the normalizing factor via simulation.

The alternative which was first proposed by Besag (1975) is to use the pseudo-likelihood in (2), rather than the likelihood. Note that the pseudo-likelihood is simply the product of the local characteristics (using the Markov property), and that the conditional distributions at a single site make the normalizing factor in (4) much easier to compute. Parameter estimates derived by maximizing the pseudo-likelihood are called maximum pseudo-likelihood estimates (MPLEs).

Asymptotics in a random field setting are traditionally taken on a single realization of an $n \times n$ region as n goes to infinity. The existence, uniqueness, and strong consistency of the MPLE were established by Geman and Graffigne (1986), while Comets (1992) established that the convergence rate of the MPLE (and, incidentally, of the MLE) is on the order of e^{-n^2} as $n \to \infty$. Consistency of the MPLE is not restricted to a grid: Jensen and Møller (1991) show the MPLE to be consistent for point processes, and Mase (1995) shows it to be consistent for continuous-state-space Gibbs processes. The moderate deviation probabilities for the MPLE have been shown to decay as $n^{-\alpha}$ as $n \to \infty$, where α is not necessarily less than one (Ji and Seymour, 1996). Each of these convergence results have no requirements on the strength of the dependence in the random field.

In addition, a model selection criterion based on penalized pseudo-likelihood, similar to the Bayesian information criterion (BIC) of time series, was shown to be weakly consistent without assuming any conditions on the strength of dependence. In contrast, the traditional BIC for GMRFs using

penalized likelihood (and using the MCMC approximation to the likelihood) was shown to approximate the Bayesian solution (i.e., the solution which minimizes the risk under a 0-1 loss function) to the model selection problem only under a weak-dependence condition (Seymour and Ji, 1996). This would seem to imply that the MPLE in some way makes more optimal use of the dependence structure in the observed GMRF than the MLE or its MCMC approximation. Further study is needed into the sufficiency of these estimates.

Guyon (1987) has shown the asymptotic normality of the MPLE under weak dependence conditions (specifically, under spatial mixing with an exponentially decaying mixing coefficient). Jensen and Künsch (1994) established the asymptotic normality of a stochastically normed MPLE for certain point processes, regardless of the strength of dependence. Janzura and Lachout (1995) showed that the sampling distribution of the MPLE is a normal mixture when the strength of spatial dependence is large, and Cométs and Janzura (1998) more recently establish the asymptotic normality of the stochastically normed MPLE in complete generality for conditionally centered random fields.

Grasping the variance of the MPLE is a difficult proposition - even for the simplest GMRF models (Cressie, 1993) - though some results are known. Guyon (1987) derived the asymptotic variance under weak spatial dependence, but it depends upon the intractible joint distribution of the field. The (restricted) mean square error has been shown to decrease asymptotically like n^{-2} (Ji and Seymour, 1996) regardless of the strength of dependence. Sherman (1996) developed a moving-block bootstrap, or sub-sampling, method which can be used for estimating the variance, which is $\mathcal{L}_2$-consistent as long as the MPLE itself is asymptotically normal (i.e., under weak mixing conditions). The variance is an important component for MPLE-based inference, but much remains to be done in both deriving and estimating it.

Once the variance of the MPLE is better-understood, efficiency of the MPLE relative to the MLE may be investigated. The MLE has been shown to be efficient for GMRFs (Mase, 1984). However, the efficiency of the MPLE relative to the MLE has been tabulated (Besag, 1977; Kashyap and Chellappa, 1983), and it appears that the MPLE is not as efficient as the MLE.

4 The Variance of the MPLE via Simulation

Before one can evaluate methods for estimating the variance of the MPLE, one must have an idea of what the target value is. Thus a simulation study was conducted in an effort to understand what the variance of the MPLE

looks like.

