Ann. Inst. Statist. Math.
Vol. 54, No. 1, 1–18 (2002)

GENERALIZED PSEUDO-LIKELIHOOD ESTIMATES FOR MARKOV RANDOM FIELDS ON LATTICE

FUCHUN HUANG[1] AND YOSIHIKO OGATA[2]

[1] *School of Computing and Mathematics, Deakin University, Victoria 3168, Australia,*
e-mail: fuchun@deakin.edu.au
[2] *The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan,*
e-mail: ogata@ism.ac.jp

(Received January 30, 1998; revised July 17, 2000)

Abstract. In this paper we generalize Besag's pseudo-likelihood function for spatial statistical models on a region of a lattice. The correspondingly defined maximum generalized pseudo-likelihood estimates (MGPLEs) are natural extensions of Besag's maximum pseudo-likelihood estimate (MPLE). The MGPLEs connect the MPLE and the maximum likelihood estimate. We carry out experimental calculations of the MGPLEs for spatial processes on the lattice. These simulation results clearly show better performances of the MGPLEs than the MPLE, and the performances of differently defined MGPLEs are compared. These are also illustrated by the application to two real data sets.

Key words and phrases: Auto-normal model, Gibbs field, Ising model, pseudo-likelihood.

1. Introduction

Because the maximum likelihood estimate (MLE) is generally difficult to compute due to the normalizing factor of the probability function of spatial statistical models, several alternative estimates have been proposed. Besag (1974, 1975) first proposed the coding method, and then proposed the maximum pseudo-likelihood estimate (MPLE) both owing to their merit of being easy to compute. In Besag (1977), the MPLE is proved to be more efficient than the coding method. Then, many researchers have proved that under suitable conditions the MPLE is consistent and asymptotically normally distributed around the true parameter value for large samples of various spatial processes (Jensen and Møller (1991); Comets (1992); Guyon and Künsch (1992); Jensen and Künsch (1994); Guyon (1995); Mase (1995); etc.). Furthermore, it is applicable to wide range of models. For instance, Besag (1986) applied it to image restoration, and Goulard *et al.* (1996) applied it to marked Gibbs point processes. On the other hand, it has also been shown that the MPLE is not efficient in comparison with the MLE (Besag (1977); Geyer (1991); Geyer and Thompson (1992); Guyon and Künsch (1992); Diggle *et al.* (1994); Huang and Ogata (1999); etc.). According to these studies, the MPLE is as good as MLE in the weak interaction case, but the difference between the two becomes substantial as the interaction becomes strong. For stochastic iteration algorithms to reach the MLE by the Markov chain Monte Carlo (MCMC) method (Penttinen (1984); Younes (1988, 1991); Moyeed and Baddeley (1991); Geyer (1991); Geyer and Thompson

(1992); Seymour and Ji (1996); etc.), a good initial value such as the MPLE is important (c.f. Huang and Ogata (1999)). Therefore, an estimate which is easy to calculate like the MPLE but better than the MPLE in performance will provide a better initial value to reach the MLE.

In this paper, we propose a generalized pseudo-likelihood (GPL) and correspondingly the maximum GPL estimate (MGPLE). The MGPLE is a class of estimates which connect the MLE and MPLE. We implement the calculation of the MGPLE for an Ising model and auto-normal models. In our simulation study for the two models, we consider several GPLs of different sizes. Incidentally we learned that Guyon (1995) suggested the possible extension of the pseudo-likelihood to the GPL in the refereeing process.

Mase (1995) proposed a pseudo-likelihood estimate of the second order for spatial point processes, and proved its strong consistency under some conditions. This pseudo-likelihood of the second order can be similarly defined on a Markov field on lattice. However, this is different from the GPL estimates.

In Section 2, the GPL is defined, and in Sections 3 and 4, the MGPLE is implemented for an Ising model and auto-normal models, and some simulation results are obtained to evaluate the performance of the MGPLE in comparison with the MPLE and MLE. In Section 5, we compute the MGPLEs of two real data sets. Section 6 concludes the paper with some remarks.

2. Generalized pseudo-likelihood

We usually consider a spatial model in terms of energy function U which exhibits the interactions between random variables on the space. The Gibbsian distribution induced by the parameterized energy function $U(\boldsymbol{x};\theta)$ is

$$P_\theta(\boldsymbol{x}) = Z(\theta)^{-1}\exp\{-U(\boldsymbol{x};\theta)\} \tag{2.1}$$

where

$$Z(\theta) = \int_{A^I} \exp\{-U(\boldsymbol{x};\theta)\}\mu(\boldsymbol{dx}) \tag{2.2}$$

is the normalizing factor with $\mu(\boldsymbol{dx})$ being either the Dirac's delta measure $\delta_{\boldsymbol{x}}(\boldsymbol{dx})$ or Lebesgue measure $\boldsymbol{dx}$, according to that $\boldsymbol{x}$ takes discrete or continuous value, respectively, and $\boldsymbol{x} = \{x_{s(i)} \in A : s(1),\ldots,s(I) \in D\}$ is a set of random variables taking values in a subset of $\boldsymbol{R}^s$ on a spatial region D which is a subset of $\boldsymbol{R}^d$. For simplicity we denote $x_{s(i)}$ by x_i in the following. Here we assume the *admissibility condition* $Z(\theta) < \infty$ holds for a set of parameters in order to define the likelihood. Thus the *log likelihood* of parameter θ is given by

$$l(\theta;\boldsymbol{x}) = \log f_\theta(\boldsymbol{x}) = -U(\boldsymbol{x};\theta) - \log Z(\theta). \tag{2.3}$$

Because of the difficulty in evaluating the normalizing factor, the *maximum likelihood estimate* (MLE) has either not been numerically available or needed very intensive computation. For this and other reasons, Besag (1975) proposed the *maximum pseudo-likelihood estimate* (MPLE) which maximizes the direct product of conditional probabilities or conditional probability densities of variable at each site on those at the rest of the sites, namely maximizing the log pseudo-likelihood

$$l_p(\theta;\boldsymbol{x}) = \sum_{i=1}^{I} \log f_\theta(x_i \mid x_j, j \neq i)$$

with respect to θ.

In order to generalize the pseudo-likelihood, define a group $g(i)$ of sites adjacent to each site i, and let $\boldsymbol{x}_{g(i)} := \{x_k : k \in g(i)\}$ and $\boldsymbol{x}^{g(i)} := \{x_k : k \notin g(i)\}$ be the sets of random variables in and out the adjacent sites group $g(i)$, respectively.

For instance, in our simulation experiments in Sections 3 and 4, we consider the following six kinds of adjacent sites to each site $(m, n) \in \mathcal{Z}^2_{M,N}$ of the $M \times N$ 2-dimensional rectangle lattice:

1. $g_{2h}(m, n) = \{(m, n), (m, n + 1)\}$;
2. $g_{2v}(m, n) = \{(m, n), (m + 1, n)\}$;
3. $g_4(m, n) = \{(m, n), (m, n + 1), (m + 1, n + 1), (m + 1, n)\}$;
4. $g_5(m, n) = \{(m, n), (m, n \pm 1), (m \pm 1, n)\}$;
5. $g_9(m, n) = g_5(m, n) \cup \{(m \pm 1, n \pm 1)\}$;
6. $g_{13}(m, n) = g_9(m, n) \cup \{(m, n \pm 2), (m \pm 2, n)\}$.

The generalized pseudo-likelihood (GPL) for the spatial process $\boldsymbol{x}$ is defined by the product of conditional probabilities (or densities) of random variable $\boldsymbol{x}_{g(i)}$ on the rest random variables $\boldsymbol{x}^{g(i)}$, that is,

$$L_g(\theta; \boldsymbol{x}) = \prod_{i=1}^{I} f_\theta(\boldsymbol{x}_{g(i)} \mid \boldsymbol{x}^{g(i)})^{1/|g(i)|} = \prod_{i=1}^{I} f_\theta(\boldsymbol{x}_{g(i)} \mid \boldsymbol{x}^{g(i)})^{1/|g(i)|},$$

where $|g(i)|$ denotes the number of sites in the set $g(i)$. Maximizing the GPL or its logarithm

$$l_g(\theta; \boldsymbol{x}) = \sum_{i=1}^{I} |g(i)|^{-1} \log f_\theta(x_k : k \in g(i) \mid x_k : k \notin g(i))$$

with respect to θ provides the *maximum GPL estimator* (MGPLE). In the case where $g(i) = \{i\}$, the MGPLE is nothing but the MPLE of Besag. In the case where $g(i)$ is the set of all sites for any i, then the MGPLE is the MLE. As $|g(i)|$ becomes larger, the performance of the MGPLE is expected to be closer to that of the MLE, but the calculation complexity will increase exponentially in $|g(i)|$. If variables of a random field are independent to each other, then $l_g(\theta; \boldsymbol{x}) = l_p(\theta; \boldsymbol{x}) = l(\theta; \boldsymbol{x})$, thus all the above estimates become the same.

