Bayesian Estimation of Repulsive Interaction Potential Models for Spatial Point Patterns

OKABE Masahilo

DOCTOR OF PHILOSOPHY

Department of Statistical Science
School of Multidisciplinary Sciences
The Graduate University for Advanced Studies

2006

Bayesian Estimation of Repulsive Interaction Potential Models for Spatial Point Patterns

OKABE Masahilo

DOCTOR OF PHILOSOPHY

Department of Statistical Science
School of Multidisciplinary Sciences
The Graduate University for Advanced Studies

2006

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A *spatial point pattern* is a set of locations of points (objects), irregularly distributed within a designated region and presumed to have been generated by some form of stochastic mechanism (Diggle (2003)). Each point is considered as a particle, individual of animals or plants, and so on. During a few decades, the methods of statistical analysis for spatial point patterns have been developed: various diagnostic statistics and graphs have been studied by using the second-order and the nearest-neighbor distance methods (Ripley (1977, 1979a, 2004), Besag (1977), Diggle (1979, 2003)); by clumping indices based on the quadrat methods (David and Moore (1954), Morisita (1954), Lloyd (1967)) or based on distance methods (Hopkins and Skellam (1954)). Modelling spatial point patterns for which interactions exist between individuals has been studied by some authors (see for example Matérn (1960), Ripley (1977), Ogata and Tanemura (1981, 1984)).

Spatial point patterns are generally classified into three types: *completely random, clustered* (*aggregated*) and *regular*. If we observe a point pattern where a certain repulsive force is acting between individuals, the pattern is called a regular type. For example, if territorial animals or plants live in a habitat, a certain spacing out among them happens. If birds fly in formation or fish swim in shoals, inhibitions between the individuals are realized. Besides, in the microscopic world, we can often see regular point patterns. If few or many nanometer- or micrometer-sized dust particles are immersed in a plasma, the particles with charge form two- or three-dimensional dust Coulomb crystals. Then the behavior of charged dust particles and the structure of the crystals have been investigated (*e.g.* Melzer *et al.* (1994), Nitter (1996), Juan *et al.* (1998), Thomas and Watson (1999), Lai and Lin (1999)).

In this thesis, we are particularly interested in the interaction between individuals and it will be interesting to describe this certain spacing out by a repulsive interaction potential. Then we consider these interactions between individuals by parametric repulsive interaction potential models.

We assume that a given regular point pattern is in equilibrium under a certain repulsive interaction potential in a finite two-dimensional region. It is known that such an equilibrium point pattern is statistically represented by the Gibbs distribution. The likelihood of parameters which characterizes the interaction potential can be described by the Gibbs distribution for a given equilibrium point pattern. Since the form of the normalizing factor of the Gibbs distribution is a high multiplicity of integral, it is very difficult to obtain the likelihood function in principle. For this reason, Bayesian analysis for these spatial point patterns has been hardly studied (*e.g.* Heikkinen and Penttinen (1999)). Then, we use the useful approximate log-likelihood (Ogata and Tanemura (1989)), which will be described in §3.2, and consider our Bayesian estimation of various regular point patterns. Bayesian inference may help us to sensitively estimate parameters of the interaction potentials. The essential characteristic of Bayesian methods is their explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis (Gelman *et al.* (2004)). Because of the development of recent computational methodology, the complex posterior density can be simulated by using MCMC (Markov chain Monte Carlo) methods.

In the thesis, our main purpose is as follows. For a point pattern of repulsively interacting points in a finite two-dimensional region, we propose a method to obtain the posterior density of the parameters of the parameterized interaction potential functions by using MCMC methods. There, the effective approximate log-likelihood for the models (Ogata and Tanemura (1989)) plays an important role in the Metropolis-Hastings algorithm. Then two types of prior densities corresponding to the parameters of the repulsive interaction potential models are considered. Jumping (proposal) densities with similar type as prior density are applied in Markov chain simulations. Our Baysian procedure is confirmed by applying to various simulated equilibrium point patterns which are generated from MCMC of the Soft-Core models. In order to obtain posterior densities for real data sets, we perform our Bayesian procedure and consider posterior inference by fitting a certain parametric function to the posterior densities.

Moreover, MCMC convergence of iterative simulation is also investigated in detail. In the thesis, the approach of single long run is adopted. Before an iterative simulation is run in the Metropolis-Hastings algorithm, the following important problems arise: when should we begin and finish sampling?, *i.e.* when does the distribution of the chain begin to get close to the stationary density and when should we terminate the run? To solve these problems, we evaluate the burn-in and the stopping time of our single long run based on independent simulated multiple short runs with various starting points (Gelman and Rubin (1992), Geyer (1992), Cowles and Carlin (1996), Gelman *et al.* (2004)), which will be remarked in §5.3 and 8.2.

The layout of the thesis is as follows. In Chapter 2, a log-likelihood of parameters for equilibrium point pattern is given. In Chapter 3, the repulsive interaction potential models (Soft-Core potential models) with two parameters and their effective approximate log-likelihood are introduced. In Chapter 4, the fundamentals of Bayesian inference for the Soft-Core models are described. In Chapter 5, the Metropolis-Hastings algorithm for Bayesian inference, its jumping rule and the assessment of convergence (the burn-in and the stopping time) from iterative simulation are explained. In Chapter 6, firstly, our Baysian procedure is applied to various simulated equilibrium point patterns which are generated by MCMC methods of the Soft-Core models for the cases of large and relatively small number of points. Then MCMC convergence is evaluated and the comparison of marginal posterior densities of parameters under two types of the prior densities is also given. In Chapter 7, four real data sets are illustrated and their preliminary analysis is given in order to categorize the type of distribution of the point patterns. In Chapter 8, the results of our Bayesian estimation of the Soft-Core models for these real data sets are shown. There, the assessment of MCMC convergence is also investigated. In order to obtain posterior inference from iterative simulation, parametric fitting of the generalized gamma distribution to marginal posterior densities is considered. To examine the validity of our results, the L-statistics for observed data is compared graphically with the envelopes of simulated point patterns for the posterior mode of parameters. Finally, in Chapter 9, some concluding remarks are given.

Chapter 2

Likelihood for equilibrium point patterns

Let us consider a system consisting of N interacting particles (points, individuals or objects) in equilibrium in a finite two-dimensional region V, typically a rectangle. We call the system an equilibrium point pattern. A typical equilibrium point pattern is shown in Fig. 2.1.

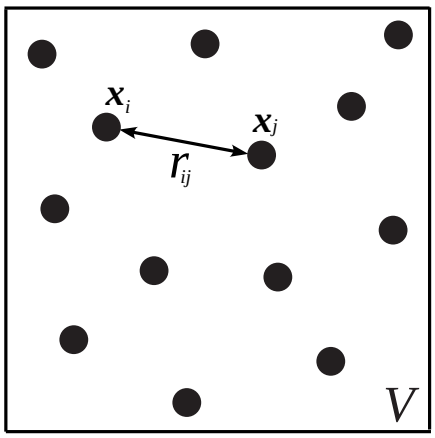

Figure 2.1: *Example of equilibrium point pattern.*

For describing the system, let the Cartesian coordinates of the observed points be $X = \{\boldsymbol{x}_i = (x_i, y_i) \in V, i = 1, 2, ..., N\}$, and we consider a family of parameterized interaction potential functions $\{\phi_\theta(r_{ij}); \theta \in \Theta\}$, defined in some parameter space Θ, where r_{ij} is the mutual distance $\mid \boldsymbol{x}_i - \boldsymbol{x}_j \mid$ between the particle centers $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ (see Fig. 2.1). The system of N particles $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N)$ which has reached the stationary state is characterized by the

Hamiltonian $H(X)$:

$$H(X) = \sum_{i<j}^{N} \phi_\theta(r_{ij}), \tag{2.1}$$

where $H(X)$ is equal to the total interaction potential energy. Then we can suppose the observed data X to follow the *Gibbs canonical distribution*:

$$f(X) = \exp\{-H(X)\}/Z(\phi_\theta; N, V), \tag{2.2}$$

where

$$Z(\phi_\theta; N, V) = \int_V \cdots \int_V \exp\{-H(X)\} d\boldsymbol{x}_1 \cdots d\boldsymbol{x}_N \tag{2.3}$$

is the *normalizing factor*, which is called the *partition function* in statistical mechanics (*e.g.* Ruelle (1977)). We have put $H(X)$ instead of $H(X)/k_BT$, where k_B is the *Boltzmann constant* and T is the *absolute temperature*. This means that the parameterized interaction potential functions $\phi_\theta(r_{ij})$ virtually includes the effect of the temperature.

The aim of the present thesis is to obtain, through Bayesian procedure, the posterior densities of the parameterized interaction potential functions $\phi_\theta(r_{ij})$ from the observed point pattern X. Then the Gibbs distribution (2.2) is used as the likelihood $L(\phi_\theta; X)$ of the potential $\phi_\theta(r_{ij})$ which is a function of parameter θ. Since the normalizing factor (2.3) has the high multiplicity of integral, it is very difficult to obtain the exact form of the likelihood (2.2) as a function of parameters except for the case of the *Poisson model*, *i.e.* the case of $\phi_\theta(r) \equiv 0$. For the Poisson model, we obtain the normalizing factor $Z = \mid V \mid^N$ from Eq. (2.3). Considering the Poisson model as the standard, we hereafter make use of the log-likelihood (ratio) function:

$$\ln L(\phi_\theta; X) = -\sum_{i<j}^{N} \phi_\theta(r_{ij}) - \ln \overline{Z}(\phi_\theta; N, V), \tag{2.4}$$

where $\overline{Z} = Z/ \mid V \mid^N$.

Chapter 3

Repulsive interaction potential models

3.1 The two-parameter Soft-Core models

We can often observe the regular point patterns in nature. In such a point pattern, a certain spacing out among individuals happens. This spacing out might be due to the competitions among individuals for territories, foods and so on. It will be useful to describe these competitions by a repulsive interaction potential. We here have the assumptions of homogeneity and isotropy for data presented as N individuals in a finite two-dimensional region V. In order to represent the range and the softness of the interactions, the so-called *Soft-Core potential models*:

$$\phi_\theta(r) = \left(\frac{\sigma}{r}\right)^n;\ n = 2/\alpha,\ 0 \leq \sigma < \infty,\ 0 \leq \alpha < 1, \tag{3.1}$$

is the most convenient. Here, θ has two components (α, σ). The parameter α represents the softness of the potential. The shape of potential function depends on the parameter α. As the special case, $\alpha \to +0$ corresponds to the *Hard-Core potential model* (or, *rigid sphere model*) with diameter σ of the particle:

$$\phi_{(0,\sigma)}(r) = \begin{cases} \infty & (r \leq \sigma), \\ 0 & (r > \sigma). \end{cases} \tag{3.2}$$

The Hard-Core potential model is known as typical of the repulsive interaction potential models. On the other hand, the parameter σ represents the range of potential and we call σ the *scale parameter*. In particular, $\sigma = 0$ corresponds to the Poisson model. Comparison of Hard-Core and Soft-Core potential curves for $\sigma = 1$ is shown in Fig. 3.1.