This small simulation study employed the well-known Ising model, in which the neighborhood around each site consists of its four nearest neighbors, and the interactions between sites i and j behave according to $\theta x_i x_j$. The parameter θ governs how strongly the data are dependent, and in fact the Ising model is the only Gibbs distribution for which there are necessary and sufficient conditions for no phase transition in terms of an explicit critical value: $|\theta| < \frac{1}{2}\ln\left(1+\sqrt{2}\right) \approx 0.44$ (Ellis, 1985). Thus, in the case that $\theta = 0.1$, there is no phase transition and the data are weakly dependent, as seen in Figure 1; if $\theta = 1$ then there is phase transition and they are very strongly dependent, as seen in Figure 2. The Gibbs sampler (Geman and Geman, 1984) was used to generate 100 random fields of several sizes from these Ising models. The MPLE was then calculated on each of these fields to simulate its sampling distribution, and the sample variance of this sampling distribution was computed.

Figure 1: A Realization for $\theta = 0.1$

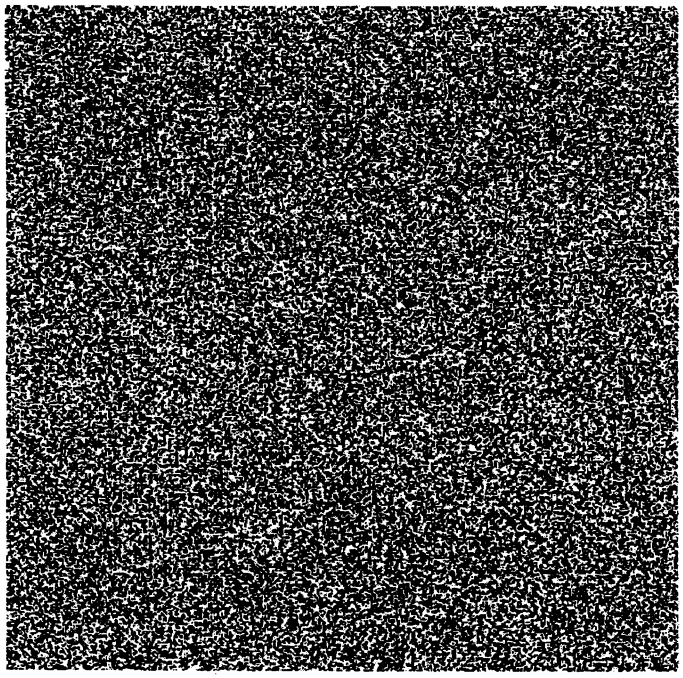


Figure 2: A Realization for $\theta = 1$

Table 1 and Table 2 give the average of 10 replications of the sample variances for the simulated sampling distribution of the MPLE for several $n \times n$ random fields. The third column of each table shows (variance)$\cdot n^2$ in an effort to understand whether the variance is proportional to n^{-2}. Figure 3 and Figure 4 plot the 10 sample variances, as well as these values multiplied by n^2, against n. In the weak-dependence case (Table 1; Figure 4), it is clear that the variance behaves as $O(n^2)$. As is to be expected, this same rate is not as apparent in the strong-dependence case (Table 2; Figure 4); however, if one discounts the case $n = 100$ as possibly being too small for the asymptotics, then the relationship is more clear.

Table 1: Average Simulated Variances, $\theta = 0.1$

n	average variance	(average variance)*n^2
50	0.00021125	0.5281
100	0.00004664	0.4664
300	0.00000570	0.5120
500	0.00000200	0.4889
1000	0.00000040	0.4437