Hereafter, the set of the adjacent sites $\{g(i) : i = 1, \ldots, I\}$ are taken with the same shape and same size for each site in our study where the periodic boundary condition is assumed. When the edge effect has to be considered, the set of adjacent sites with different sizes may be taken.

Two questions arise naturally, that is, whether the performance of the MGPLE is better than the MPLE and whether its calculation is feasible like the MPLE. In the following sections we calculated MGPLEs for an Ising model and two auto-normal models to compare their performances with those of the MLE and MPLE. Before proceeding to the detailed models, we here note some general remarks as follows. In any case, we need to calculate $f_\theta(\boldsymbol{x}_{g(i)} \mid \boldsymbol{x}^{g(i)})$ for each site i given a parameter θ. Consider the case where the model is given by an energy function $U(\boldsymbol{x}; \theta)$ which is rewritten by $U_i(\boldsymbol{x}_{g(i)}, \boldsymbol{x}^{g(i)}; \theta)$ for each site i. Then we have

$$f_\theta(\boldsymbol{x}_{g(i)} \mid \boldsymbol{x}^{g(i)}) = \frac{\exp\{-U_i(\boldsymbol{x}_{g(i)}, \boldsymbol{x}^{g(i)}; \theta)\}}{\int \exp\{-U_i(\boldsymbol{y}, \boldsymbol{x}^{g(i)}; \theta)\} \mu(\boldsymbol{dy})}.$$

In the case when $U(\boldsymbol{x};\theta)$ is a linear function of the parameter θ: $U(\boldsymbol{x};\theta) = \theta^t V(\boldsymbol{x})$ where θ^t indicates the transpose of a row vector and $V(\boldsymbol{x})$ is a vector of the same dimension of θ, it is easy to see that $\log f_\theta(\boldsymbol{x}_{g(i)} \mid \boldsymbol{x}^{g(i)})$ is concave in θ by the same argument given in Guyon ((1995), §5.1.1) for the likelihood and conditional pseudo-likelihood, and hence the GPL $l_g(\theta;\boldsymbol{x})$ since it's a finite sum of concave functions. This assures the uniqueness of the MGPLE if it exists, and the convergence of the gradient algorithms used in the associated optimization problem.

In the case of the Markov random field model, the conditional probability $f(x_j,\ j \in g(i) \mid$ all other site values) only depends upon $\boldsymbol{x}_{g(i)}$ and also upon the values at sites of a boundary set of $g(i)^c$ to $g(i)$ associated with the Markovian property; specifically, denote this by $\partial g(i)$. Then we have $f(\boldsymbol{x}_{g(i)}|\boldsymbol{x}^{g(i)}) = f(\boldsymbol{x}_{g(i)}|\boldsymbol{x}_{\partial g(i)})$.

Comets (1992) proved the strong consistency of the MPLE and MLE for exponential families of Markov random fields on the lattice without any mixing or ergodicity conditions. The asymptotic normality of the MPLE and MLE is proved under weak dependency condition (see also Jensen and Künsch (1994); Guyon (1995), §4-5). With similar conditions, we expect the consistency and asymptotic normality of the MGPLEs hold. It is important to know the asymptotic variances of the estimators which imply the theoretical asymptotic efficiency relative to the MLE. This is formally given by the following matrix

$$E\left\{-\frac{\partial^2 l_g}{\partial\theta^t\partial\theta}\right\}^{-1} \operatorname{Cov}\left\{\frac{\partial l_g}{\partial\theta^t}, \frac{\partial l_g}{\partial\theta}\right\} \; E\left\{-\frac{\partial^2 l_g}{\partial\theta\partial\theta^t}\right\}^{-1} \tag{2.4}$$

where l_g is log GPL. This is surmised from the classical paradigm of the M-estimators which minimizes a contrast function (Huber (1964, 1967); e.g., Ogata (1980); and Guyon (1995), §3). For the MLE and MPLE of the Ising model and the auto-normal model described in Sections 3 and 4 in this paper, some closed or semiclosed results exist in the literature showing their asymptotic variances and their relative efficiency values over the MLE as functions of correlations between neighbouring variables (e.g., Besag (1977); Guyon and Künsch (1992); Guyon (1995), §4.3.2). To show the superiority of the MGPLE over the MPLE we have also calculated the efficiency values (2.4) of all estimators for the two models studied in Sections 3 and 4, besides the simulation results of their empirical variances.

In the following sections we calculated MGPLEs corresponding to the adjacent patterns $g_{2h}(i)$, $g_{2v}(i)$, $g_4(i)$, $g_5(i)$, $g_9(i)$, $g_{13}(i)$ as shown above for spatial processes on lattice. The corresponding GPLs are denoted by l_{2h}, l_{2v}, l_4, l_5, l_9 and l_{13}, respectively, and the corresponding MGPLEs are denoted by $\hat\theta_{2h}$, $\hat\theta_{2v}$, $\hat\theta_4$, $\hat\theta_5$, $\hat\theta_9$ and $\hat\theta_{13}$, respectively. Since $\hat\theta_{2h}$ and $\hat\theta_{2v}$ are based on anisotropic adjacent sites $g_{2h}(m,n)$ and $g_{2v}(m,n)$, we eventually consider the estimate $\hat\theta_2$ which maximize $l_2 := (l_{2h} + l_{2v})/2$. Hereafter $\hat\theta_m$ and $\hat\theta_p$ denote the MLE and MPLE, respectively.

3. An Ising model

3.1 *Description of the model*

Consider a binary random variable $X_{m,n}(=\pm 1)$ at each site (m,n) on a regular $M \times N$ lattice, $\mathcal{Z}^2_{M,N}$. For realizations $\boldsymbol{x} = \{x_{m,n} : (m,n) \in \mathcal{Z}^2_{M,N}\}$ of the resulting stochastic process $\boldsymbol{X} = \{X_{m,n} : (m,n) \in \mathcal{Z}^2_{M,N}\}$, consider the statistic

$$Q = Q(\boldsymbol{x}) = \sum_{nn} x_{m,n} x_{u,v}$$

where "nn" indicates that only nearest-neighbour pairs of sites contribute the summands. This lattice statistic, which counts the excess of *like*, over *unlike*, nearest-neighbour bonds on the lattice, is of central importance. The joint (Gibbs) distribution for X is then defined by

$$P(\boldsymbol{X} = \boldsymbol{x}) = Z(\theta)^{-1} e^{\theta Q(\boldsymbol{x})} \tag{3.1}$$

where θ is a parameter so that the energy function is $U(\boldsymbol{x};\theta) = -\theta Q(\boldsymbol{x})$. The normalizing constant (or, partition function) is obtained by summing over all configurations, i.e.,

$$Z(\theta) = \sum_{\boldsymbol{x}} e^{\theta Q(\boldsymbol{x})}. \tag{3.2}$$

This Gibbs random field is actually a nearest-neighbour Markov field. To see this, let $\eta_{m,n}$ be the set of sites that are nearest-neighbours of site (m,n). Then direct computation yields

$$\begin{aligned} &P(X_{m,n} = x_{m,n} \mid X_{u,v} = x_{u,v}; (u,v) \neq (m,n)) \\ &= \exp\left\{ \theta x_{m,n} \sum_{\eta_{m,n}} x_{u,v} \right\} \Big/ 2\cosh\left\{ \theta \sum_{\eta_{m,n}} x_{u,v} \right\}, \end{aligned}$$

and hence we can compute the MPLE. To compute the MLE, we use the asymptotic formula given by Onsager (1944) (see Feynman (1972), pp. 148; also Pickard (1987) and Huang and Ogata (2001)). Also note a theoretical comparison of the MPLE and MLE is given in Guyon and Künsch (1992) and Guyon (1995).