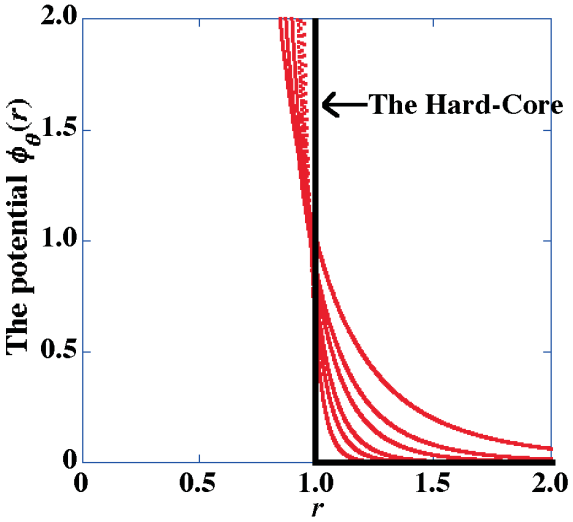

Figure 3.1: *Curves of the potential models for* $\sigma = 1$. *Each curve near the Hard-Core model corresponds to the Soft-Core models with* $n = 4$, *6, 8, 12, 16 and 24 from outside.*

3.2 Approximate likelihood for the two-parameter Soft-Core models

In statistical physics, all thermodynamic quantities are derived from the value of normalizing factor (2.3) in principle. Computation of the normalizing factor is necessary to maximize the likelihood of parameters. However, in general, it is hard to obtain the explicit form of the normalizing factor as a function of θ because of a high multiplicity of the integral in Eq. (2.3). To avoid such a difficulty, Ogata and Tanemura (1981) have developed the approximate normalizing factor under the assumption that a spatial point pattern is sparse. Furthermore, Ogata and Tanemura (1984, 1989) have devised the following Monte Carlo methods for the estimation of the normalizing factor for the case of point patterns with higher density. We here define the *number density* by

$$\rho = N/V, \tag{3.3}$$

assuming that the number of particles N and the volume of the region V are not too small. Since the crampedness of a point pattern is related to the scale parameter σ, we here introduce a parameter τ:

$$\tau = \rho\sigma^2, \tag{3.4}$$

which is called the *reduced density*. In thermodynamics, it is known that the *compressibility factor* $\psi(\tau) = P(\tau)V/N - 1$ ($P(\tau)V/N = \psi(\tau) + 1$ is the *equation*

of state, where $P(\tau)$ is the *pressure*) measures the deviation from the ideal gas. It is also known that $\psi(\tau)$ is represented as a function of τ for the potential models with the scaling property described in §3.1, and that the normalizing factor is given by $\psi(\tau)$ in the form (Ogata and Tanemura (1984, 1989)):

$$\frac{1}{N}\ln \overline{Z}(\tau) = -\int_0^\tau \frac{\psi(t)}{t}dt. \tag{3.5}$$

To obtain the useful approximated log-likelihood, Ogata and Tanemura (1984, 1989) performed the Monte Carlo experiments to estimate the compressibility factor $\psi(\tau)$ by generating a great number of equilibrium point patterns for various values of τ and n. For the Soft-Core models, the function $\psi(\tau)$ is estimated as follows:

$$\hat{\psi}(\tau) = \frac{n}{2N}\frac{1}{M}\sum_{t=1}^{M}\sum_{i<j}^{N}\left(\frac{\sqrt{\tau}}{r_{ij}^*(t)}\right)^n. \tag{3.6}$$

Here $r_{ij}^*(t) = \mid \boldsymbol{x}_i(t) - \boldsymbol{x}_j(t) \mid /\sqrt{V/N}$ and the patterns $[\ X(t) = \{\boldsymbol{x}_i(t); i = 1, 2, ..., N\}; t = 1, 2, ..., M\ ]$ are obtained by the Monte Carlo simulations with M steps. After some computer simulations, Ogata and Tanemura (1984, 1989) obtained the compressibility factor (sample means) $\hat{\psi}(\tau_k)$ for $\tau_k = 0.05k$, $k = 1, 2, ..., 15$ and for $n = 4, 6, 8, 12, 16$ and 24 together with the standard deviations $\hat{s}_{k,n}$ of the time series. We regard the log-normalizing factor $\ln \overline{Z}$ of (2.4) as a function of (α, τ), and consider $\psi(\tau)/\tau$ instead of $\psi(\tau)$. To estimate the function $\psi(\tau)/\tau$ in two parameter space (α, τ), Ogata and Tanemura (1989) fitted the *2-dimensional cubic spline function* to the Monte Carlo simulated data for $\psi(\tau)/\tau$. In a subdivided rectangle $[\alpha_m, \alpha_{m+1}] \times [\tau_n, \tau_{n+1}]$ for (α, τ)-plane, the 2-dimensional cubic spline function $h(\alpha, \tau \mid C)$ is defined as follows:

$$h(\alpha, \tau \mid C) = \sum_{i=0}^{3}\sum_{j=0}^{3} c_{m+i,n+j} B_{4-i}(r_\alpha) B_{4-j}(r_\tau), \tag{3.7}$$

$$\begin{cases} B_1(r) = r^3/6, \\ B_2(r) = (-3r^3 + 3r^2 + 3r + 1)/6, \\ B_3(r) = (3r^3 - 6r^2 + 4)/6, \\ B_4(r) = (-r^3 + 3r^2 - 3r + 1)/6, \end{cases} \tag{3.8}$$

where $\{B_i(r); i = 1, 2, 3, 4\}$ is the *B-spline basis* (Ahlberg *et al.* (1967)) on $[0, 1]$, C represents the coefficients $\{c_{m+i,n+j}\}$, and m and n are respectively specified by the range of (α, τ), that is, $r_\alpha = 3(\alpha - m/3)$, $(m = -3, -2, ..., 6)$ and $r_\tau = 4(\tau - n/4)$, $(n = -3, -2, ..., 7)$. Then Ogata and Tanemura (1989) considered a linear regression model:

$$\frac{\psi(\alpha_i, \tau_j)}{\tau_j} = h(\alpha_i, \tau_j \mid C) + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \varsigma^2 s_{ij}^2), \tag{3.9}$$

where ς^2 is a parameter to be minimized. The coefficient matrix $C = \{c_{ij}\}$ is obtained as to minimize the penalized log-likelihood $Q(C, w)$ for total number $IJ = 7 \times 16$ of simulated data set (α_i, τ_j):

$$Q(C, w) = \ln q(C) - [w_1 \Psi_1(h) + w_2 \Psi_2(h)], \tag{3.10}$$

$$\Psi_1(h) = \iint_S \left(\left(\frac{\partial h}{\partial x} \right)^2 + \left(\frac{\partial h}{\partial y} \right)^2 \right) dxdy, \tag{3.11}$$

$$\Psi_2(h) = \iint_S \left(\left(\frac{\partial^2 h}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 h}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 h}{\partial y^2} \right)^2 \right) dxdy, \tag{3.12}$$

$$\ln q(C) = -(IJ/2) \ln \varsigma^2 + \sum_{i=1}^{I} \sum_{j=1}^{J} [\psi(\alpha_i, \tau_j)/\tau_j - h(\alpha_i, \tau_j \mid C)]^2 / (2\varsigma^2 s_{ij}^2), \tag{3.13}$$

where the values of (w_1, w_2) are the weights which control the roughness penalties (smoothness functionals) (Ψ_1, Ψ_2) and the combination of first- and second-order derivatives are used to estimate the roughness of the function. Thus we can use the log-likelihood function for the Soft-Core models:

$$\ln L(\alpha, \sigma; X) = -\sum_{i<j}^{N} \left(\frac{\sigma}{r_{ij}} \right)^{2/\alpha} + N \int_0^{\tau} h(\alpha, \tau \mid \hat{C}) dt, \tag{3.14}$$

where $n = 2/\alpha$ and the log-likelihood function is effective within the region $0 < \tau \leq 0.75$ and $0 < \alpha \leq 0.5$ (*i.e.* $4 \leq n < \infty$). Since $h(\alpha, \tau \mid \hat{C})$ in Eq. (3.7) corresponds to $\psi(t)/t$ in Eq. (3.5), the integral form of $h(\alpha, \tau \mid \hat{C})$ appears in the second term of the rhs in Eq. (3.14). The estimated coefficient matrix $\hat{C}$ is given in Table 2 of Ogata and Tanemura (1989) under the values of $(w_1, w_2) = (4.42 \times 10^{-6}, 5.53 \times 10^{-3})$ (Ogata and Katsura (1988)). We apply the approximate log-likelihood (3.14) to our Bayesian procedure.

Chapter 4

Bayesian inference

4.1 Bayesian paradigm

Bayesian inference for the statistical model is as follows (*e.g.* Box and Tiao (1973), Gelman *et al.* (2004), O'Hagan and Forster (2004), Robert (2004)). Prior information about the interested parameter θ is quantified as a density $p(\theta)$. The likelihood function for θ can be represented as $L(\theta; X)$ given the observed data (point pattern) X. From the Bayes theorem, a product of the prior density $p(\theta)$ and the likelihood function $L(\theta; X)$ results the posterior density of θ:

$$p(\theta \mid X) \propto p(\theta)L(\theta; X), \tag{4.1}$$

where the rhs is the *unnormalized posterior density*. Any inference about θ is then based on the posterior density $p(\theta \mid X)$ after observing the data X. From Eq. (4.1), the full generality of the posterior density of θ is equivalent to

$$p(\theta \mid X) = \frac{p(\theta)L(\theta; X)}{\int_\Theta p(\theta)L(\theta; X)d\theta}, \tag{4.2}$$

where the denominator of the rhs is called the *normalizing constant*.

Suppose that the prior density is chosen from a parametric family $\mathcal{F}$. If for every prior $p(\theta) \in \mathcal{F}$ the posterior $p(\theta \mid X)$ also belongs to $\mathcal{F}$, then the family $\mathcal{F}$ of prior is said to be *closed under sampling* for the sampling distribution (likelihood). This property is called *conjugacy*. The conjugate prior is often used in some standard models such as binomial, Poisson and normal, *etc* for the likelihood. The use of conjugate prior have the practical advantage for mathematical convenience. Then it is easy to understand the results which can be often put in analytic form and are often good approximation. In practice, conjugate prior densities may not be possible for complicated models including high dimensions, then it is difficult to obtain posterior densities. Even if, in early days, the conjugate priors could be applied, computational power was not

available. However, since the advancement of recent computational methodology, through MCMC methods, we can simulate the complex posterior density without knowing the normalizing constant. The idea of MCMC methods will be explained in Chapter 5. Note that the normalizing constant is different from the normalizing factor Z (2.3) of the likelihood for θ in Eq. (2.2) (*i.e.* $\overline{Z}$ of the log-likelihood in Eq. (2.4)). As we shall see in Chapter 5, the value of the normalizing factor, which is the second term in the rhs of Eq. (3.14), is needed in the *Metropolis-Hastings algorithm*.

Our purpose is to obtain Bayesian inference about the Soft-Core potential models (3.1), that is, we estimate the posterior density $p(\theta \mid X)$ from the observed point pattern X. Then we run Markov chain simulation for a long time in the Metropolis-Hastings algorithm, regard it as samples of θ from the posterior density $p(\theta \mid X)$ and use the samples to estimate the properties of posterior density.

We will describe our Bayesian procedure about the Soft-Core potential models with two parameters $\theta = (\alpha, \sigma)$ in §4.2, 4.3 and Chapter 5.

4.2 Prior densities

We specify a prior density $p(\theta)$ as

$$p(\theta) = p(\alpha)p(\sigma), \tag{4.3}$$

where $p(\alpha)$ and $p(\sigma)$ are respectively the prior densities of α and σ.

In this thesis, Bayesian analyses are carried out under following two prior specifications: noninformative and informative prior densities. For noninformative priors, we select the *uniform* prior densities which are flat distributions. The justification of using noninformative priors is often said to be 'to let the data speak for themselves', so that inferences are unaffected by information external to the current data (Gelman *et al.* (2004)). Noninformative priors can be often taken as reference or default ones. For informative priors, we consider the following distributions. The *truncated normal* prior densities are chosen. The truncated normal densities are not flat and not symmetric. When the number of points N is large or relatively small, we ascertain the posterior to be sensitive to the choice of the priors. Since Bayesian inference is based on the posterior density for a particular choice of priors, the uniform priors may be useful for comparison with the asymmetric priors. Two types of prior densities for the parameters of the Soft-Core models are as follows:

1. Uniform prior densities (the type (i) prior densities):

$$p(\alpha) = \mathrm{U}(\alpha \mid a, b), \tag{4.4}$$

$$p(\sigma) = \mathrm{U}(\sigma \mid c, d), \tag{4.5}$$

where $\mathrm{U}(\cdot \mid \xi, \eta)$ is a uniform distribution on the interval $[\xi, \eta]$; a, b, c and d are the hyperparameters.