Figure 3: Simulated Variances $\theta = 0.1$

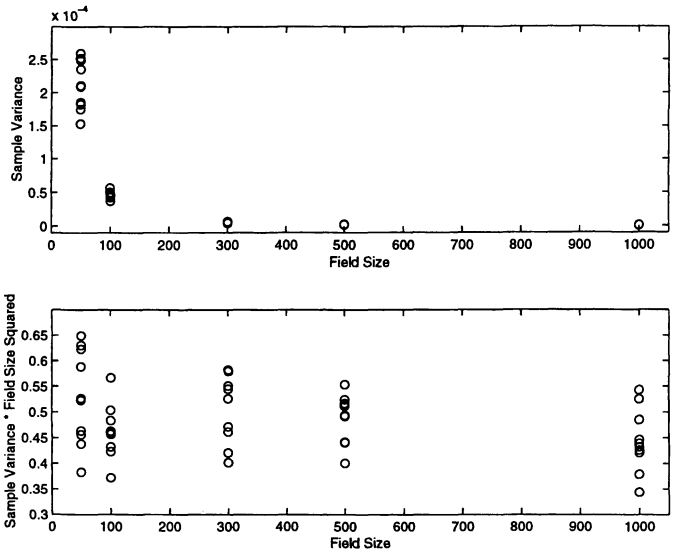
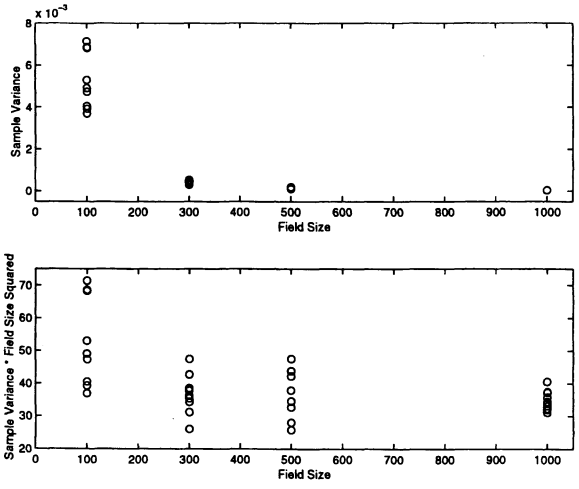


Table 2: Average Simulated Variances, $\theta = 1$

n	average variance	(average variance)*n^2
100	0.00542753	54.26
300	0.00040862	36.78
500	0.00014896	37.24
1000	0.00003478	34.78

Figure 4: Simulated Variances, $\theta = 1$



5 Estimating the Variance via Sub-sampling

In Section 4, the variance of the MPLE was estimated by sampling many random fields. Rather than sampling repeatedly, one may simply sub-sample the random field in hand, potentially saving much computing time (though not necessarily). This will yield accurate results in many cases, most notably when dependence is weak (Sherman, 1996).

Sherman and Carlstein (1994) introduced a sub-sampling procedure to estimate a general statistic (cf. also Politis and Romano, 1993) which is useful in estimating the variance of the MPLE. However, the MPLE must be asymptotically normal for their estimate to be consistent, which is the case only under weak dependence. The conjecture explored herein is that a sub-sampling scheme may be used to estimate the variance of the MPLE when there is a limiting distribution for the MPLE, as is the case whether dependence is weak or strong.

In the following examples, we again look at the two cases of the Ising model described in Section 4. In each case, we sample one $n \times n$ GMRF, $n = 100, 300, 500, 1000$. We then partition that GMRF into sub-blocks that are 10%, 15%, 20%, and 25% of the size of the sampled GMRF. In addition, we allow three different degrees of overlap: None (shifting a whole sub-block width), half (shifting a half sub-block width), and one pixel (shifting one pixel, the maximum overlap possible while still shifting).

Table 3 and Table 4 summarize the results of the sub-blocking schemes given above and compare the results to the simulated variances found in Section 4. Several interesting phenomena may be observed.

Table 3 contains the results for the weak dependence case - the case for which much is known. As is easily seen, the size of the sub-blocks and the amount of overlap matter very little, and all sub-sampled estimates are close to the simulated variance. Note that for the larger cases of $n = 500$ and $n = 1000$, the one-pixel shift is omitted - this is due to the fact that the number of sub-blocks for the one-pixel shift is prohibitively large in these cases.