For a site $(m,n) \in \mathcal{Z}^2_{M,N}$ and the corresponding set of adjacent sites $g(m,n)$, let $\partial g(m,n)$ be the boundary set of sites $(m',n') \in g(m,n)^c$ which are the nearest neighbours of $g(m,n)$. In the following we also use simplified notations g and ∂_g for $g(m,n)$ and $\partial g(m,n)$, respectively. Let $(\boldsymbol{x}_g, \boldsymbol{x}_{\partial g})$ denote the row vector of variables located in $g \cup \partial g$. Let $Q(\boldsymbol{x})$ be rewritten by $Q_{m,n}(\boldsymbol{x}_g, \boldsymbol{x}^g)$ for each site (m,n). Also let $Q^g_{m,n}(\boldsymbol{x}_g, \boldsymbol{x}_{\partial g}) = \sum_{gnn} x_{m,n} x_{u,v}$ where "gnn" means that only the nearest neighbour pairs of sites within $g \cup \partial g$ contribute summands. Then the conditional probability of variables in g given all the other variables is expressed by

$$\begin{aligned} P(x_{u,v} : (u,v) \in g \mid x_{u,v} : (u,v) \notin g; \theta) &= \frac{\exp\{\theta Q_{m,n}(\boldsymbol{x}_g, \boldsymbol{x}^g)\}}{\sum_{\boldsymbol{y}} \exp\{\theta Q_{m,n}(\boldsymbol{y}, \boldsymbol{x}^g)\}} \\ &= \frac{\exp\{\theta Q^g_{m,n}(\boldsymbol{x}_g, \boldsymbol{x}_{\partial g})\}}{\sum_{\boldsymbol{y}} \exp\{\theta Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}_{\partial g})\}}. \end{aligned}$$

Thus the log GPL is given by

$$l_g(\theta) = \sum_{m,n} |g(m,n)|^{-1} \log \frac{\exp\{\theta Q^g_{m,n}(\boldsymbol{x}_{g(m,n)}, \boldsymbol{x}_{\partial g(m,n)})\}}{\sum_{\boldsymbol{y}} \exp\{\theta Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}_{\partial g(m,n)})\}}.$$

Since the sum in the denominator consists of $2^{|g(m,n)|}$ terms, it is easy to calculate by computer as far as $|g(m,n)|$ is not too large. For example, the largest set of adjacent sites taken in our simulation consists of 13 sites, so that the denominator totally consists of $2^{13} (= 8192)$ summands.

3.2 *Simulation results*

In this simulation study, the Ising model is considered on an $M \times M$ square lattice $\mathcal{Z}^2_{M,M}$ on the plane: $\{X_{m,n} : m, n = 1, \ldots, M\}$ with $M = 64$. We assume the periodic boundary for the square lattice, namely equivalent to a torus such that $X_{0,n} = X_{M,n}$, $X_{M+1,n} = X_{1,n}$, $X_{m,0} = X_{m,M}$, $X_{m,M+1} = X_{m,1}$, for all m and n.

To simulate this process, we use the Metropolis algorithm. For each site (m, n) selected in the lexicographical order, let the current value of the process be $X_{m,n}$, and let the current energy value be U. Take the alternative value here $X^*_{m,n} = -X_{m,n}$ with the corresponding energy value U^*. Then, calculate the ratio $r = \exp(U - U^*)$ and accept the transition or reject it according to the Metropolis rule, i.e., accept with probability $\min(r, 1)$. The initial state of the process is taken at random such that $X_{m,n}$ at each site (m, n) is taken independently to be $+1$ or -1 with probability $1/2$. This process was repeated $320 \times M^2$ times (320 Monte Carlo steps) to ensure that the equilibrium states are achieved.

We take the six kinds of sets of adjacent sites as defined in Section 2. For each parameter we extracted 500 realizations, and from these we obtained 500 respective estimates of the MLE, MPLE and MGPLEs. The five kinds of MGPLEs $\hat{\theta}_2$, $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_9$ and $\hat{\theta}_{13}$ are also defined in Section 2. To calculate the MLE, MPLE and MGPLEs, we used the DALL optimization program by Ishiguro and Akaike (1989) which is an implementation of Davidon's variance algorithm with a numerical derivative evaluation procedure.

Figure 1 shows empirical distributions of the values of $l(\hat{\theta}_m) - l(\hat{\theta}_p)$, $l(\hat{\theta}_m) - l(\hat{\theta}_2)$, $l(\hat{\theta}_m) - l(\hat{\theta}_4)$, $l(\hat{\theta}_m) - l(\hat{\theta}_5)$, $l(\hat{\theta}_m) - l(\hat{\theta}_9)$ and $l(\hat{\theta}_m) - l(\hat{\theta}_{13})$ which are marked by "p", "2", "4", "5", "9", "B", respectively. Since $l(\hat{\theta}_m)$ takes the maximum values; all the values in the diagrams are positive, and smaller value means closer to the MLE in the sense of log-likelihood increment. We see that all the five MGPLEs are closer to the MLE

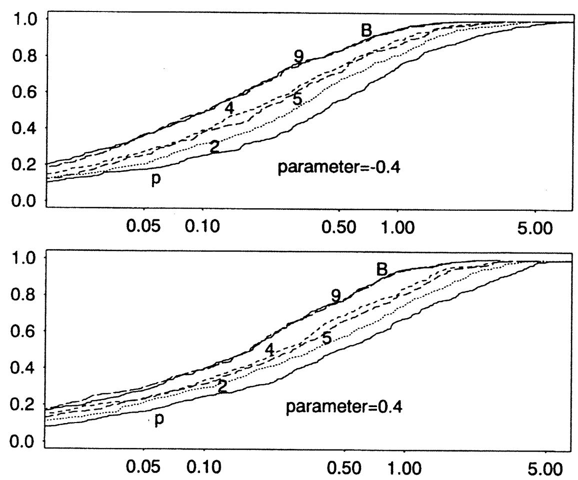

Fig. 1. Empirical distributions of the deviations of the log-likelihood of the MPLE and the five MGPLEs from that of the MLE of the Ising model, that is, empirical distributions of the values of $l(\hat{\theta}_m) - l(\hat{\theta}_p)$, $l(\hat{\theta}_m) - l(\hat{\theta}_2)$, $l(\hat{\theta}_m) - l(\hat{\theta}_4)$, $l(\hat{\theta}_m) - l(\hat{\theta}_5)$, $l(\hat{\theta}_m) - l(\hat{\theta}_9)$ and $l(\hat{\theta}_m) - l(\hat{\theta}_{13})$ which are marked by "p", "2", "4", "5", "9" and "B", respectively, in the cases where $\theta_0 = -0.4$ (upper) and 0.4 (lower).

Table 1. Root mean square errors ($\times 10^2$) and relative efficiency values of the MLE, MPLE and five other MGPLE estimators of the Ising model.

θ		$\hat{\theta}_p$	$\hat{\theta}_2$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_9$	$\hat{\theta}_{13}$	$\hat{\theta}_m$
−0.400	*r.m.s.e.*	1.092	0.975	0.884	0.905	0.807	0.803	0.641
	re.eff.	0.588	0.658	0.726	0.709	0.795	0.799	1.000
−0.300	*r.m.s.e.*	1.098	1.005	0.929	0.956	0.893	0.900	0.873
	re.eff.	0.795	0.869	0.940	0.913	0.978	0.970	1.000
−0.200	*r.m.s.e.*	1.120	1.066	1.040	1.051	1.034	1.038	1.038
	re.eff.	0.927	0.974	0.998	0.988	1.004	1.000	1.000
−0.100	*r.m.s.e.*	1.061	1.047	1.039	1.042	1.036	1.037	1.034
	re.eff.	0.974	0.987	0.994	0.992	0.998	0.997	1.000
0.000	*r.m.s.e.*	1.115	1.112	1.111	1.112	1.110	1.111	1.110
	re.eff.	0.996	0.998	0.999	0.998	1.000	0.999	1.000
0.100	*r.m.s.e.*	1.098	1.077	1.065	1.073	1.063	1.066	1.062
	re.eff.	0.967	0.986	0.997	0.990	0.999	0.996	1.000
0.200	*r.m.s.e.*	1.127	1.080	1.052	1.061	1.045	1.046	1.045
	re.eff.	0.928	0.968	0.994	0.985	1.000	1.000	1.000
0.300	*r.m.s.e.*	1.102	1.013	0.949	0.971	0.912	0.918	0.887
	re.eff.	0.805	0.876	0.935	0.913	0.973	0.967	1.000
0.400	*r.m.s.e.*	1.183	1.053	0.941	0.974	0.844	0.846	0.667
	re.eff.	0.563	0.633	0.708	0.685	0.790	0.788	1.000

than the MPLE. As the number of adjacent sites $|g(i)|$ increases, the corresponding estimate appears to get closer to the MLE. But it should be noted that we see the similar performance between $\hat{\theta}_4$ and $\hat{\theta}_5$ and also between $\hat{\theta}_9$ and $\hat{\theta}_{13}$.