2. Truncated normal prior densities (the type (ii) prior densities):

$$p(\alpha) = \mathrm{TN}_{[0,\infty)}(\alpha \mid \mu_\alpha, s_\alpha^2), \tag{4.6}$$

$$p(\sigma) = \mathrm{TN}_{[0,\infty)}(\sigma \mid \mu_\sigma, s_\sigma^2), \tag{4.7}$$

where $\mathrm{TN}_{[\xi,\eta)}(\cdot \mid \mu, s^2)$ is a truncated normal distribution which is a normal distribution with mean μ and variance s^2 defined on the truncated interval $[\xi, \eta)$. We take here the truncated interval $[0, \infty)$ because the parameters α and σ take only non-negative values. Hereafter, we will abbreviate "truncated normal" to "normal" for simplicity unless otherwise stated.

In two cases, parameters α and σ are initially assumed to be respectively independent.

4.3 Posterior densities

For each of the above two types of prior specifications, the joint posterior density of α and σ, given the data X is obtained as

$$p(\alpha, \sigma \mid X) \propto p(\alpha)p(\sigma)L(\alpha, \sigma; X), \tag{4.8}$$

where $L(\alpha, \sigma; X)$ is the likelihood of parameters (α, σ) for spatial point pattern X (the approximate log-likelihood (3.14) is used). And the marginal posterior densities of α and σ are respectively expressed by

$$p(\alpha \mid X) \propto \int p(\alpha)p(\sigma)L(\alpha, \sigma; X)d\sigma, \tag{4.9}$$

$$p(\sigma \mid X) \propto \int p(\alpha)p(\sigma)L(\alpha, \sigma; X)d\alpha. \tag{4.10}$$

So we can estimate the two marginal posterior densities $p(\alpha \mid X)$ and $p(\sigma \mid X)$.

Chapter 5

Markov chain Monte Carlo methods

5.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm (Hastings (1970)) generates a general family of MCMC simulation methods. This method is widely used in different fields of science and is usually simple for computer programming. We note that a Metropolis method (Metropolis *et al.* (1953), Gelman *et al.* (2004)) is known as a particular type of Metropolis-Hastings algorithm. For the time $t = 0, 1, 2, \cdots$, the Metropolis-Hastings algorithm produces a sequence of random points $(\theta^0, \theta^1, \theta^2, \cdots)$ whose distributions converge to the desired target posterior density $p(\theta \mid X)$, given the observed point pattern X. As we shall see later, a candidate point θ^* is generated from an asymmetric transition function, given the current point θ^{t-1}. The Metropolis-Hastings algorithm proceeds as follows:

1. Draw a starting point θ^0 from a *starting density* (prior density) $p(\theta)$:

$$\theta^0 \sim p(\theta). \tag{5.1}$$

2. For the *iteration time* $t = 1, 2, \cdots, T$ (T is the *stopping time*):

   (a) Sample a *proposal* (known as a new candidate) point θ^* from a *jumping (proposal or transition) density* $J_t(\theta^* \mid \theta^{t-1})$ at time t, given the current point θ^{t-1}:

$$\theta^* \sim J_t(\theta^* \mid \theta^{t-1}). \tag{5.2}$$

   In Markov chain simulation, a candidate point θ^* does not depend on $(\theta^0, \theta^1, \ldots, \theta^{t-2})$ but on the current point θ^{t-1}.

(b) Compute the ratio w:

$$w = \frac{p(\theta^* \mid X)/J_t(\theta^* \mid \theta^{t-1})}{p(\theta^{t-1} \mid X)/J_t(\theta^{t-1} \mid \theta^*)}. \tag{5.3}$$

The ratio w is always defined, since a jump from θ^{t-1} to θ^* can only occur if both $p(\theta^{t-1} \mid X)$ and $J_t(\theta^* \mid \theta^{t-1})$ are not zero. Instead of calculating w itself, we compute its logarithm $\ln(w)$: if we make use of the log-likelihood function (3.14) and Eq. (4.1), it is easier to compute $\ln(w) = \ln[p(\theta^* \mid X)/J_t(\theta^* \mid \theta^{t-1})] - \ln[p(\theta^{t-1} \mid X)/J_t(\theta^{t-1} \mid \theta^*)]$ than to compute w, as described in detail below. Note that if the jumping density is symmetric (for the case of the Metropolis algorithm), namely $J_t(\theta_a \mid \theta_b) = J_t(\theta_b \mid \theta_a)$ for all θ_a, θ_b and t, then the ratio w turns out $p(\theta^* \mid X)/p(\theta^{t-1} \mid X)$.

(c) If $\ln(w) > 0$, then set $\theta^t = \theta^*$, else if $\ln(w) < 0$, then generate an independent uniform random number u from the uniform distribution on $[0, 1]$ and set

$$\theta^t = \begin{cases} \theta^* & (\ln(w) \geq \ln(u)), \\ \theta^{t-1} & (\ln(w) < \ln(u)). \end{cases}$$

Here, we describe the advantage of the use of the Metropolis-Hastings algorithm. It is known that the Metropolis algorithm is proposed as the Monte Carlo method of simulating the Gibbsian equilibrium point patterns without calculating the normalizing factor (2.3). On the other hand, in Bayesian inference, the Metropolis-Hastings algorithm is used for simulating the target posterior density $p(\theta \mid X)$ without calculating the normalizing constant of $p(\theta \mid X)$, not of the normalizing factor (2.3) of $L(\theta; X)$ in Eq. (5.1). In the Metropolis-Hastings algorithm for our Bayesian procedure, we need to calculate the ratio $w = p(\theta^*)L(\theta^*; X)J_t(\theta^{t-1} \mid \theta^*)/p(\theta^{t-1})L(\theta^{t-1}; X)J_t(\theta^* \mid \theta^{t-1})$ in Eq. (5.3). In this ratio, the normalizing factor (2.3) of $L(\theta; X)$ is not cancelled. It means that the values of $L(\theta^*; X)$ and $L(\theta^{t-1}; X)$ each are necessary to compute w. Therefore, in the thesis, it is essential to use the explicit form of the approximated log-likelihood.

5.2 Jumping rule

5.2.1 Jumping densities

For Markov chain simulations, to draw a candidate point θ^* given the current point θ^{t-1} of the chain, we assume the following independent jumping density $J_t(\theta^* \mid \theta^{t-1})$:

$$J_t(\theta^* \mid \theta^{t-1}) = J_t(\alpha^* \mid \alpha^{t-1})J_t(\sigma^* \mid \sigma^{t-1}). \tag{5.4}$$

For the type (i) priors, we adopt the two jumping densities $J_t(\alpha^* \mid \alpha^{t-1})$ and $J_t(\sigma^* \mid \sigma^{t-1})$ as uniform, centered on the current point and with the interval length of $2\delta_\alpha$ and $2\delta_\sigma$, respectively:

$$J_t(\alpha^* \mid \alpha^{t-1}) = \begin{cases} \mathrm{U}(\alpha^* \mid \alpha^{t-1} - \delta_\alpha, \alpha^{t-1} + \delta_\alpha) & (a + \delta_\alpha \leq \theta^{t-1} \leq b - \delta_\alpha), \\ \mathrm{U}(\alpha^* \mid a, \alpha^{t-1} + \delta_\alpha) & (\theta^{t-1} < a + \delta_\alpha);\ \delta_\alpha > 0, \end{cases} \tag{5.5}$$

$$J_t(\sigma^* \mid \sigma^{t-1}) = \begin{cases} \mathrm{U}(\sigma^* \mid \sigma^{t-1} - \delta_\sigma, \sigma^{t-1} + \delta_\sigma) & (c + \delta_\sigma \leq \theta^{t-1} \leq d - \delta_\sigma), \\ \mathrm{U}(\sigma^* \mid c, \sigma^{t-1} + \delta_\sigma) & (\theta^{t-1} < c + \delta_\sigma);\ \delta_\sigma > 0. \end{cases} \tag{5.6}$$

The values of δ should be adjusted for the optimal proportion of acceptable jumps. So, δ_α and δ_σ are called the *adjusting parameters* for the jumping of α and σ, respectively.

For the type (ii) priors, we specify two jumping densities as truncated normal (we call it "normal" as noted before), whose mean and variance are the current point and δ^2, respectively:

$$J_t(\alpha^* \mid \alpha^{t-1}) = \mathrm{TN}_{[0,\infty)}(\alpha^* \mid \alpha^{t-1}, \delta_\alpha^2);\ \delta_\alpha > 0, \tag{5.7}$$

$$J_t(\sigma^* \mid \sigma^{t-1}) = \mathrm{TN}_{[0,\infty)}(\sigma^* \mid \sigma^{t-1}, \delta_\sigma^2);\ \delta_\sigma > 0, \tag{5.8}$$

where δ_α^2 and δ_σ^2 are the adjusting parameters for the jumping of α and σ, respectively. For the normal jumping densities (5.9)-(5.10), given the current point θ^{t-1}, draw a candidate point θ^* from a normal density $\mathrm{N}(\theta^{t-1}, \delta_\theta^2)$ recurrently until θ^* is sampled in the range $[0, \infty)$ (*e.g.* Geweke (1991)).

For comparison, the *independence sampler* will be also examined. We choose the two uniform jumping densities of the parameters as independence samplers under the type (i) priors:

$$J_t(\alpha^* \mid \alpha^{t-1}) = \mathrm{U}(\alpha^* \mid a, b), \tag{5.9}$$

$$J_t(\sigma^* \mid \sigma^{t-1}) = \mathrm{U}(\sigma^* \mid c, d), \tag{5.10}$$

where the form of the densities is the same as the type (i) prior, respectively (*e.g.* Møller (2003)). Hereafter, we will abbreviate "independent uniform sampler" to "independence sampler" for simplicity unless otherwise remarked.

5.2.2 Relation between jumping rule and efficiency of Markov chain simulations

We describe relation between jumping rule and efficiency of Markov chain simulations. The ideal jumping density of the Metropolis-Hastings algorithm is the target density, that is, $J_t(\theta^* \mid \theta^{t-1}) = p(\theta^* \mid X)$ for all θ. Then the ratio

w in Eq. (5.3) is always 1 strictly, and the all iterations θ^t are a sequence of independent draws from $p(\theta \mid X)$. However, in practice, iterative simulation is applied to problems for which direct sampling is impossible. A good jumping density has following properties: it is desirable that it is easy to sample θ^* from $J_t(\theta^* \mid \theta^{t-1})$ for any θ; it is easy to compute the ratio w; each jump goes a reasonable distance in the parameter space Θ; the jump are not rejected too frequently.

In this thesis, we must make a careful choice of the values of the two parameters δ_α and δ_σ so that each transition keeps moving in two-dimensional parameter space Θ, and the transitions are not accepted or rejected too frequently in Markov chain simulations; we carried out simulation experiments to have the rate of acceptance of candidate point θ^* around $0.40 \sim 0.44$ in two dimensions within Metropolis-Hastings steps in the application to the simulated data (see Chapter 6) and several real data (see Chapter 8) (Gelman *et al.* (2004)).

5.3 Assessing convergence from iterative simulation

5.3.1 Assessing convergence and stopping time

In §5.1 and 5.2, we have presented the Metropolis-Hastings algorithm. An important problem concerning the implementation of MCMC methods is to assess when *convergence* has been achieved. Convergence is considered as follows: the distribution of the Markov chain at time $t \geq T^*$ (T^* is the initial *burn-in time*, which will be described in §5.3.2 and 8.2.3) is sufficiently close to the target density, then Markov chain samples adequately describe the characteristics of the target density. Then, to assess the convergence from iterative simulation, *convergence diagnostics* are considered. Convergence diagnostics are the quantities that assess how long to run a chain to obtain observations approximately from the stationary density (*e.g.* Gelfand and Smith (1990), Gilks *et al.* (1996), Gamerman (1997), Brooks (1998), Roberts and Tweedie (2001), Møller (2003), Robert and Casella (2004)). Various methods for assessing convergence have been proposed: for single long run, diagnostics based on the spectral analysis (Geweke (1992)), based on the convergence rate estimate (Raftery and Lewis (1992)) and based on the graphical cusum path plots (Yu and Mykland (1998)), *etc*; for multiple short runs, diagnostics based on the large-sample normal theory (Gelman and Rubin (1992)) and based on the probability theory (Roberts (1992)), *etc.* For extensive reviews, see Cowles and Carlin (1996), Brooks and Roberts (1998) and Mengersen *et al.* (1999).