Table 3: Variance Estimates when $\theta = 0.1$

N	Size of Sub-block	Shift	Number of Sub-blocks	Average MPLE	Estimated Var(MPLE)	Simulated Var(MPLE)
100	10	Whole	100	0.1152	80×10^{-6}	46.6×10^{-6}
		Half	361	0.1128	89×10^{-6}	
		One	8281	0.1157	83×10^{-6}	
	15	Whole	36	0.1182	69×10^{-6}	
		Half	121	0.1107	54×10^{-6}	
		One	7396	0.1113	59×10^{-6}	
	20	Whole	25	0.1097	63×10^{-6}	
		Half	81	0.1081	62×10^{-6}	
		One	6561	0.1100	49×10^{-6}	
	25	Whole	16	0.1110	74×10^{-6}	
		Half	36	0.1102	50×10^{-6}	
		One	5776	0.1091	42×10^{-6}	
300	30	Whole	100	0.0993	6×10^{-6}	5.7×10^{-6}
		Half	361	0.0988	6×10^{-6}	
		One	73441	0.0991	6×10^{-6}	
	45	Whole	36	0.1006	6×10^{-6}	
		Half	144	0.0988	6×10^{-6}	
		One	65536	0.0989	5×10^{-6}	
	60	Whole	25	0.0994	4×10^{-6}	
		Half	81	0.987	5×10^{-6}	
		One	58081	0.0991	5×10^{-6}	
	75	Whole	16	0.0991	6×10^{-6}	
		Half	36	0.0997	4×10^{-6}	
		One	51076	0.0994	5×10^{-6}	
500	50	Whole	100	0.1003	2×10^{-6}	2.0×10^{-6}
		Half	361	0.1013	2×10^{-6}	
	75	Whole	36	0.1012	2×10^{-6}	
		Half	144	0.1016	2×10^{-6}	
	100	Whole	25	0.1002	2×10^{-6}	
		Half	81	0.1010	2×10^{-6}	
	125	Whole	16	0.1003	2×10^{-6}	
		Half	36	0.1033	2×10^{-6}	
1000	100	Whole	100	0.0998	0.5×10^{-6}	0.4×10^{-6}
		Half	361	0.0997	0.5×10^{-6}	
	150	Whole	36	0.1003	0.6×10^{-6}	
		Half	144	0.1000	0.4×10^{-6}	
	200	Whole	25	0.0997	0.3×10^{-6}	
		Half	81	0.1000	0.5×10^{-6}	
	250	Whole	16	0.0997	0.6×10^{-6}	
		Half	49	0.1000	0.5×10^{-6}	

Table 4, which presents the results in the strong dependence case, contains much more interesting phenomena. Notice how the schemes with the smaller sub-blocks give bad estimates of the variance, and indeed of the MPLE itself. This may be understood intuitively by realizing that if the sub-blocks don't contain an accurate representation of the dependence structure in the entire field, the sub-sampling scheme will fail to provide an accurate estimate. For $\theta = 1$, it is clear that good estimates of the MPLE can be obtained from fields of size $n = 75$ or larger. Once the sub-block size becomes large enough to capture the dependence structure in the entire field, then the degree of overlap begins to play a role in the accuracy of the estimate. Even that, however, seems to fade as is seen in the $n = 500$ and $n = 1000$ cases. Again, the cases employing maximum overlap were too computationally intense for inclusion here.