In order to compare the performance in more detail, we hereafter consider the *mean square error* of an estimator $\hat{\theta}$ which is defined by $\sum |\hat{\theta}_i - \theta_0|^2/n$ where θ_0 is the true parameter and $|\cdot|$ means the Euclidean distance, and also the *relative efficiency* of an estimator which is the ratio of the mean square error of the MLE to that of the considered estimator. Table 1 gives the mean square errors and relative efficiency values of the seven estimators of parameter values $\theta = 0.0, \pm 0.1, \ldots, \pm 0.4$ for the Ising model, and Fig. 2 shows the results graphically, with the upper plot showing the mean square errors and the lower one the relative efficiency values. Note that for the Ising model on the whole space $\mathcal{Z}^2$ there are phase transitions at the critical values $\theta \approx 0.441$ or -0.441, and the interaction becomes strong when the parameter θ approaches the critical values. We can clearly see by Fig. 2 that, as the interaction becomes strong, the differences among the seven estimates becomes substantial. The MLE and MPLE are the best and the worst among the seven, respectively, and the performances of the estimators roughly ranked in the reverse order of the sizes of their adjacent sites. In the cases of the strongest interaction ($\theta_0 = 0.4$ or -0.4) among the simulated parameters, the relative efficiency values of the MPLE and the MGPLE $\hat{\theta}_{13}$ are about 0.6 and 0.8, respectively. Finally, it should be noted that we again see similar performances between $\hat{\theta}_4$ and $\hat{\theta}_5$ and also between $\hat{\theta}_9$ and $\hat{\theta}_{13}$.

To examine the behaviour of the log GPL functions, we have also calculated the surmised theoretical efficiency (2.4) by the MCMC method. Here we generated 1000 samples

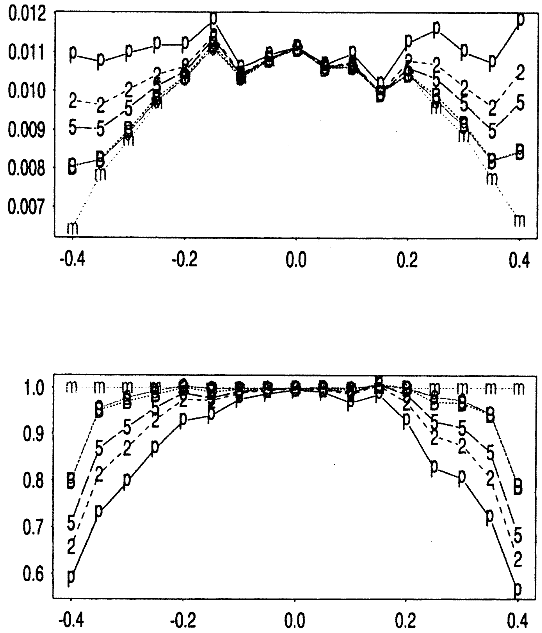

Fig. 2. The mean square errors (upper) and relative efficiency (lower) values of the MLE, the MPLE and five MGPLEs of the Ising model marked by "m", "p", "2", "4", "5", "9" and "B" for estimators $\hat{\theta}_m$, $\hat{\theta}_p$, $\hat{\theta}_2$, $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_9$ and $\hat{\theta}_{13}$, respectively. The performances of $\hat{\theta}_9$ and $\hat{\theta}_{13}$ are almost the same, so most of their curves are overlapped in the figure.

$\{\boldsymbol{x}^{(t)} : t = 1, \dots, 1000\}$ corresponding to the true parameter $\theta = 0.0, 0.05, 0.1, \dots, 0.4$ in the same way as stated above and for each of them we calculated the first and second order differentials of the log GPL functions

$$\frac{dl_g}{d\theta} = \sum_{m,n} |g(m,n)|^{-1} \Bigg\{ Q^g_{m,n}(\boldsymbol{x}^{(t)}_{g(m,n)}, \boldsymbol{x}^{(t)}_{\partial g(m,n)}) - \frac{\sum_{\boldsymbol{y}} \{Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)}) \exp\{\theta Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)})\}\}}{\sum_{\boldsymbol{y}} \exp\{\theta Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)})\}} \Bigg\},$$

$$\frac{d^2 l_g}{d\theta^2} = \sum_{m,n} |g(m,n)|^{-1} \Bigg\{ \frac{\sum_{\boldsymbol{y}} Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)})^2 \exp\{\theta Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)})\}}{\sum_{\boldsymbol{y}} \exp\{\theta Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)})\}} - \left\{ \frac{\sum_{\boldsymbol{y}} \{Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)})^2 \exp\{\theta Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)})\}\}}{\sum_{\boldsymbol{y}} \exp\{\theta Q^g_{m,n}(\boldsymbol{y}, \boldsymbol{x}^{(t)}_{\partial g(m,n)})\}} \right\}^2 \Bigg\},$$

and hence the approximated efficiency values using sample average instead of expectation in formula (2.4). Figure 3 shows the calculation results. We can see that they are consistent with the results in Fig. 2.

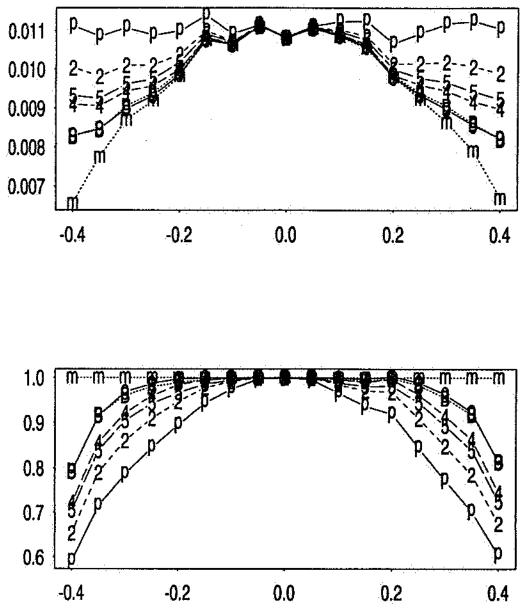

Fig. 3. The asymptotic variance (upper) and asymptotic relative efficiency (lower) values calculated by formula (2.4) of the MLE, the MPLE and five MGPLEs of the Ising model marked by "m", "p", "2", "4", "5", "9" and "B" for estimators $\hat{\theta}_m$, $\hat{\theta}_p$, $\hat{\theta}_2$, $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_9$ and $\hat{\theta}_{13}$, respectively.

4. Auto-normal models

4.1 *Description of the model*

Consider an auto-normal model (Besag (1974)) on a lattice of $M \times N$ grids, $\mathcal{Z}^2_{M,N}$, whose conditional probability density at a site $(m,n) \in \mathcal{Z}^2_{M,N}$ on the rest of sites of the lattice $(u,v) \in \mathcal{Z}^2_{M,N}$ is given by

$$(4.1)\quad f(x_{m,n} \mid x_{u,v} : (u,v) \neq (m,n)) = (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} \left\{ x_{m,n} - \mu_{m,n} - \sum_{(u,v)} \beta_{m,n;u,v}(x_{u,v} - \mu_{u,v}) \right\}^2 \right\},$$

where σ, $\mu_{m,n}$ and $\beta_{m,n;u,v}$'s are parameters. The joint probability density of the process X on the lattice can be written by

$$(4.2)\qquad f(\boldsymbol{x}) = (2\pi\sigma^2)^{-MN/2}|B|^{1/2} \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{x} - \mu)' B(\boldsymbol{x} - \mu) \right\}$$

where $B = (-\beta_{m,n;u,v})$ with $-\beta_{m,n;m,n} = 1$ for $m = 0, \ldots, M$ and $n = 0, \ldots, N$. The maximum likelihood estimator for a general Markov random field model of the form (4.2) is not easy to calculate due to the difficulty of evaluating the normalizing constant, since generally B is a very high dimensional matrix (e.g., 4096×4096 matrix corresponding to the small model of 64×64 square lattice). However, if we assume a torus lattice process, i.e., identifying $(0,n)$ with (M,n) and $(m,0)$ with (m,N), then $|B|$ can be represented by

$$|B| = \prod_{m=0}^{M-1}\prod_{n=0}^{N-1}\left\{1 - \sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\beta_{m,n;u,v}\cos(\omega_m u + \eta_n v)\right\} \tag{4.3}$$

where $\omega_m = 2\pi m/M$ and $\eta_n = 2\pi n/N$, owing to Besag and Moran (1975). See also Cressie (1993) for more details.