There is a controversy about MCMC convergence: single long run (Geyer

(1992)) versus multiple short runs (Gelman and Rubin (1992)). An approach based on the single long run is as follows (Geyer (1992)). One very long run is a valuable diagnostic. If the run does not seem to be stationary, it is too short. As the iteration time t is getting larger, a distribution of a sequence of simulated random points θ^t may be closer to the target density. Running one long time iterative simulation is of higher efficiency and smaller bias than running multiple short runs and will have the chance of getting posterior modes. However, since it is possible that the simulated points θ^t keep on moving in the small part of the parameter space, single long run may fail to detect metastability.

On the other hand, an approach based on the multiple short runs proceeds as follows (Gelman and Rubin (1992)). Multiple short runs are the parallel simulation of several $k(\geq 2)$ independent sequences $\theta^t_{(k)}$ with various starting points $\theta^0_{(k)}$. By using the simulated multiple sequences, it is easy and effective to control convergence to the stationary density comparing the estimation of quantities of interest without knowing the time series structure of the simulations. Each of these sequences may indicate slow mixing and sticky regions. Then we can regard these various independent sequences as samples from the posterior density. For the implementation of this approach, note that the slower sequence governs convergence and that the choice of the starting points $\theta^0_{(k)}$ is very important in guaranteeing that the various sequences are well dispersed. It is waste of discarding initial burn-in time from each of multiple short runs.

Each approach has its advantages and disadvantages. Asmussen *et al.* (1992) pointed out that no convergence diagnostic technique will successfully diagnose convergence in all settings. In particular, for slow mixing sequences, convergence diagnostics may be often unreliable. We should carry out the exploration of the target density beforehand. Whenever the approach based on multiple short run is adopted, we should run very long time iterative simulation once at least.

In our case, the log-likelihood for the Soft-Core models (3.14) is effective within $0 < \alpha \leq 0.5$ and $0 < \tau \leq 0.75$, as described in §3.2. Within the region, the system consisting of N interacting points does not make a phase transition. When we run a long time iterative simulation under this assumption, the simulated posterior density will not be multimodal. Then, in the thesis, we adopt the approach of single long run. Multiple short runs with various starting points are also simulated. We employ these multiple short runs only for estimating the convergence and the stopping time T of the simulated single long run.

Our approach of the assessment of convergence proceeds as follows. After we have run the long time iterative simulation in the above Metropolis-Hastings algorithm, we investigated when the convergence had been reached and when we have to terminate the simulation. To estimate the stopping time T of our

single long run, we computed the diagnostics quantity (*potential scale reduction factor*) $\hat{R}$, defined below, in the approach of Gelman and Rubin (1992) (see Gelman *et al.* (2004), Cowles and Carlin (1996)). Gelman and Rubin's $\hat{R}$ is one of the quantities in the MCMC convergence diagnostics. In particular, since it is easy for computer programming, it is useful for users as a convergence diagnostics. Gelman and Rubin proposed that the prescribed several sequences for each scalar estimand of interest are run independently with starting points drawn from an overdispersed distribution (either from rough estimates or from a more elaborate approximate distribution). Then a comparison of calculating the *between- and within-sequence variances* (*i.e.* the potential scale reduction factor $\hat{R}$) will reveal whether the sequences converge or not. Gelman and Rubin's method can be stated as follows.

1. Simulate independently the prescribed $k(\geq 2)$ parallel sequences whose number of iteration is (l^*+l) each, with starting points dispersed throughout the parameter space. We throw away the first l^* iterations as a burn-in time. We label the simulation draws as φ_{ij} $(i = 1, ..., l;\ j = 1, ..., k)$ for the scalar estimand φ of interest.

2. Compute the between- and within-sequence variances, B and W, for simulated k parallel sequences:

$$B = \frac{l}{k-1}\sum_{j=1}^{k}(\bar{\varphi}_{\cdot j} - \bar{\varphi}_{\cdot\cdot})^2, \quad \text{where} \quad \bar{\varphi}_{\cdot j} = \frac{1}{l}\sum_{i=1}^{l}\varphi_{ij},\ \bar{\varphi}_{\cdot\cdot} = \frac{1}{k}\sum_{j=1}^{k}\bar{\varphi}_{\cdot j}, \tag{5.11}$$

$$W = \frac{1}{k}\sum_{j=1}^{k}s_j^2, \quad \text{where} \quad s_j^2 = \frac{1}{l-1}\sum_{i=1}^{l}(\varphi_{ij} - \bar{\varphi}_{\cdot j})^2. \tag{5.12}$$

   Note that B/l is the variance between the means from k parallel sequences, and W is the mean of the k within-sequence variances.

3. Calculate the potential scale reduction factor $\hat{R}$:

$$\hat{R} = \sqrt{\frac{l-1}{l} + \frac{1}{l}\cdot\frac{B}{W}} \tag{5.13}$$

   to estimate the factor by which the variance of the estimand of interest might be reduced by continuing simulation. If the value of $\hat{R}$ is near 1 for all scalar estimands of interest, then we can assume that enough simulations have been run and just collect $(k \times l)$ simulations from the second parts of all the sequences. And we can regard the $(k \times l)$ simulations as samples from the posterior distribution. The value of $\hat{R}$ below 1.1 are

acceptable in practice. A cut-off of 1.1 was suggested by Gelman in a round table discussion on MCMC methods (Kass *et al.* (1997)).

As an estimate of the stopping time T, we take $T = T^* + (k \times l)$ for our single long run approach. Here $T^*(\equiv l^*)$ is the initial burn-in time, which will be described in the next section. Then these $(k \times l)$ simulations are regarded as samples from the posterior distribution.

5.3.2 Assessing burn-in time

In Markov chain simulation, the practice of discarding initial iterations is mentioned to as burn-in and it is necessary to use burn-in time T^* in order to diminish the effect of the starting point θ^0. Then, we have the following question; for large enough iteration time, when should we begin sampling?, that is, how long does it take the Markov chain to get sufficiently close to the stationary (target) density $p(\theta \mid X)$? Some methods of effective burn-in have been discussed. Geyer (1992) proposed routinely throwing away the initial 1 or 2% of runs will suffice. Jones and Hobert (2001) suggested appropriate burn-in is the total variation distance between the density of the simulated value and the stationary density is less than 1%. Gelman *et al* (2004) pointed out discarding the first half of sequence and focus attention on the second half. In general, discarding the initial $100 \sim 1000$ steps of the simulation runs is often used for the appropriate burn-in time, then dependence on the starting point is supposed to be lost.

In this thesis, the burn-in time T^* of a single long run is investigated and evaluated by simulating multiple short runs and applying the above Gelman and Rubin's method as remarked in §5.2.2.

Chapter 6

Bayesian inference for simulated equilibrium point patterns

In order to verify our Bayesian procedure, we performed Bayesian inference for simulated equilibrium point patterns. Then, to investigate the influence of priors upon posterior, we applied to the cases of the number of points $N = 500$ and $N = 100$.

6.1 Large sample point patterns

6.1.1 Example: the single-parameter Soft-Core models

Firstly, our Bayesian procedure was applied to the simulated data for the case of $N = 500$. These simulated data were generated from MCMC of the Soft-Core models, through the Metropolis algorithm, for the cases of $N = 500$, $V = \sqrt{500} \times \sqrt{500}$, $\alpha = 1/3$, $\tau = 0.1, 0.3, 0.5$. These simulated equilibrium point patterns are relatively large size and illustrated in Figs. 6.1(a)-(c). These three data sets (a), (b) and (c) are named D1,D2 and D3, respectively. From Figs. 6.1(a)-(c), we can see that as the reduced density τ is getting larger, the degree of regularity increases.

By employing the simulated data, we can obtain the posterior densities of the interaction potential functions and then verify our Bayesian procedure. In this subsection, to simplify the two-parameter Soft-Core models of Eq. (3.1), we begin with considering the case of the single-parameter ones for the three simulated data sets (D1,D2,D3). The method is as follows. Each parameter α or τ will be estimated separately by fixing the other parameter. Then, in Eq. (3.1), we can define as the Soft-Core models with the parameter α fixing τ and with the parameter τ fixing α, respectively. Each of them is called the *single-parameter Soft-Core models*. Then, we adopted the uniform prior (type (i) prior) density with the uniform jumping density and with the independence

sampler. Then the log-likelihood Eq. (3.14) was used. By considering the effective region of the log-likelihood, we set the hyperparameters of the type (i) prior density. To specify the jumping density for the type (i) prior, the adjusting parameters were chosen such that Markov chain simulation should have the acceptance rate about 0.5. Then, for estimating α by fixing τ (*i.e.* σ), we set the hyperparameters (a, b) in Eq. (4.4) and the optimal adjusting parameter δ_α of Eq. (5.5). These values are given in Table 6.1. For estimating τ by fixing α, we put the hyperparameters of (c, d) in Eq. (4.5) (using Eq. (3.4)) and the optimal adjusting parameter δ_σ of Eq. (5.6). The values are given in Table 6.2. When we used the independence sampler as the jumping density, the type (i) prior was chosen, as described in §5.2.1 (see Tables 6.1-6.2). Then we performed a single long run simulation of 31000 steps in the Metropolis-Hastings algorithm for the data sets (D1,D2,D3). We calculated the total potential energy in Eq. (3.14) by using the *periodic boundary condition*, *i.e.* the region (rectangle) V was regarded as a torus (see §8.5).

MCMC methods rely on the generation of random numbers. In the implementation of the methods, it is important to choose a good and fast random number generator. If we use a deterministic *pseudo-random number generator*, a periodicity usually appears. Then MCMC random samples will suffer some biases. Contrary to this, a *physical random number generator* is known to be free from periodicity. Therefore, we employed the reliable and fast physical random number generator (133Mbytes per second) equipped in the Institute of Statistical Mathematics.

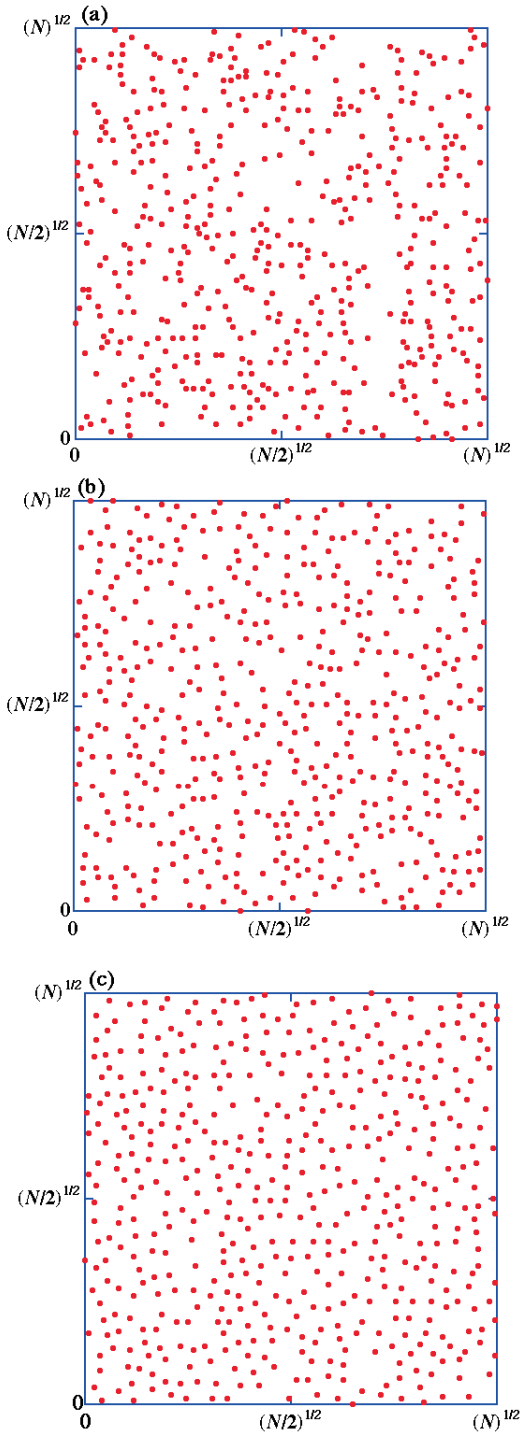

Figure 6.1: *Simulated equilibrium point patterns* ($N = 500, V = \sqrt{500} \times \sqrt{500}, \alpha = 1/3$). *(a):* $\tau = 0.1$ *(D1); (b):* $\tau = 0.3$ *(D2); (c):* $\tau = 0.5$ *(D3).*

Table 6.1: *Prior and jumping specification (estimating α by fixing τ (i.e. σ)): the values of hyperparameters (a, b) in Eq. (4.4), the adjusting parameters δ_α in Eq. (5.5) and the independence sampler in Eq. (5.9) for the type (i) prior.*

Data	(a, b)	δ_α
D1	(0, 0.50)	0.15
D2	(0, 0.50)	0.08
D3	(0, 0.50)	0.05

Table 6.2: *Prior and jumping specification (estimating τ(i.e. σ) by fixing α): the values of hyperparameters (c, d) in Eq. (4.5), the adjusting parameters δ_σ in Eq. (5.6) and the independence sampler in Eq. (5.10) for the type (i) prior.*

Data	(c, d)	δ_σ
D1	(0, 0.866)	0.030
D2	(0, 0.866)	0.029
D3	(0, 0.866)	0.028

The convergence of our single long run was evaluated as follows. We divided the single long runs (initial $T^* = 1000$ steps were discarded as burn-in in each case) into five sequences of equal length for all data sets. Then the potential scale reduction factor $\hat{R}$ was calculated for each case (see Tables 6.3-6.4). In the Tables 6.3-6.4, the values of $\hat{R}$ are here well below 1.1 for all cases. This indicates that the length of each single long run is sufficient for sampling from the posterior density. Thus, for these single long runs, we have used 30000 samples for the data sets (D1,D2,D3) as samples from the posterior density $p(\alpha \mid X)$ or $p(\sigma \mid X)$.