Table 4: Variance Estimates when $\theta = 1$

N	Size of Sub-block	Shift	Number of Sub-blocks	Average MPLE	Estimated Var(MPLE)	Simulated Var(MPLE)
100	10	Whole	100	4.5685	0.021924	0.00542753
		Half	361	4.6020	0.019078	
		One	8281	5.2807	0.013587	
	15	Whole	36	4.2588	0.067236	
		Half	121	4.6205	0.066193	
		One	7396	4.6828	0.064116	
	20	Whole	25	3.7488	0.160362	
		Half	81	3.5458	0.161188	
		One	6561	4.2046	0.170413	
	25	Whole	16	3.0882	0.313485	
		Half	36	2.6370	0.191927	
		One	5776	3.5785	0.324719	
300	30	Whole	100	2.8113	0.045737	0.00040862
		Half	361	2.9351	0.048044	
		One	73441	3.1094	0.051915	
	45	Whole	36	2.1644	0.087682	
		Half	144	2.0011	0.078927	
		One	65536	1.9984	0.079410	
	60	Whole	25	1.3051	0.043379	
		Half	81	1.2719	0.042598	
		One	58081	1.3693	0.060681	
	75	Whole	16	1.0137	0.000794	
		Half	36	0.9917	0.000387	
		One	51076	1.0665	0.015967	
500	50	Whole	100	1.4124	0.015656	0.00014896
		Half	361	1.4113	0.016536	
	75	Whole	36	1.0015	0.000129	
		Half	144	1.0388	0.002883	
	100	Whole	25	1.0032	0.000159	
		Half	81	1.0008	0.000139	
	125	Whole	16	0.9984	0.000118	
		Half	36	0.9955	0.000130	
1000	100	Whole	100	1.0077	0.000042	0.00003478
		Half	361	1.0084	0.000042	
	150	Whole	36	1.0014	0.000044	
		Half	144	1.0000	0.000033	
	200	Whole	25	0.9990	0.000034	
		Half	81	0.9988	0.000041	
	250	Whole	16	0.9972	0.000040	
		Half	49	0.9954	0.000039	

6 Conclusion

Several fundamental questions about the properties of the MPLE remain unresolved - including the form of its sampling variance. However, because of its simplicity and ease of use, it is an appealing estimator. The minimal accomplishment of discovering the variance of the distribution of the MPLE will aid scientists in drawing inferences from their data, and in some cases render MCMC estimation redundant.

This simulation study demonstrates that the variance of the MPLE can be estimated in spite of the fact that it is unknown. Weaknesses in the sub-sampling strategy which result from overlapping sub-blocks are currently being addressed by recasting the sub-sampled values as a stochastic process in which correlation (induced by overlap) is routine.

Acknowledgement

The author would like to thank Michael Sherman of the Department of Statistics at Texas A&M University for helpful comments and suggestions. The author would also like to thank Ingram Olkin and the Department of Statistics at Stanford University for sponsorship under which this work was completed, as well as to acknowledge partial support of this work by the National Science Foundation Grant DMS-9631278 and SES-0074456.

References

Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society*, Series B 6, 192-236.

Besag, J. (1975), "Statistical analysis of non-lattice data," *The Statistician* 24, 179-195.

Besag, J. (1977), "Efficiency of pseudolikelihood estimators for simple Gaussian fields," *Biometrika* 64, 616-618.

Cométs, F. (1992) "On consistency of a class of estimators for exponential families of Markov random fields on the lattice," *Annals of Statistics* 20, 455-468.

Cométs, F. and Janzura, M. (1998), "A central limit theorem for conditionally centered random fields with an application to Markov fields," *Journal of Applied Probability* 35, 608-621.

Cressie, N. (1993), *Statistics for Spatial Data*, Wiley, New York.

Ellis, R. S. (1985), *Entropy, Large Deviations, and Statistical Mechanics*, Springer-Verlag, New York.

Geman, D. (1991), "Random fields and inverse problems in imaging," *Lecture Notes in Mathematics* 1427, 113-193, Springer-Verlag, New York.

Geman, D., Geman, S., Graffigne, C., and Dong, P. (1990) "Boundary detection by constrained optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence 12*, 609-628.

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

Geman, S., and Graffigne, C., (1986) "Markov random field image models and their applications to computer vision," *Proceedings of the International Congress of Mathematicians*, 1496-1517, Berkeley, CA.

Geyer, C. J., and Thompson, E. A., (1992) "Constrained Monte Carlo maximum likelihood for dependent data," *Journal of the Royal Statistical Society*, Series B 54, 657-699.