The pseudo-likelihood of the model (4.1) is easily given by the mere products of the transition probabilities

$$\prod_{m,n} f(x_{m,n} \mid x_{u,v} : (u,v) \neq (m,n)) = (2\pi\sigma^2)^{-MN/2} \tag{4.4}$$

$$\cdot \exp\left\{-\frac{1}{2}\sigma^{-2}\sum_{m,n}\left\{x_{m,n} - \mu - \sum_{(u,v)\neq(m,n)}\beta_{m,n;u,v}(x_{u,v} - \mu_{u,v})\right\}^2\right\},$$

so that the MPLE of $\beta_{m,n;u,v}$ in this case is obtained by minimizing the following sum of squares,

$$\sum_{m,n}\left\{x_{m,n} - \mu - \sum_{(u,v)\neq(m,n)}\beta_{m,n;u,v}(x_{u,v} - \mu_{u,v})\right\}^2.$$

For a site $(m,n) \in \mathcal{Z}^2_{M,N}$ and the corresponding set of adjacent sites $g(m,n)$, let $\partial g(m,n)$ be the set of sites $(m',n') \in g(m,n)^c$ which are the nearest neighbours of $g(m,n)$. In the following we also use simplified notations g and ∂_g for $g(m,n)$ and $\partial g(m,n)$, respectively. Let $(\boldsymbol{x}_g, \boldsymbol{x}_{\partial g})$ be the row vector of variables located in $g \cup \partial g$. For simplicity, we here assume that the mean parameter of the process $\mu = 0$. Let

$$B_{m,n} = \begin{pmatrix} B_{gg} & B_{g\partial g} \\ B_{\partial gg} & B_{\partial g\partial g}\end{pmatrix}$$

be the matrix composed of the elements of B in (4.2) but rearranged corresponding to the random vector $(\boldsymbol{x}_g, \boldsymbol{x}_{\partial g})$, then

$$\begin{aligned} f(\boldsymbol{x}_g \mid \boldsymbol{x}^g) &= f(\boldsymbol{x}_g \mid \boldsymbol{x}_{\partial g}) = \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x}_g, \boldsymbol{x}_{\partial g})B_{m,n}(\boldsymbol{x}_g, \boldsymbol{x}_{\partial g})^t\right\} \\ &\qquad \bigg/ \int \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}, \boldsymbol{x}_{\partial g})B_{m,n}(\boldsymbol{y}, \boldsymbol{x}_{\partial g})^t\right\} d\boldsymbol{y} \\ &= (2\pi\sigma^2)^{-|g|/2}|B_{gg}|^{1/2} \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\{\boldsymbol{x}_g + \boldsymbol{x}_{\partial g}B_{\partial gg}B_{gg}^{-1}\}B_{gg}\{\boldsymbol{x}_g + \boldsymbol{x}_{\partial g}^t B_{g\partial g}^t B_{gg}^{-t}\}^t\right\}, \end{aligned}$$

where V^t and V^{-t} mean transposes of matrixes V and V^{-1} respectively. Thus the GPL in this case becomes

$$L_g(B) = \prod_{m,n} f(\boldsymbol{x}_{g(m,n)} \mid \boldsymbol{x}^{g(m,n)})^{1/|g(m,n)|} = (2\pi\sigma^2)^{-MN/2}\prod_{m,n}|B_{gg}|^{1/|g(m,n)|} \tag{4.5}$$

$$\cdot \exp\left\{ -\frac{1}{2\sigma^2|g(m,n)|} \times \{\boldsymbol{x}_{g(m,n)} + \boldsymbol{x}_{\partial g(m,n)} B_{\partial gg} B_{gg}^{-1}\} \cdot B_{gg}\{\boldsymbol{x}_{g(m,n)} + \boldsymbol{x}^t_{\partial g(m,n)} B^t_{g\partial g} B_{gg}^{-t}\}^t \right\}.$$

Note here that B_{gg} and $B_{g\partial g}$ depend on $\{\beta_{m,n;u,v} : (m,n) \in g, (u,v) \in g\}$ and $\{\beta_{m,n;u,v} : (m,n) \in g, (u,v) \in \partial g\}$, respectively. Therefore the log GPL is generally not quadratic with respect to the parameters $\beta_{m,n;u,v}$, except in the special case of the log pseudo-likelihood. Nevertheless the log GPL of the models in the following sections is concave in the canonical parameter, which is $(\frac{1}{\sigma^2}, \frac{\beta}{\sigma^2})$ and $(\frac{1}{\sigma^2}, \frac{\beta_1}{\sigma^2}, \frac{\beta_2}{\sigma^2})$ for the first order model and the second order model in Subsections 4.2 and 4.3 respectively, which assures the convergence of the numerical optimization needed to compute the MGPLE.

4.2 *Simulation results of the first order auto-normal model*

In the simulation study, we take the Gaussian Markov process on $M \times N$ lattice with $M = N = 64$. Again, in order to avoid the edge effect, we identify the finite region of lattice with a torus like in the case of Ising model. Assume here that $\mu = 0$, $\beta_{m,n;u,v} = \beta$ for the nearest neighbour sites of (m,n) and $\beta_{m,n;u,v} = 0$ otherwise. The conditional probability density of $X_{m,n}$ given the values of the rest variables for this particular model is

$$(2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2}(X_{m,n} - \beta X^*_{m,n})^2 \right\}$$

where $X^*_{m,n} := X_{m-1,n} + X_{m+1,n} + X_{m,n-1} + X_{m,n+1}$. Thus there are two parameters β and σ to be estimated. In our simulation, β ranges from 0 to 0.245, while the true σ value is set to be always 1.0. Both β and σ are assumed unknown when these are estimated. Notice that stationarity of the auto-normal model requires $|\beta| < 1/4$, thus $\beta = 0.245$ is close to the critical value and represents strong interaction among nearest neighbours.

We use the Gibbs sampling to generate the process. That is, we choose a site (m,n) in the lexicographical order and replace $X_{m,n}$ by the value of $\beta X^*_{m,n} + \epsilon_{m,n}$ where $\epsilon_{m,n}$ is independently generated from $N(0, \sigma^2)$. We start with setting $X_{m,n} = 0$ for all sites (m,n) and then iterate this process 320 times to get equilibrium sample realizations.

We take the six kinds of sets of adjacent sites as defined in Section 2. For each parameter β, we generated 500 realizations of the first order auto-normal model and thus obtained 500 MLE $\theta_m = (\hat{\beta}_m, \hat{\sigma}_m)$, MPLE $\theta_p = (\hat{\beta}_p, \hat{\sigma}_p)$ and MGPLE estimates of $\theta_2 = (\hat{\beta}_2, \hat{\sigma}_2)$, $\theta_4 = (\hat{\beta}_4, \hat{\sigma}_4)$, $\theta_5 = (\hat{\beta}_5, \hat{\sigma}_5)$, $\theta_9 = (\hat{\beta}_9, \hat{\sigma}_9)$ and $\theta_{13} = (\hat{\beta}_{13}, \hat{\sigma}_{13})$ which are defined in Section 2.

Figure 4 shows empirical distributions of deviations of the log-likelihood values of the MPLE and the five MGPLEs from the corresponding maximum log-likelihood values for the parameter $\beta_0 = 0.245$, that is, the empirical distributions of the values $l(\theta_m) - l(\theta_.)$ for each estimator. We see that all MGPLEs are better than the MPLE, and their performances are roughly ranked in the reverse order of the number of the taken adjacent sites.