Table 6.3: *Results of the values of the potential scale reduction factor $\hat{R}_\alpha$ as the convergence diagnostics for all data sets; '$\hat{R}_\alpha$ (U)' and '$\hat{R}_\alpha$ (I)' stand for the uniform jumping density and the independence sampler, respectively.*

Data	$\hat{R}_\alpha$ (U)	$\hat{R}_\alpha$ (I)
D1	1.0008	1.0003
D2	1.0007	1.0002
D3	1.0004	1.0016

Table 6.4: *Results of the values of the potential scale reduction factor* $\hat{R}_\sigma$ *as the convergence diagnostics for all data sets; '*$\hat{R}_\sigma$ *(U)' and '*$\hat{R}_\sigma$ *(I)' represent the uniform jumping density and the independence sampler, respectively.*

Data	$\hat{R}_\sigma$ (U)	$\hat{R}_\sigma$ (I)
D1	1.0001	1.0062
D2	1.0000	1.0017
D3	1.0000	1.0011

For the estimation α under the fixed τ, we show in Figs. 6.2(a)-(f) the time series plots of the Monte Carlo output of α for all data sets under the type (i) prior with the uniform jumping, and their histograms of α for the simulation. Figs. 6.3(a)-(f) show the time series plots of the Monte Carlo output of α for data D1 for the independence sampler, and their simulated histograms of α. Table 6.5 gives simulated posterior mean of α under the type (i) prior with the uniform jumping and with the independence sampler.

For the estimation τ under the fixed α, we give in Figs. 6.4(a)-(f) the time series plots of the Monte Carlo output of τ for all data sets under the type (i) prior with the uniform jumping, and their histograms of τ for the simulation. Figs. 6.5(a)-(f) show the time series plots of the Monte Carlo output of τ for all data sets for the independence sampler, and their simulated histograms of τ. Table 6.6 gives simulated posterior mean of τ under the type (i) prior with the uniform jumping and with the independence sampler. In the last column of Tables 6.5-6.6, for all data sets, the *maximum likelihood estimates* (MLE) are also given. We calculated the values of $\hat{\alpha}$ and $\hat{\tau}$ from the log-likelihood (3.14) by using the *quasi-Newton method*. For calculating the total potential energy in Eq. (3.14), we employed the periodic boundary condition again.

Table 6.5: *Posterior means of* α *under the type (i) prior for all data sets; '*$\overline{\alpha}_U$*' and '*$\overline{\alpha}_I$*' stand for mean for the case of the uniform jumping and of the independence sampler, respectively. The last column gives the maximum likelihood estimates (*$\hat{\alpha}$*).*

Data	$\overline{\alpha}_U$	$\overline{\alpha}_I$	$\hat{\alpha}$
D1	0.34477	0.34349	0.33865
D2	0.33297	0.33285	0.32906
D3	0.33762	0.33796	0.33587

Table 6.6: *Posterior means of τ under the type (i) prior for all data sets; '$\overline{\tau}_U$' and '$\overline{\tau}_I$' indicate mean for the case of the uniform jumping and of the independence sampler, respectively. The last column gives the maximum likelihood estimates ($\hat{\tau}$).*

Data	$\overline{\tau}_U$	$\overline{\tau}_I$	$\hat{\tau}$
D1	0.10948	0.10900	0.10984
D2	0.30694	0.30679	0.30723
D3	0.52315	0.52391	0.52360

From Table 6.5, we can see that, for the data D1 whose τ is relatively small, the posterior mean of α is slightly greater than the MLE for both types jumping density. It seems that the coincidence between the values of each posterior means and MLE is good for other cases. From Table 6.6, it is clear that, for all data sets, each posterior mean and MLE of τ are very close. From Figs. 6.2(b), (d) and (f), it seems that, as the degree of regularity of the data is getting lower, the posterior density of α tends to be more slightly spread.

From Figs. 6.2(a), (c) and (e), and Figs. 6.4(a), (c) and (e), for estimating α or τ using the uniform jumping, each run seems to be stationary. In these simulations, the acceptance rate were about 0.5. Then the values of $\hat{R}$ are well below 1.1 for the uniform jumping (see Tables 6.3-6.4). From the results, we consider that the use of the uniform jumping is suitable for inference. On the other hand, from Figs. 6.3(a), (c) and (e), and Figs. 6.5(a), (c) and (e), for the independence sampler, each run looks stationary. However, each acceptance rate was about 0.3, 0.2 and 0.1 for the data D1, D2 and D3, respectively. Since the independence sampler has no adjusting parameter, each value was lower than 0.5. Although the values of $\hat{R}$ are well below 1.1 for the independence sampler (see Tables 6.3-6.4), we found that the use of the independence sampler is not appropriate for inference.

In this subsection, we have presented the validity of our Bayesian procedure for the single-parameter Soft-Core models by using MCMC methods. In the next subsection, we will apply our procedure to the two-parameter Soft-Core models.

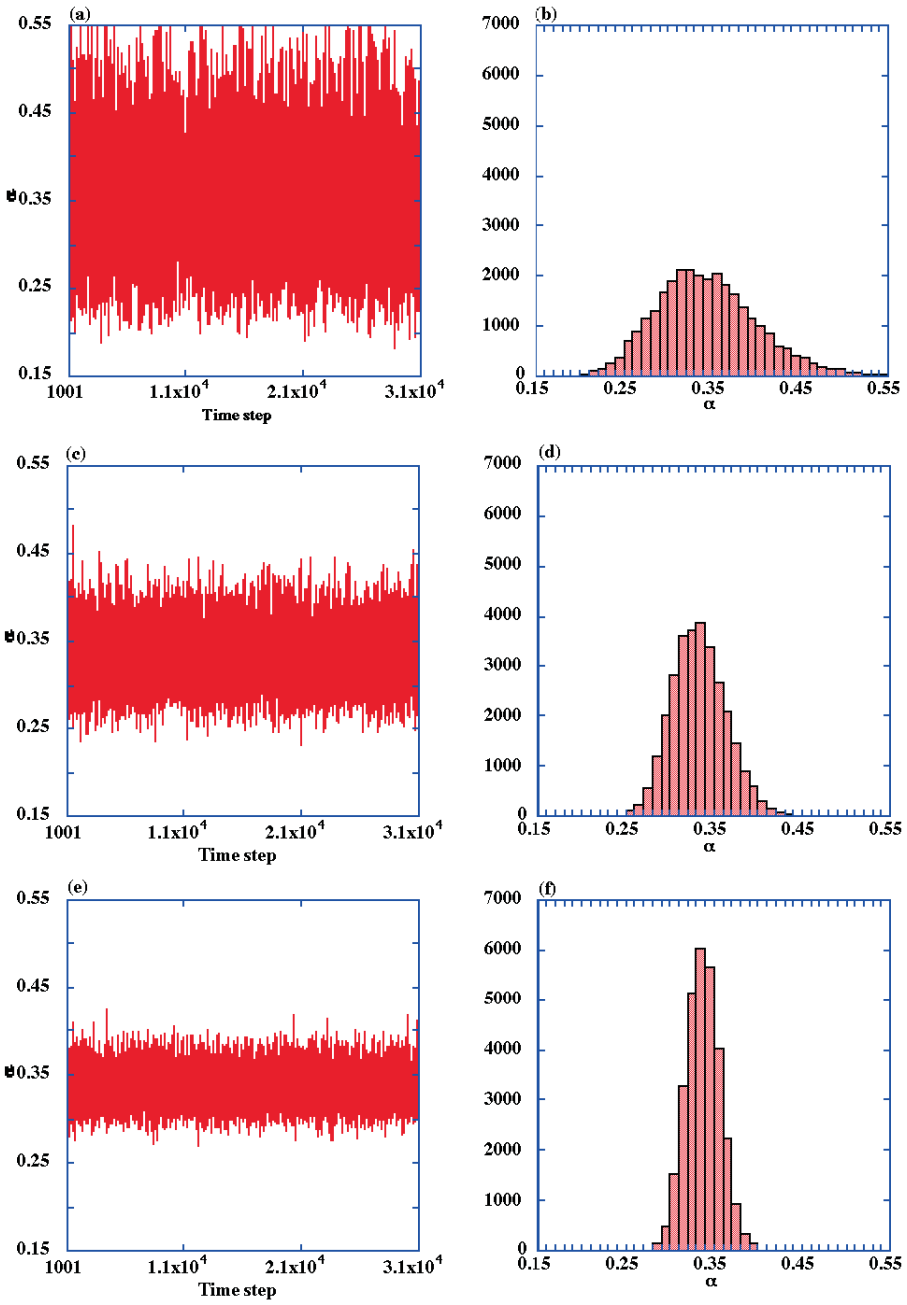

Figure 6.2: *Time series plots for α of simulation (after burn-in) for all data sets under the type (i) prior with the uniform jumping and their histograms of α of 30000 simulation draws; (a),(b): D1; (c),(d): D2; (e),(f): D3. For all histograms (b), (d) and (f), the class interval is 0.01.*