Gibbs, J. W. (1902) *Elementary Principles of Statistical Mechanics*, Yale University Press.

Guyon, X., (1987), "Estimation d'un champ par pseudo-vraisemblance conditionelle: Etude asymptotique et application au cas de Markovien," *Spatial Processes and Spatial Time Series Analysis (Proceedings of the Sixth Franco-Belgian Meeting of Statisticians*, 1985), F. Droesbeke, ed., Publications des Facultes Universitaires Saint-Louis, Brussels, Belgium, 16-62.

Guyon, X. and H. R. Künsch (1992), "Asymptotic comparison of estimators in the Ising model," in: *Stochastic Models, Statistical Methods and Algorithms in Image Analysis* (P. Barone, A. Frigessi, and M. Piccioni eds.), Lecture Notes in Statist. 74 177-198. Springer, New York.

Ising, E. (1925) "Beitrag zur Theorie des Ferromagnetismus," *Zeitschrift Physik* 31, 253-258.

Janzura, M. (1994), "Asymptotic normality of the maximum pseudo-likelihood estimate of parameters for Gibbs random fields," *Transactions of the 12th Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes*, 119-121.

Janzura, M., and P. Lachout (1995) "A central limit theorem for stationary random fields," *Mathematical Methods of Statistics* 4 463-471.

Jensen, J. L., and Künsch, (1994), "On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes," *Annals of the Institute of Statistical Mathematics* 46, 475-486.

Jensen, J. L., and Møller, J. (1991), "Pseudolikelihood for exponential family models of spatial processes," *Annals of Applied Probability* 1, 445-461.

Ji, C., and Seymour, L. (1996), "A consistent model selection procedure for Markov random fields based on penalized pseudo-likelihood," *Annals of Applied Probability* 6, 423-443.

Kashyap, R., and Chellappa, R. (1983), "Estimation and choice of neighbors in spatial-interaction models of images," *IEEE Transactions on Information Theory* 29, 60-72.

Mase, S., (1984), "Locally asymptotic normality of Gibbs models on a lattice," *Advances in Applied Probability* 16, 585-602.

Mase, S. (1995), "Consistency of the maximum pseudo-likelihood estimator for continuous state space Gibbsian processes," *Annals of Applied Probability* 5, 603-612.

Politis, D. N., and Romano, J. P. (1993). "On the sample variance of linear statistics derived from mixing sequences," *Stochastic Processes and Their Applications* 45, 155-167.

Propp, J. G., and Wilson, D. G. (1995), "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random Structures and Algorithms* 9, 223-252.

Seymour, L. (1993), "Parameter estimation and model selection in image analysis using Gibbs-Markov random fields, Dissertation, The University of North Carolina.

Seymour, L., and Ji, C. (1996), "Approximate Bayes model selection procedures for Gibbs-Markov random fields," *Journal of Statistical Planning and Inference Special Issue on Spatial Statistics* 51, 75-97.

Seymour, L. (2000), "Gibbs Regression and a Test for Goodness of Fit," in *Goodness-of-Fit Tests and Model Validity* (C. Huber-Carol, N. Balakrishnan, M. S. Nikulin, and M. Mesbah, eds.) Birkhäuser, Boston. To appear.

Sherman, M. (1996), "Variance Estimation for Statistics Computed from Spatial Lattice Data," *Journal of the Royal Statistical Society*, Series B 58 509-523.

Sherman, M., and Carlstein, E. (1994), "Nonparametric estimation of the moments of a general statistic computed from spatial data," *Journal of the American Statistical Association 89*, 496-500.

Seymour, L., Smith, R. L., Calloway, M. O., and Morrissey, J. (2000), "Lattice Models for Social Networks With Binary Data," Technical Report #2000-24, Department of Statistics, University of Georgia.

Tjelmeland and Besag, J. (1998), "Markov random fields with higher-order interactions," *Scandinavian Journal of Statistics* 25, 415-433.