For this particular model with true parameters β and σ, the asymptotic variances of the MLEs $\hat{\beta}_m$ and $\hat{\sigma}_m$, and the MPLEs $\hat{\beta}_p$ and $\hat{\sigma}_p$ are analytically given,

$$\lim_{M,N\to\infty} (MN \operatorname{var}(\hat{\beta}_m)) = \frac{1}{2}(\varphi(\beta) - 4V(\beta)^2)^{-1}, \tag{4.6}$$

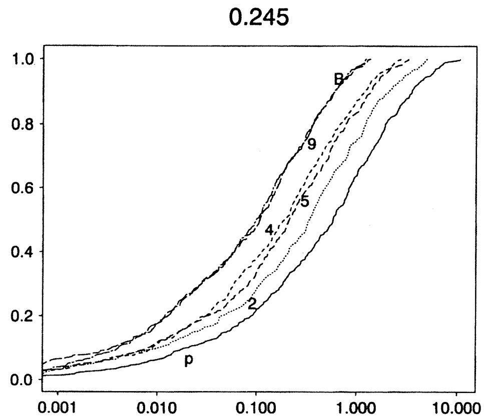

Fig. 4. Empirical distributions of deviations of the log-likelihood of the MPLE and the five MGPLEs from that of the MLE of the first order auto-normal model, that is, empirical distributions of the values of $l(\hat{\theta}_m)-l(\hat{\theta}_p)$, $l(\hat{\theta}_m)-l(\hat{\theta}_2)$, $l(\hat{\theta}_m)-l(\hat{\theta}_4)$, $l(\hat{\theta}_m)-l(\hat{\theta}_5)$, $l(\hat{\theta}_m)-l(\hat{\theta}_9)$ and $l(\hat{\theta}_m) - l(\hat{\theta}_{13})$ which are marked by "m", "p", "2", "4", "5", "9" and "B", respectively, in the case where $(\beta, \sigma) = (0.245, 1.0)$.

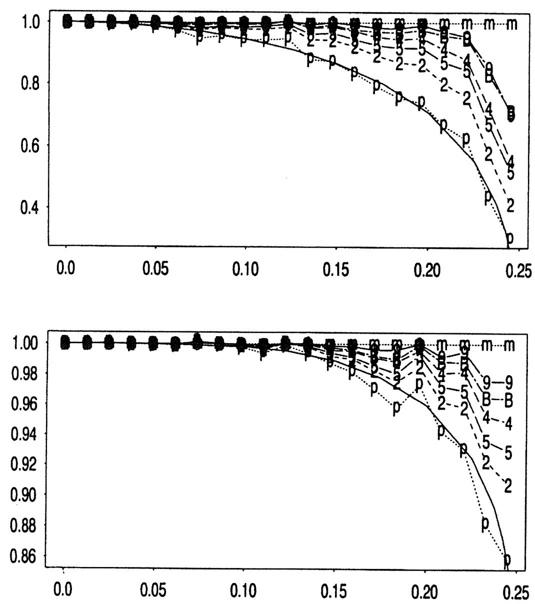

Fig. 5. The relative efficiency values of the MPLE and the five MGPLEs of parameter β (upper) and σ (lower), marked by "m", "p", "2", "4", "5", "9" and "B" for the estimators $\hat{\theta}_m$, $\hat{\theta}_p$, $\hat{\theta}_2$, $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_9$ and $\hat{\theta}_{13}$, respectively, in the first order auto-normal model. The smooth real curve near to the plots of the MPLE is due to the theoretical values given in Besag (1977).

$$\lim_{M,N\to\infty}(MN \operatorname{var}(\hat{\sigma}_m)) = 2\sigma^2\varphi(\beta)(\varphi(\beta) - 4V(\beta)^2)^{-1}, \tag{4.7}$$

where

$$\varphi(\beta) = (4\pi^2)^{-1}\int_0^{2\pi}\int_0^{2\pi}\frac{(\cos x + \cos y)^2}{\{1 - 2\beta(\cos x + \cos y)\}^2}dxdy,$$

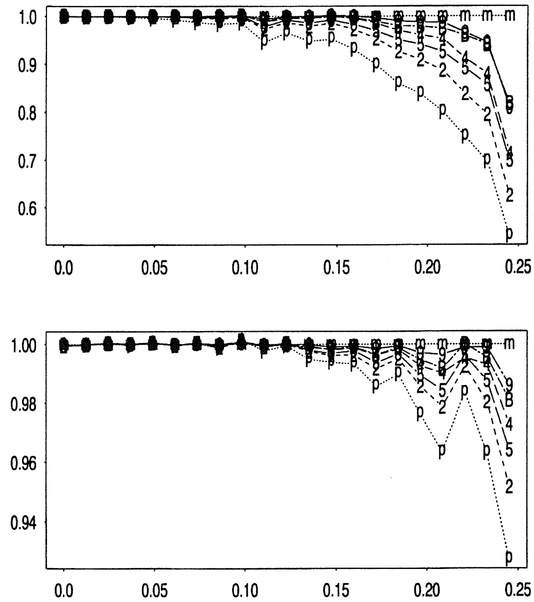

Fig. 6. The asymptotic relative efficiency values calculated by formula (2.4), of the MPLE and the five MGPLEs of parameter β (upper) and σ (lower), marked by "m", "p", "2", "4", "5", "9" and "B" for the estimators $\hat{\theta}_m$, $\hat{\theta}_p$, $\hat{\theta}_2$, $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_9$ and $\hat{\theta}_{13}$, respectively, in the first order auto-normal model.

$$V(\beta) = (4\pi^2)^{-1} \int_0^{2\pi} \int_0^{2\pi} \frac{\cos x}{1 - 2\beta(\cos x + \cos y)} dxdy,$$

and

$$\lim_{M,N\to\infty} (MN \operatorname{var}(\hat{\beta}_p)) = \frac{1}{2}\beta^2(1 - 4\beta\rho(\beta))^2/\rho(\beta)^2 \text{ if } \beta \neq 0, = \frac{1}{2} \text{ if } \beta = 0, \tag{4.8}$$

$$\lim_{M,N\to\infty} (MN \operatorname{var}(\hat{\sigma}_p)) = 2\sigma^2(1 + 4\beta^2), \tag{4.9}$$

where $\rho(\beta)$ is the correlation between two nearest neighbours which is 0 if $\beta = 0$, or given as $\rho(\beta) = (4\beta)^{-1} - \pi(8\beta K(4\beta))^{-1}$ if $\beta \neq 0$ with $K(x) = \int_0^{\pi/2}(1 - x^2 \sin^2 t)^{-1/2} dt$ (for $0 \leq x < 1$) being the complete elliptic integral of the first kind. (See Besag and Moran (1975); Besag (1977); Guyon (1995), §5.3.4.1, §1.4.2. There are some typos in the formulae for the asymptotic variances of the MLEs and MPLEs in these references.). Therefore the relative efficiencies of the MPLE $\hat{\beta}_p$ over the MLE $\hat{\beta}_m$, which is $\lim \operatorname{var}(\hat{\beta}_m)/\operatorname{var}(\hat{\beta}_p)$, and of the MPLE $\hat{\sigma}_p$ over the MLE $\hat{\sigma}_m$, which is $\lim \operatorname{var}(\hat{\sigma}_m)/\operatorname{var}(\hat{\sigma}_p)$, are also available numerically.

Figure 5 shows relative efficiency values of the seven estimators for β (upper) and σ (lower) corresponding to parameter values $\beta_0 = i \times 0.245/20$ ($i = 0, \ldots, 20$) in the first order auto-normal model. The smooth real curve near the plot of the MPLE is got from the theoretical results (4.6)–(4.9).

We can see clearly that, as the interaction becomes strong, the differences among the seven estimators becomes larger. The MLE and MPLE are the best one and the worst one among the seven, respectively, and the performances are roughly ranked in the reverse order of the sizes of their corresponding adjacent sites. It may be surprising that $\hat{\theta}_9$ has better performance than $\hat{\theta}_{13}$, and $\hat{\theta}_4$ has better performance than $\hat{\theta}_5$, despite the latter ones take more adjacent sites than the formers.

To examine the behaviour of the log GPL functions, we have also calculated the surmised theoretical efficiency (2.4) by the MCMC method. Here we generated 1000 samples $\{\boldsymbol{x}^{(t)} : t = 1, \dots, 1000\}$ corresponding to the true parameter $\beta = 0.245 * i/20$ $(i = 0, \dots, 20)$ and $\sigma = 1$, and for each sample corresponding to β we calculated values of various log GPL $l_g(\beta + u\epsilon, \sigma + v\epsilon)$ for $u, v = -1, 0, 1$ and $\epsilon = 10^{-6}$. We then calculated various partial differentials using numerical approximation and summed them or their squared values to get approximations of the expectation values in (2.4). Here we used numerical differentials because we found the theoretical differentials are difficult to calculate unlike in the case of Ising model. Figure 6 shows the calculation result. We can see that they are consistent with the results in Fig. 5, although there are some roughness possibly due to the numerical differentials.