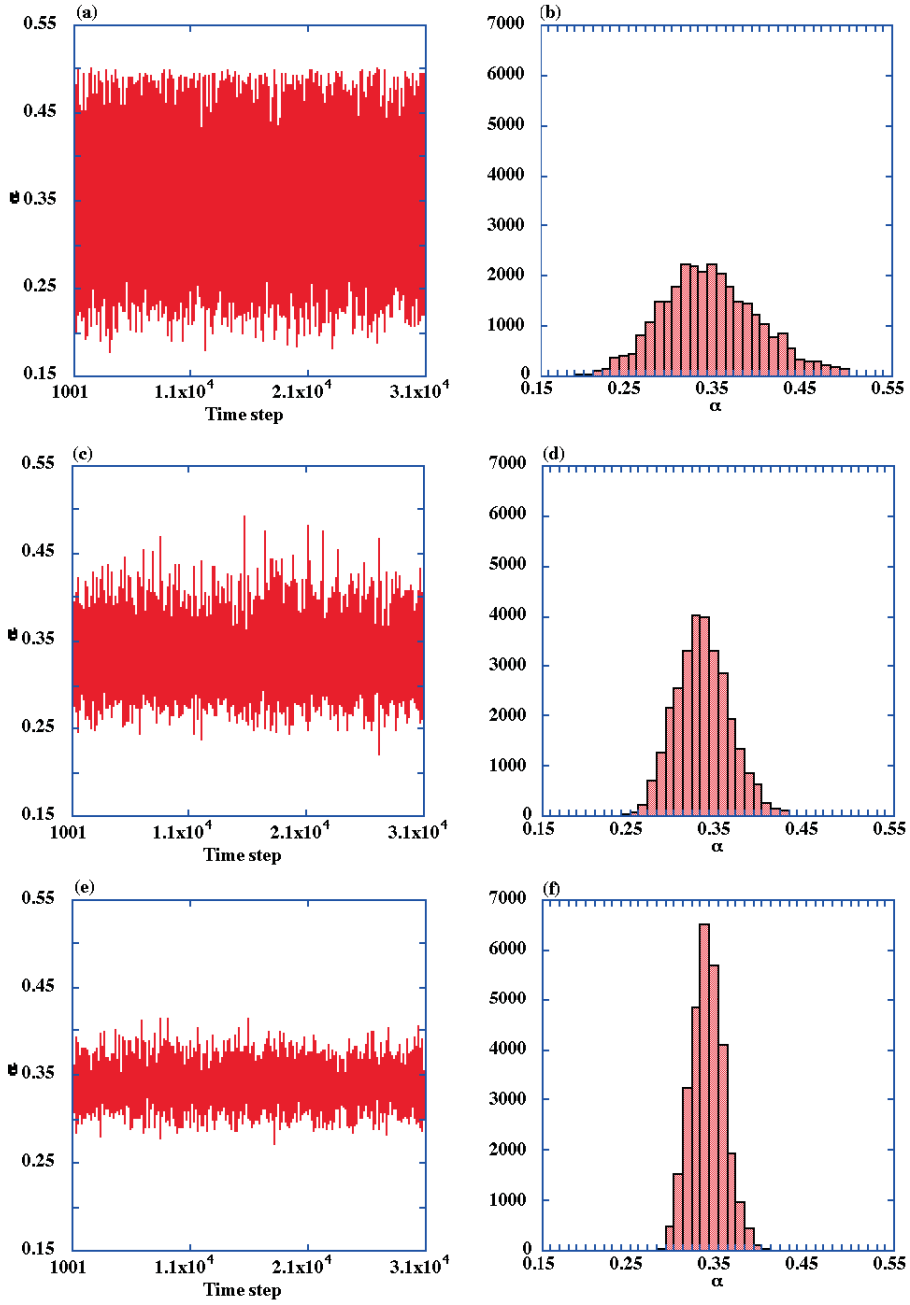

Figure 6.3: *Time series plots for α of simulation (after burn-in) for all data sets under the type (i) prior with the independence sampler and their histograms of α of 30000 simulation draws; (a),(b): D1; (c),(d): D2; (e),(f): D3. For all histograms (b), (d) and (f), the class interval is 0.01.*

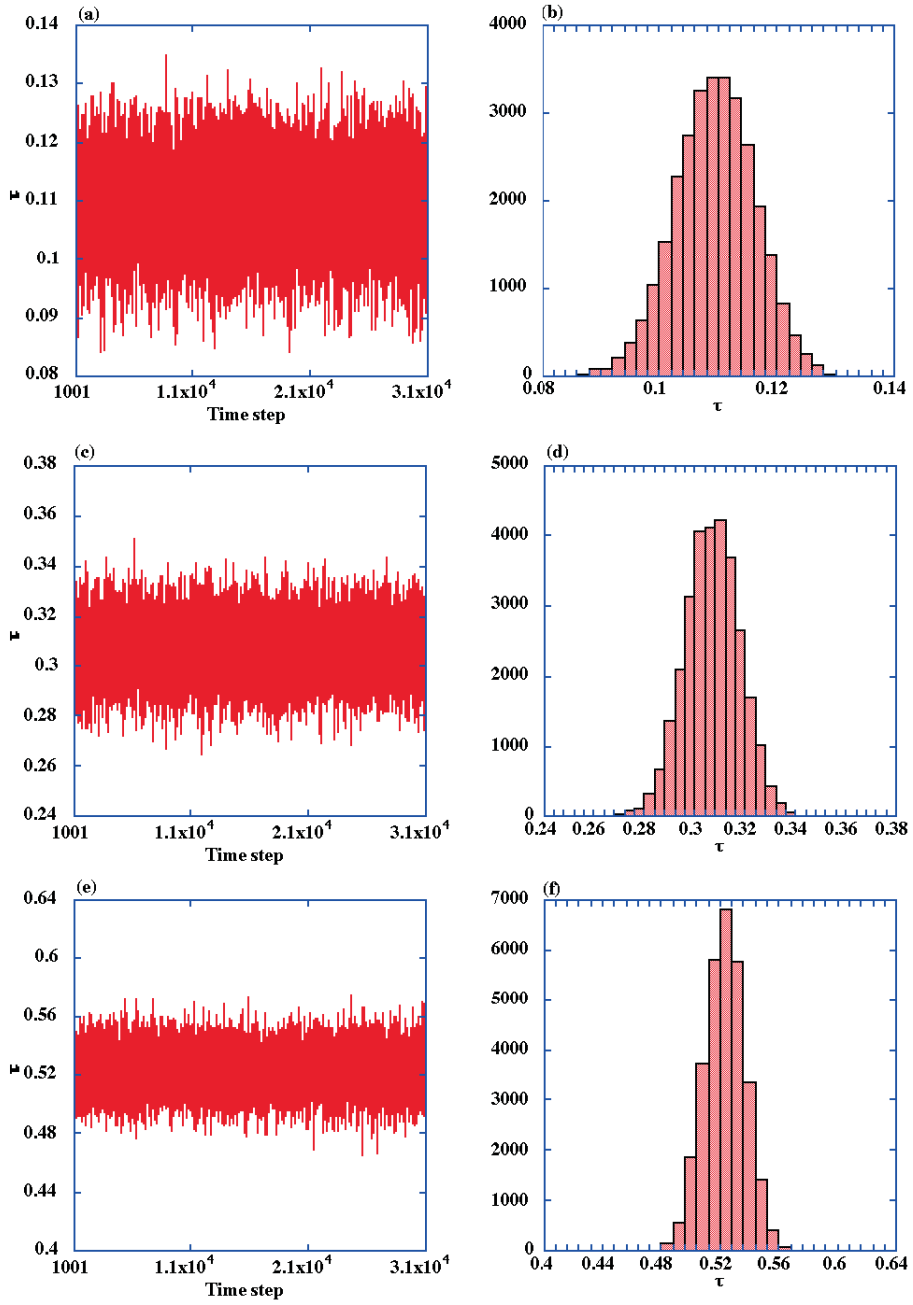

Figure 6.4: *Time series plots for τ of simulation (after burn-in) for all data sets under the type (i) prior with the uniform jumping and their histograms of τ of 30000 simulation draws; (a),(b): D1; (c),(d): D2; (e),(f): D3. For histograms (b), (d) and (f), each class interval is 0.002, 0.004 and 0.008, respectively.*

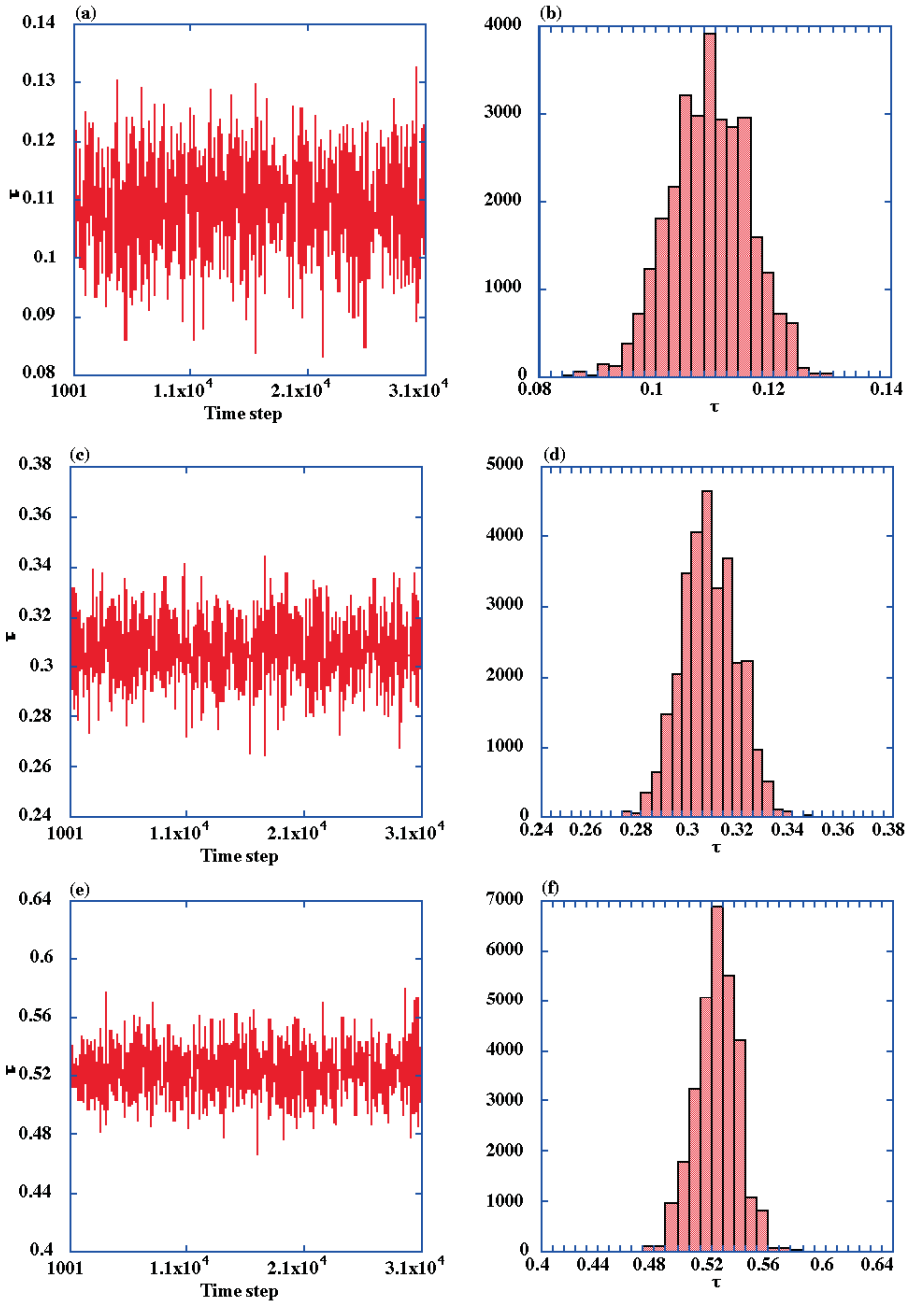

Figure 6.5: *Time series plots for τ of simulation (after burn-in) for all data sets under the type (i) prior with the independence sampler and their histograms of τ of 30000 simulation draws; (a),(b): D1; (c),(d): D2; (e),(f): D3. For histograms (b), (d) and (f), each class interval is 0.002, 0.004 and 0.008, respectively.*

6.1.2 Example: the two-parameter Soft-Core models

In previous subsection, we have verified our Bayesian procedure for the single-parameter Soft-Core models for the simulated data. In this subsection, we applied our Bayesian procedure to the two-parameter Soft-Core models (3.1) for the three simulated data sets (D1,D2,D3) illustrated in Figs. 6.1(a)-(c). The method proceeds as follows. The two parameters α and τ will be estimated simultaneously. Then, we adopted the type (i) prior densities of parameters $(\alpha, \tau$ $(i.e.\ \sigma))$ with the uniform jumping densities. We set the values of hyper-parameters (a, b, c, d) in Eqs. (4.4)-(4.5) for the type (i) prior densities of α and σ by considering the effective region. Then to specify jumping densities in Eqs. (5.5)-(5.6) for the type (i) priors, their adjusting parameters are chosen such that Markov chain simulation should have the acceptance rate about $0.40 \sim 0.44$ in two-dimension, as remarked in §5.2.2. These values are given in Table 6.7. Then, to estimate (α, τ), we performed a simulation of 31000 steps in the Metropolis-Hastings algorithm for each of the data sets (D1,D2,D3).

Table 6.7: *Prior and jumping specification: the values of hyperparameters* (a, b, c, d) *in Eqs. (4.4)-(4-5) and the adjusting parameters* $(\delta_\alpha, \delta_\sigma)$ *in Eqs. (5.5)-(5.6).*

Data	(a, b, c, d)	($\delta_\alpha, \delta_\sigma$)
D1	(0, 0.50, 0, 0.866)	(0.025, 0.038)
D2	(0, 0.50, 0, 0.866)	(0.020, 0.035)
D3	(0, 0.50, 0, 0.866)	(0.020, 0.035)

In order to investigate the convergence, the single long runs (initial $T^* = 1000$ steps were discarded as burn-in in each case) were divided into five sequences of equal length for all data sets, and the potential scale reduction factor $\hat{R}$ was calculated for parameters α and σ separately. Table 6.8 shows the values of $\hat{R}$ for respective parameters for all data sets. In the Table, we see that the values of $\hat{R}$ are well below 1.1 for all cases, that is, our single long runs are sufficient for sampling from the target posterior density. Thus, we have used 30000 samples for each of the data sets (D1,D2,D3) as samples from the joint posterior densities $p(\alpha, \sigma \mid X)$.

Table 6.8: *Results of the values of the potential scale reduction factor* $\hat{R}$ *as the convergence diagnostics for all data sets.*

Data	*parameter*	$\hat{R}$ (*single long run*)
D1	α	1.0099
	σ	1.0011
D2	α	1.0060
	σ	1.0023
D3	α	1.0190
	σ	1.0132

Table 6.9: *Posterior means of* (α, τ) *under the type (i) prior for all data sets; '*$\overline{\alpha}$*' and '*$\overline{\tau}$*' stand for mean for the case of the uniform jumping and of the independence sampler, respectively. The last column gives the maximum likelihood estimates of* (α, τ)*.*

Data	$\overline{\alpha}$	$\overline{\tau}$	$(\hat{\alpha}, \hat{\tau})$
D1	0.34347	0.10934	(0.33865, 0.10984)
D2	0.33547	0.30825	(0.32906, 0.30723)
D3	0.33833	0.52538	(0.33587, 0.52360)

Figs. 6.6, 6.7 and 6.8 display the time series plots for the Monte Carlo output of α and τ together with their histograms for all data sets under the type (i) prior. We can see that each run seems to be stationary. Table 6.9 shows posterior means of α and τ under the type (i) prior with the uniform jumping. In the last column of the table, the maximum likelihood estimates $(\hat{\alpha}, \hat{\tau})$ are also given.

We compare the results between this case and the case of the single-parameter models. From Tables 6.5 and 6.9, some small difference can be found between the values of each posterior means and each MLE, but it seems that their coincidence of the values is good for all cases. From Tables 6.6 and 6.9, we see that each posterior mean and MLE of τ are very close for all cases. Figs. 6.9(a)-(f) show posterior densities of α and τ under the type (i) prior in the cases of the single- and two-parameter models. From Figs. 6.9(a)-(b), the marginal posterior densities of the respective parameters are said to be similar for the data

D1. From Figs. 6.9(c)-(f), for the data sets D2 and D3, the marginal posterior densities of respective parameters for the independence sampler show smaller peak and slightly spread. For all data sets (D1, D2, D3), the coincidence between the shape of all marginal posterior densities of respective parameters using the uniform jumpings is good for the two models. From the results, it seems that as the degree of regularity of the data is getting lower, the posterior density of α tends to be more spread.

In this subsection, we have shown the validity of our Bayesian procedure for the two-parameter Soft-Core models. In the next subsection, we will apply our procedure to various point patterns for the two-parameter Soft-Core models.

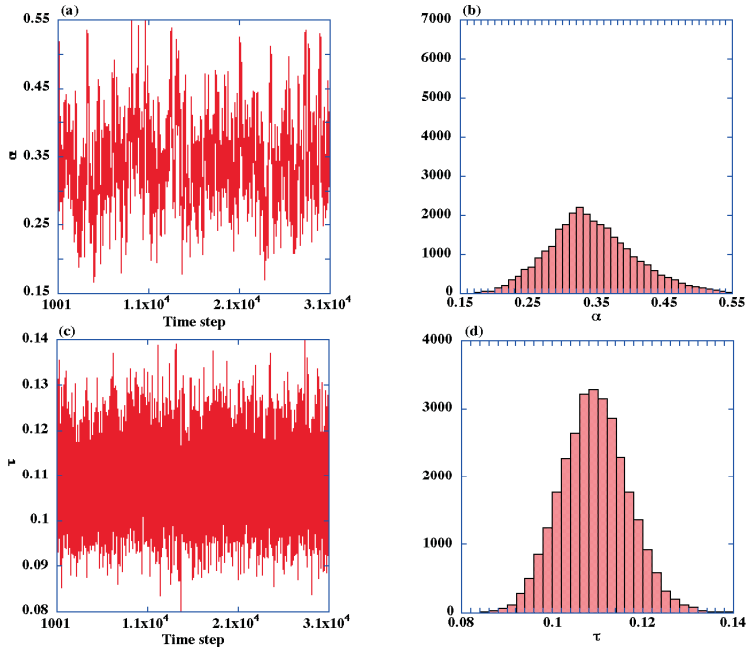

Figure 6.6: *Time series plots for α and τ of the single long run (after burn-in) for data D1 under the type (i) prior with the uniform jumping and their histograms of α of 30000 simulation draws; (a),(b): α; (c),(d): τ. For histograms (b) and (d), the class intervals of α and τ are 0.01 and 0.002, respectively.*

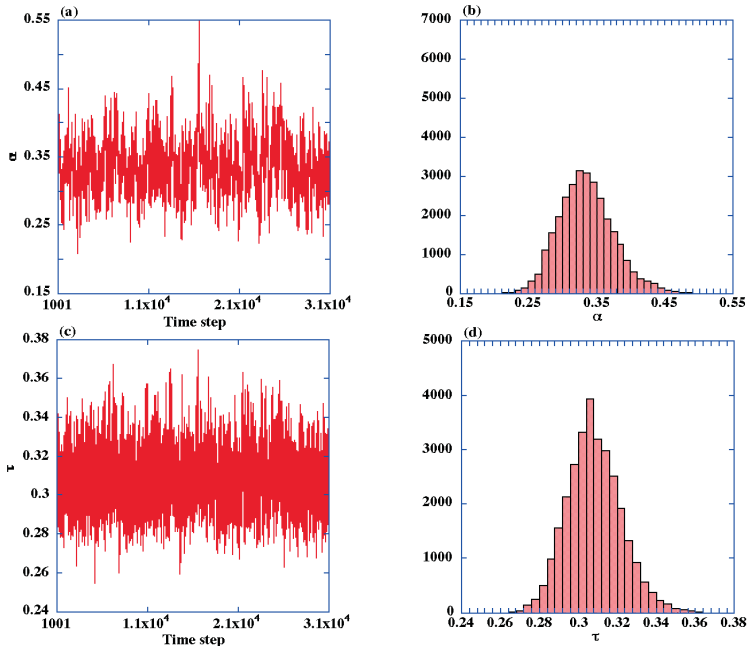

Figure 6.7: *Time series plots for α and τ of the single long run (after burn-in) for data D2 under the type (i) prior with the uniform jumping and their histograms of α of 30000 simulation draws; (a),(b): α; (c),(d): τ. For histograms (b) and (d), the class intervals of α and τ are 0.01 and 0.004, respectively.*

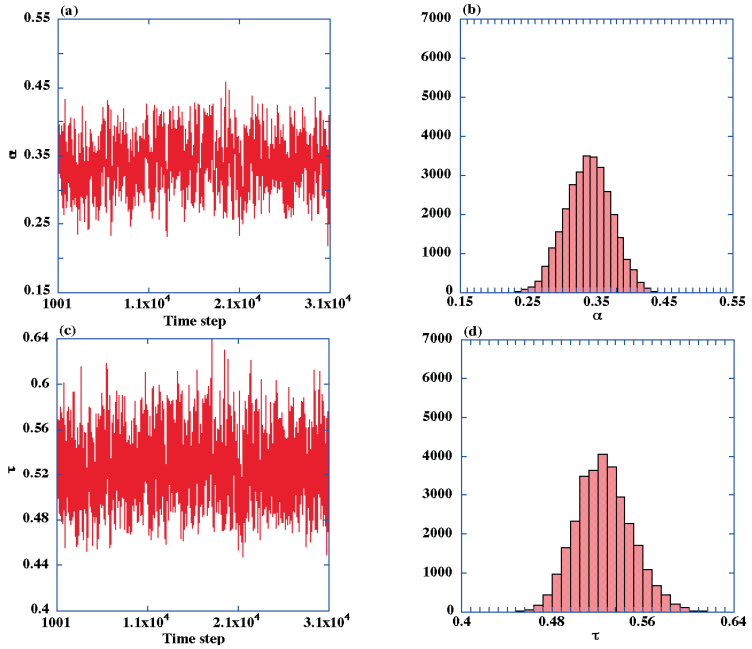

Figure 6.8: *Time series plots for α and τ of the single long run (after burn-in) for data D3 under the type (i) prior with the uniform jumping and their histograms of α of 30000 simulation draws; (a),(b): α; (c),(d): τ. For histograms (b) and (d), the class intervals of α and τ are 0.01 and 0.008, respectively.*

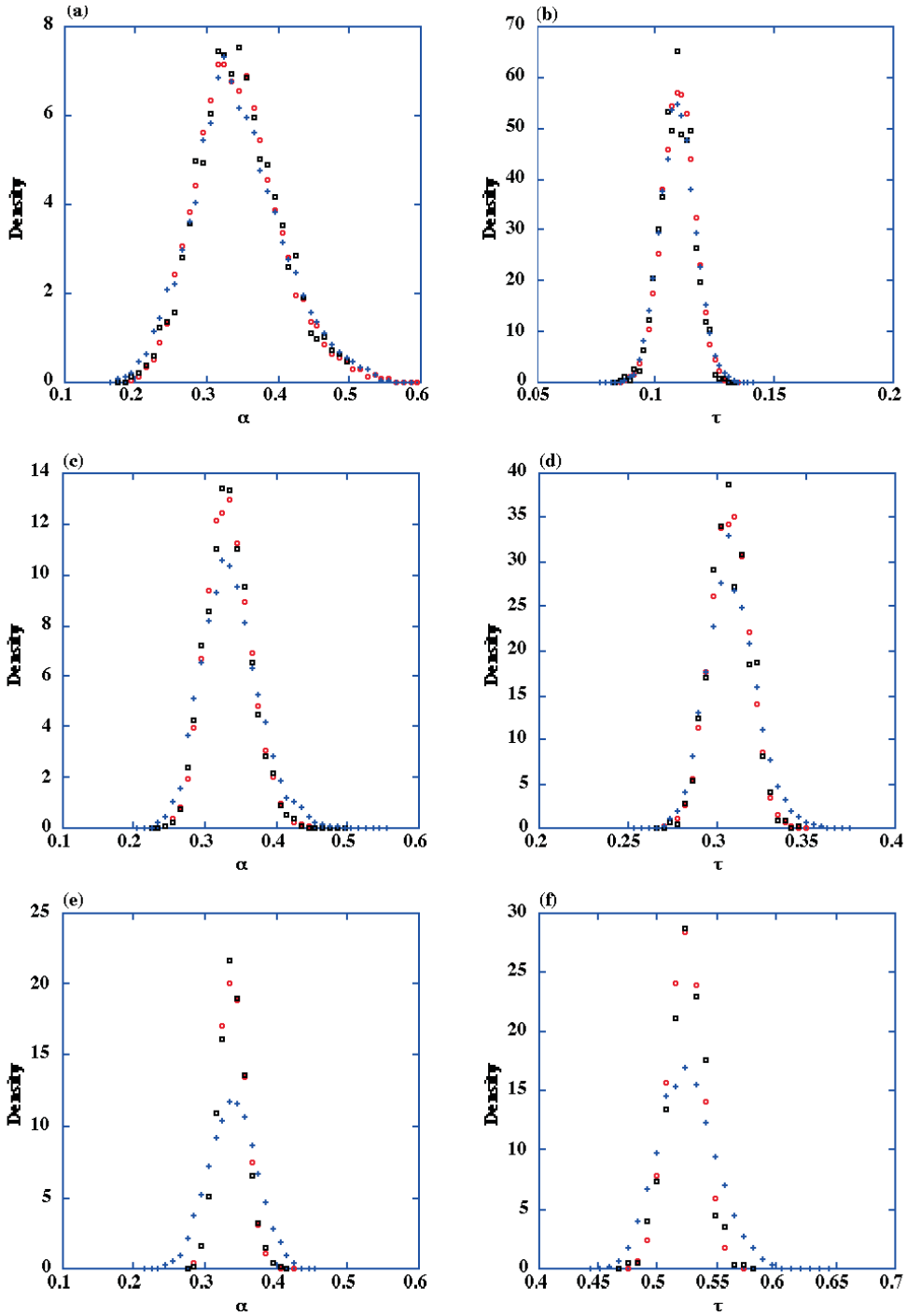

Figure 6.9: *Comparison of posterior densities of α and τ ($p(\alpha \mid X)$ and $p(\tau \mid X)$) under the type (i) prior (the symbol $\circ$ indicates the densities of the type (i) prior with uniform jumping and the symbol $+$ represents the densities of the type (i) prior with the independence sampler for the single-parameter models; the symbol $\square$ stands for the densities of the type (i) priors with the jumping densities for the two-parameter models). (a),(b): D1; (c),(d): D2; (e),(f): D3.*