4.3 *Simulation results of the second order auto-normal model*

We also carried out the simulation study for the second order auto-normal model, that is, $\beta_{m,n;u,v} = \beta_1$ for the nearest neighbour site (u, v) satisfying $\text{mod}(m-u, M) + \text{mod}(n-v, N) = 1$; $\beta_{m,n;u,v} = \beta_2$ for the second nearest neighbour site (u, v) satisfying $\text{mod}(m-u, M) = 1$ and $\text{mod}(n-v, N) = 1$; $\beta_{m,n;u,v} = 0$ for the other sites (u, v). All other specifications are the same as in the case of the first order auto-normal model in the previous subsection. The conditional probability density of $X_{m,n}$ given values of the rest variables for this second order model is

$$(2\pi\sigma^2)^{-1} \exp\left\{-\frac{1}{2\sigma^2}(X_{m,n} - \beta_1 X^*_{m,n} - \beta_2 X^{**}_{m,n})^2\right\}$$

where $X^*_{m,n} := X_{m-1,n} + X_{m+1,n} + X_{m,n-1} + X_{m,n+1}$ and $X^{**}_{m,n} := X_{m-1,n-1} + X_{m+1,n-1} + X_{m-1,n+1} + X_{m+1,n+1}$. Thus we estimate three parameters β_1, β_2 and σ simultaneously. Note that a sufficient condition for the stationarity of this model is

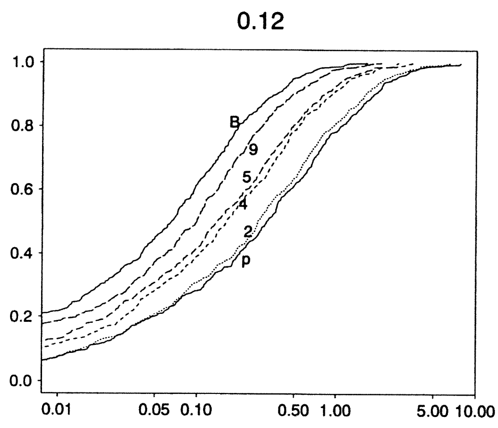

Fig. 7. Empirical distributions of deviations of the log-likelihood of the MPLE and five MGPLEs from that of the MLE of the second order auto-normal model, that is, empirical distributions of the values of $l(\hat{\theta}_m) - l(\hat{\theta}_p)$, $l(\hat{\theta}_m) - l(\hat{\theta}_2)$, $l(\hat{\theta}_m) - l(\hat{\theta}_4)$, $l(\hat{\theta}_m) - l(\hat{\theta}_5)$, $l(\hat{\theta}_m) - l(\hat{\theta}_9)$ and $l(\hat{\theta}_m) - l(\hat{\theta}_{13})$ which are marked by "p", "2", "4", "5", "9" and "B", respectively, in the case where $(\beta_1, \beta_2, \sigma) = (0.12, 0.12, 1.00)$.

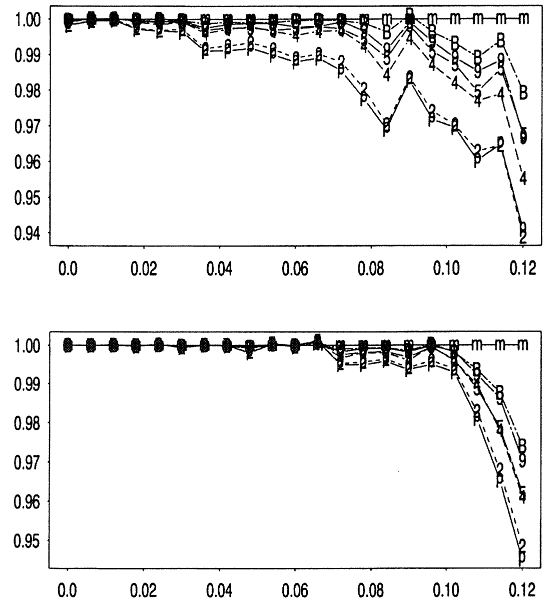

Fig. 8. The relative efficiency values of the MPLE and five MGPLEs of the (β_1, β_2) (upper) and σ (lower), marked by "m", "p", "2", "4", "5", "9" and "B" for the estimators $\hat{\theta}_m$, $\hat{\theta}_p$, $\hat{\theta}_2$, $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_9$ and $\hat{\theta}_{13}$, respectively, of the second order auto-normal model.

$|\beta_1| + |\beta_2| < 0.25$, and the canonical parameter is $(\frac{1}{\sigma^2}, \frac{\beta_1}{\sigma^2}, \frac{\beta_2}{\sigma^2})$ on which the GPLs are concave. In the generation of samples, we set $\beta_1 = \beta_2$ ranging from 0.00 to 0.12, while the true σ value is always set to be 1.0.

Figure 7 shows empirical distributions of the deviations of log-likelihood values of the MPLE and the five MGPLEs from the corresponding maximum log-likelihood values for the parameter $\beta_1 = \beta_2 = 0.12$. It should be noted, in comparison with the results in Fig. 4, that $\hat{\theta}_5$ and $\hat{\theta}_{13}$ (hereafter $\hat{\theta}.$ denotes $(\hat{\beta}_1., \hat{\beta}_2., \hat{\sigma}.)$ for various estimators) now become significantly better than $\hat{\theta}_4$ and $\hat{\theta}_9$, respectively.

Figure 8 shows the relative efficiency values of the MLE, the MPLE and five MGPLEs of the (β_1, β_2) (upper) and σ (lower) marked by "m", "p", "2", "4", "5", "9" and "B" for $\hat{\theta}_m$, $\hat{\theta}_p$, $\hat{\theta}_2$, $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_9$ and $\hat{\theta}_{13}$, respectively. We see similar features of the results to the case of the first order auto-normal processes but with contrasting result in that $\hat{\theta}_5$ and $\hat{\theta}_{13}$ become significantly better than $\hat{\theta}_4$ and $\hat{\theta}_9$, respectively.

5. Illustration by two real data sets

In this section we illustrate the behaviour of the MGPLEs by two well cited real data sets. The data sets used are Mercer and Hall's wheat yield data (Mercer and Hall (1911)) and Wiebe's wheat yield data (Wiebe (1935)). These data and references of related research are collected in Andrews and Herzberg (1985).

5.1 *Ising model applied to Wiebe's data*

We first transferred the original data of Wiebe's wheat yield values into binary data of the two values of "1" for "larger than or equal to the median value" and "−1" for "less

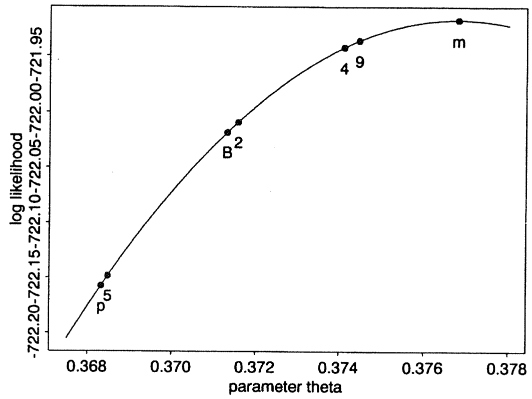

Fig. 9. The log-likelihood curve and the seven estimators of the parameter θ in a 125×12 Ising model for the binary data of the two values of "1" for "larger than or equal to the median value" and "-1" for "less than the median value" of the Wiebe's wheat yield data. The seven estimated values of the estimators $\hat{\theta}_m$, $\hat{\theta}_p$, $\hat{\theta}_2$, $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_9$ and $\hat{\theta}_{13}$ are marked by "m", "p", "2", "4", "5", "9" and "B", respectively.

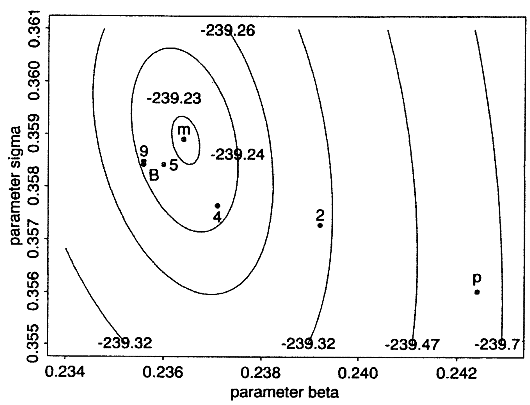

Fig. 10. The contour plot of the log-likelihood functions and the estimated values of the parameter (β, σ) in a first order 20×25 auto-normal model for Mercer and Hall's data. The estimated values of the estimators $(\hat{\beta}_m, \hat{\sigma}_m)$, $(\hat{\beta}_p, \hat{\sigma}_p)$, $(\hat{\beta}_2, \hat{\sigma}_2)$, $(\hat{\beta}_4, \hat{\sigma}_4)$, $(\hat{\beta}_5, \hat{\sigma}_5)$, $(\hat{\beta}_9, \hat{\sigma}_9)$ and $(\hat{\beta}_{13}, \hat{\sigma}_{13})$ are marked by "m", "p", "2", "4", "5", "9" and "B", respectively.

than the median value", so that we can use an Ising model on 125×12 rectangle lattice described in Subsection 3.1 of this paper. Assuming the periodic boundary condition in the data for convenience, we calculated the seven estimators described in Section 3.

Figure 9 shows the curve of the log likelihood function superimposed the locations of the MLE, MPLE and MGPLEs with the corresponding function values. We can see that all the MGPLEs are closer to the MLE than the MPLE, and so are their log likelihood values.

5.2 *The auto-normal model applied to Mercer and Hall's data*

Next we fitted the first order auto-normal model to the Mercer and Hall's wheat yield data on a 20×25 rectangle lattice. We first subtracted the data with its overall mean value, which is equivalent to adding a shift parameter to the auto-normal model (Besag and Moran (1975)). With the periodic boundary assumption, we calculated the estimators described in Subsection 4.1.

Figure 10 shows the contour plot on the parameter region, $[0.234, 0.243] \times [0.355, 0.361]$, for (β, σ), superimposed the locations of the MLE, MPLE and MGPLEs: $(\hat{\beta}_m, \hat{\sigma}_m)$, $(\hat{\beta}_p, \hat{\sigma}_p)$, $(\hat{\beta}_2, \hat{\sigma}_2)$, $(\hat{\beta}_4, \hat{\sigma}_4)$, $(\hat{\beta}_5, \hat{\sigma}_5)$, $(\hat{\beta}_9, \hat{\sigma}_9)$ and $(\hat{\beta}_{13}, \hat{\sigma}_{13})$ which are marked by "m", "p", "2", "4", "5", "9" and "B", respectively, in the figure. We can see that all MGPLEs located closer to the MLE than the MPLE, and so are their log-likelihood values.

6. Concluding remarks

The maximum generalized pseudo-likelihood estimator (MGPLE) is a class of estimators including and connecting the maximum pseudo-likelihood estimator (MPLE) and the maximum likelihood estimator (MLE). From our experiments of the Ising model and auto-normal models, the performance of the MGPLEs is clearly better than MPLE, and the performance becomes better as the size of the taken adjacent sites increases. On the other hand, as the size of the adjacent sites increases, the computing complexity for the MGPLE increases exponentially. However, for those MGPLEs that take small size of adjacent sites, the calculation is as easy as for the MPLE, and the performance of the MGPLEs can be remarkably better than the MPLE as shown in our experiments. Since the MPLE is already widely used for various models in spatial statistics, and also since one can use the MPLE as a reasonable initial value to reach the MLE by a single Newton-Raphson step using Markov chain Monte Carlo method (cf. Huang and Ogata, (1999)), the MGPLE may widen the door of such approaches.

Acknowledgements

We thank the anonymous referees for their critical comments and constructive suggestions, and Dr. H. T. Zhu of the University of Victory for some helpful communications, which improved the paper. This study was mainly carried out when the first author was studying at the Department of Statistical Science, The Graduate University for Advanced Studies, Tokyo, with the support of Monbusho Scholarship. This study was also partially supported by the ISM Cooperative Research Program (98-ISM.CRP-B-13).

References

Andrews, D. F. and Herzberg, A. M. (1985). *Data*, Springer, New York.

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

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

Besag, J. E. (1977). Efficiency of pseudolikelihood estimation for simple Gaussian fields, *Biometrika*, **64**, 616–618.

Besag, J. E. (1986). On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B*, **48**, 259–302.

Besag, J. E. and Moran, P. A. P. (1975). On the estimation and testing of spatial interaction in Gaussian lattice processes, *Biometrika*, **62**, 555–562.

Comets, F. (1992). On consistency of a class of estimators for exponential families of Markov random fields on the lattice, *Ann. Statist.*, **20**, 455–468.

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

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes, *International Statistical Review*, **62**, 99–117.

Feynman, R. P. (1972). *Statistical Mechanics: A Set of Lectures*, Benjamin/Cummings, Reading, Massachusetts.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood, *Computing Science and Statistics: 23rd Symposium on the Interface*, (ed. E. M. Keramidas), 156–163, Interface Foundation of North America, Fairfax Station.

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

Goulard, M., Särkkä, A. and Grabarnik, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudolikelihood method, *Scandinavian Journal of Statistics*, **23**, 365–379.

Guyon, X. (1995). *Random Fields on a Network: Modeling, Statistics and Applications*, Springer, New York.

Guyon, X. and Künsch, H. R. (1992). Asymptotic comparison of estimators in the Ising model, *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*, Lecture Notes in Statist., **74**, 177–198, Springer, Berlin.

Huang, F. and Ogata, Y. (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models, *J. Comput. Graph. Statist.*, **8**, 510–530.

Huang, F. and Ogata, Y. (2001). Comparison of two methods for calculating the partition functions of various spatial statistical models, *Australian and New Zealand Journal of Statistics*, **41**, 47–65.

Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73–101.

Huber, P. J. (1967). The behaviour of maximum likelihood estimates under nonstandard conditions, *Proc. 5th Berkeley Symp. on Math. Statist. Prob.*, **1**, 221–233.

Ishiguro, M. and Akaike, H. (1989). DALL: Davidon's algorithm for log likelihood maximization —A FORTRAN subroutine for statistical model builders, *Comput. Sci. Monographs*, **25**, The Institute of Statistical Mathematics, Tokyo.

Jensen, J. L. and Künsch, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes, *Ann. Inst. Statist. Math.*, **46**, 475–486.

Jensen, J. L. and Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes, *Ann. Appl. Probab.*, **1**, 445–461.

Mase, S. (1995). Consistency of the maximum pseudo-likelihood estimator of continuous state space Gibbsian processes, *Ann. Appl. Probab.*, **5**, 603–612.

Mercer, W. B. and Hall, A. D. (1911). The experimental error of field trials, *Journal of Agriculture Science*, **4**, 107–132.

Moyeed, R. A. and Baddeley, A. J. (1991). Stochastic approximation of the MLE for a spatial point pattern, *Scandinavian Journal of Statistics*, **18**, 39–50.

Ogata, Y. (1980). Maximum likelihood estimates of incorrect Markov models for time series and the derivation of AIC, *J. Appl. Probab.*, **17**, 59–72.

Onsager, L. (1944). Crystal statistics I, *Physical Review*, **65**, 117–149.

Penttinen, A. (1984). Modelling interaction in spatial point patterns: Parameter estimation by the maximum likelihood method, *Jyvaskyla Studies in Computer Science, Economics and Statistics*, **7**.

Pickard, D. K. (1987). Inference for discrete Markov fields: The simplest nontrivial case, *J. Amer. Statist. Assoc.*, **82**, 90–96.

Seymour, L. and Ji, C. (1996). Approximate Bayes model selection procedures for Gibbs-Markov random fields, *J. Statist. Plann. Inference*, **51**, 75–97.

Wiebe, G. A. (1935). Variation and correlation among 1500 wheat nursery plots, *Journal of Agriculture Research*, **50**, 331–357.

Younes, L. (1988). Estimation and annealing for Gibbsian fields, *Ann. Inst. H. Poincaré Probab. Statist.*, **24**, 269–294.

Younes, L. (1991). Maximum likelihood estimation for Gibbs fields, *Spatial Statistics and Imaging* (ed. A. Possolo), Lecture Notes-Monograph Series, Vol. 20, 403–426 Institute of Mathematical Statistics, Hayward, California.