6.1.3 Example: the two-parameter Soft-Core models for various large sample point patterns

In §6.1.1-6.1.2, we have found that our Bayesian procedure can be applied to the three simulated point patterns (D1, D2, D3). In practice, for observed data, it is probable that the parameters α and τ will take various values. In this subsection, we would like to apply our procedure to various simulated point patterns. Then, we performed our Bayesian estimation of the two-parameter Soft-Core models for simulated equilibrium point patterns generated by MCMC for the cases of $N = 500,\ V = \sqrt{500} \times \sqrt{500},\ \alpha = 0.2, 0.3, 0.4,\ \tau = 0.05, 0.1, 0.3, 0.5$. These data sets are named as follows; for $\alpha = 0.2$ corresponding to τ: (DL0.2-0.05, DL0.2-0.1, DL0.2-0.3, DL0.2-0.5), for $\alpha = 0.3$: (DL0.3-0.05, DL0.3-0.1, DL0.3-0.3, DL0.3-0.5) and for $\alpha = 0.4$: (DL0.4-0.05, DL0.4-0.1, DL0.4-0.3, DL0.4-0.5), respectively. These simulated point patterns are illustrated in Figs. 6.10, 6.11 and 6.12. From the figures, it is found that as the reduced density τ is getting larger, the degree of regularity increases.

As described in §4.2, we have considered two types of prior (the type (i) and (ii) prior) densities of the parameters. In the subsection, we carry out Bayesian inference under the two types of prior for the two-parameter Soft-Core models. Similarly as the previous subsection, to specify the type (i) prior density of α, we put the values of hyperparameters $(a, b) = (0, 0.5)$ in Eq. (4.4) for all data sets. By considering the relation between σ and τ of Eq. (3.4), we set the values of hyperparameters (c, d) for σ in Eq. (4.5): $(c, d) = (0, 0.224)$ for $\tau = 0.05$; $(c, d) = (0, 0.316)$ for $\tau = 0.1$; $(c, d) = (0, 0.548)$ for $\tau = 0.3$; $(c, d) = (0, 0.707)$ for $\tau = 0.5$, respectively. For the type (ii) prior densities of the parameters, their means (μ_α, μ_σ) were chosen as the center of the range of each parameter, and their standard deviations (s_α, s_σ) for α and σ are respectively chosen as the values of half of the mean. Then we put the values of hyperparameters $(\mu_\alpha, s_\alpha) = (0.25, 0.125)$ and $(\mu_\sigma, s_\sigma) = (0.612, 0.306)$ in Eqs. (4.6)-(4.7) for all data sets, respectively. These normal priors are supposed to cover the range $0 < \alpha \leq 0.5$ and $0 < \tau \leq 0.75$, as stated in §3.2.

For both types (i) and (ii) of prior, we applied three types of jumping densities: uniform, normal and independence sampler. For uniform and normal jumping densities, the adjusting parameters in Eqs. (5.5)-(5.8) were chosen such that Markov chain simulation should have the acceptance rate about $0.40 \sim 0.44$ in two-dimension of the Metropolis-Hastings steps, as remarked in §5.2.2. The values of the adjusting parameters are given in Tables 6.10-6.13. For the independence samplers of respective parameters, the values of the hyperparameters $(a, b) = (0, 0.5)$ were set in Eq. (5.9). And corresponding to the value of τ for each data set, the values of (c, d) were put in Eq. (5.10) as the same values of prior as described above.

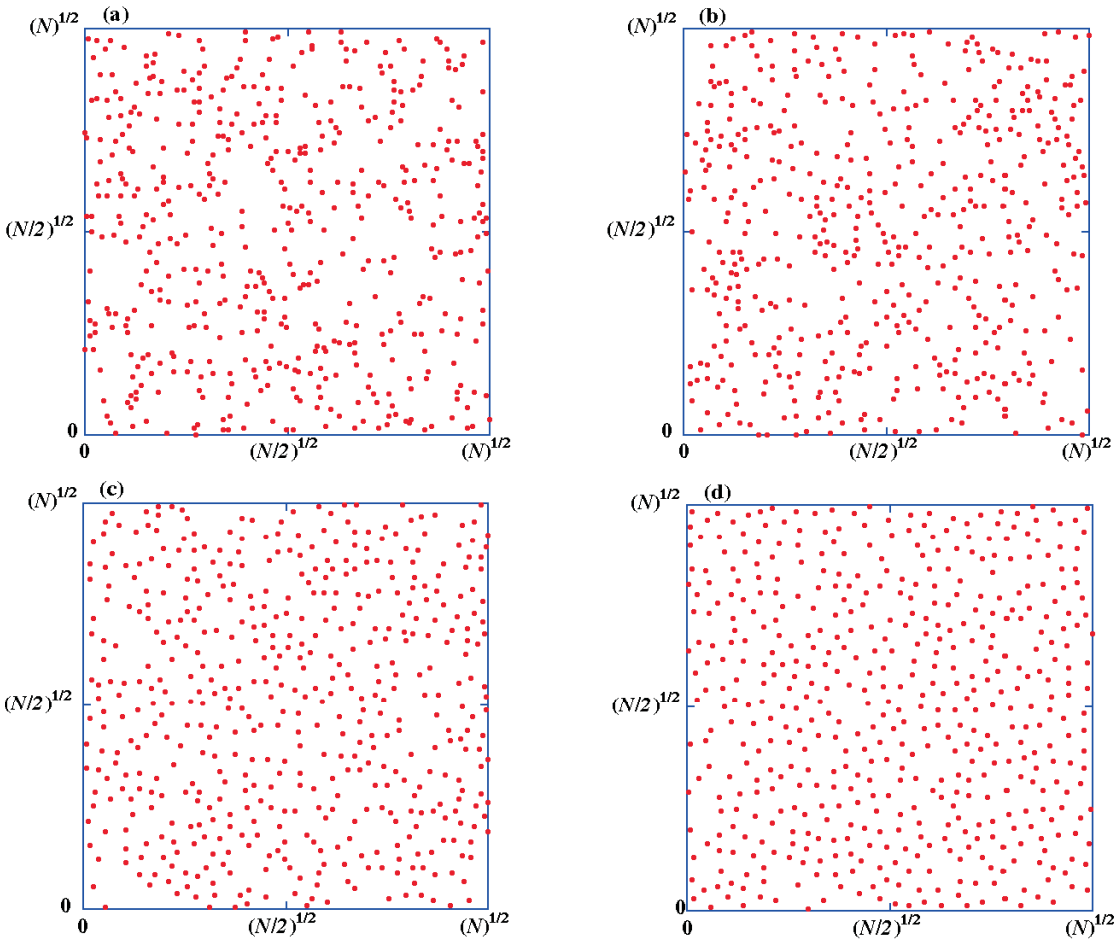

Figure 6.10: *Simulated equilibrium point patterns* ($N = 500, V = \sqrt{500} \times \sqrt{500}, \alpha = 0.2$). *(a):* $\tau = 0.05$ *(DL0.2-0.05); (b):* $\tau = 0.1$ *(DL0.2-0.1); (c):* $\tau = 0.3$ *(DL0.2-0.3) ; (d):* $\tau = 0.5$ *(DL0.2-0.5).*

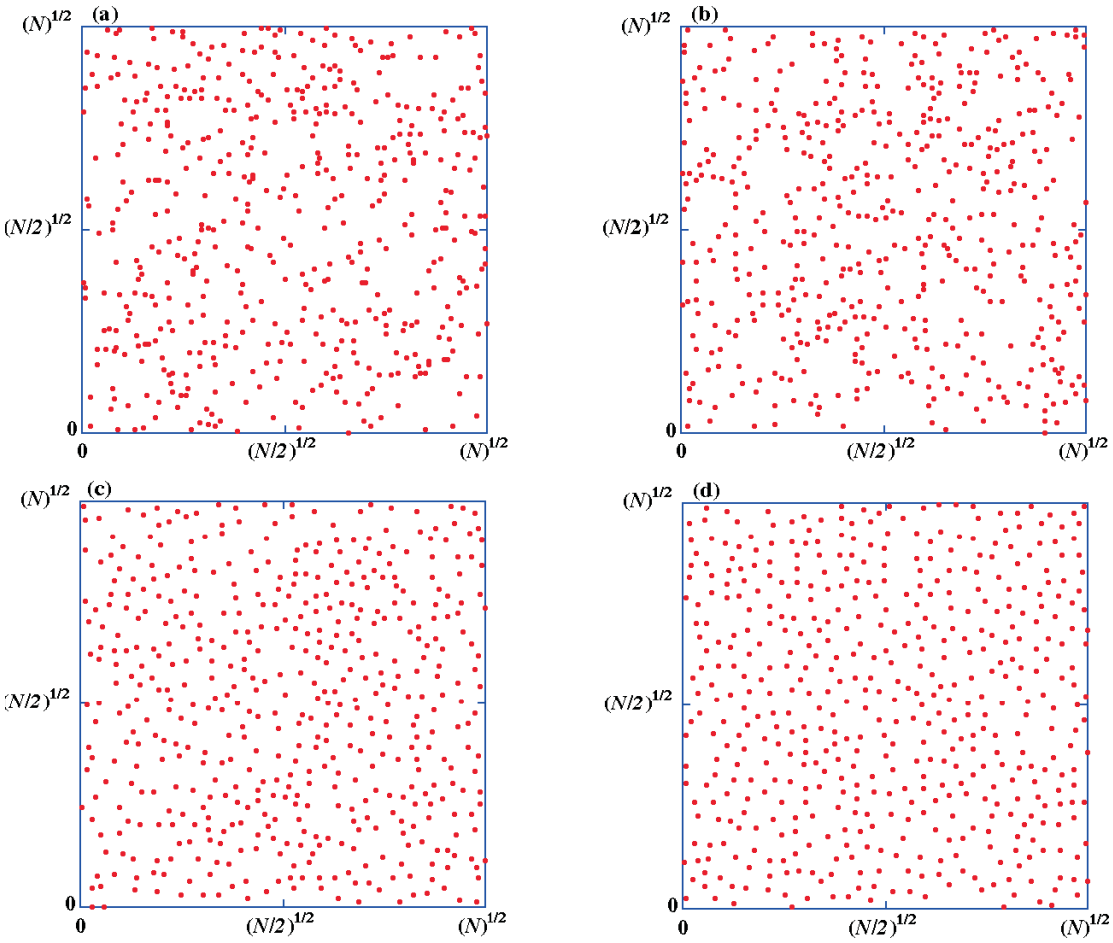

Figure 6.11: *Simulated equilibrium point patterns* ($N = 500, V = \sqrt{500} \times \sqrt{500}, \alpha = 0.3$). *(a):* $\tau = 0.05$ *(DL0.3-0.05); (b):* $\tau = 0.1$ *(DL0.3-0.1); (c):* $\tau = 0.3$ *(DL0.3-0.3) ; (d):* $\tau = 0.5$ *(DL0.3-0.5).*

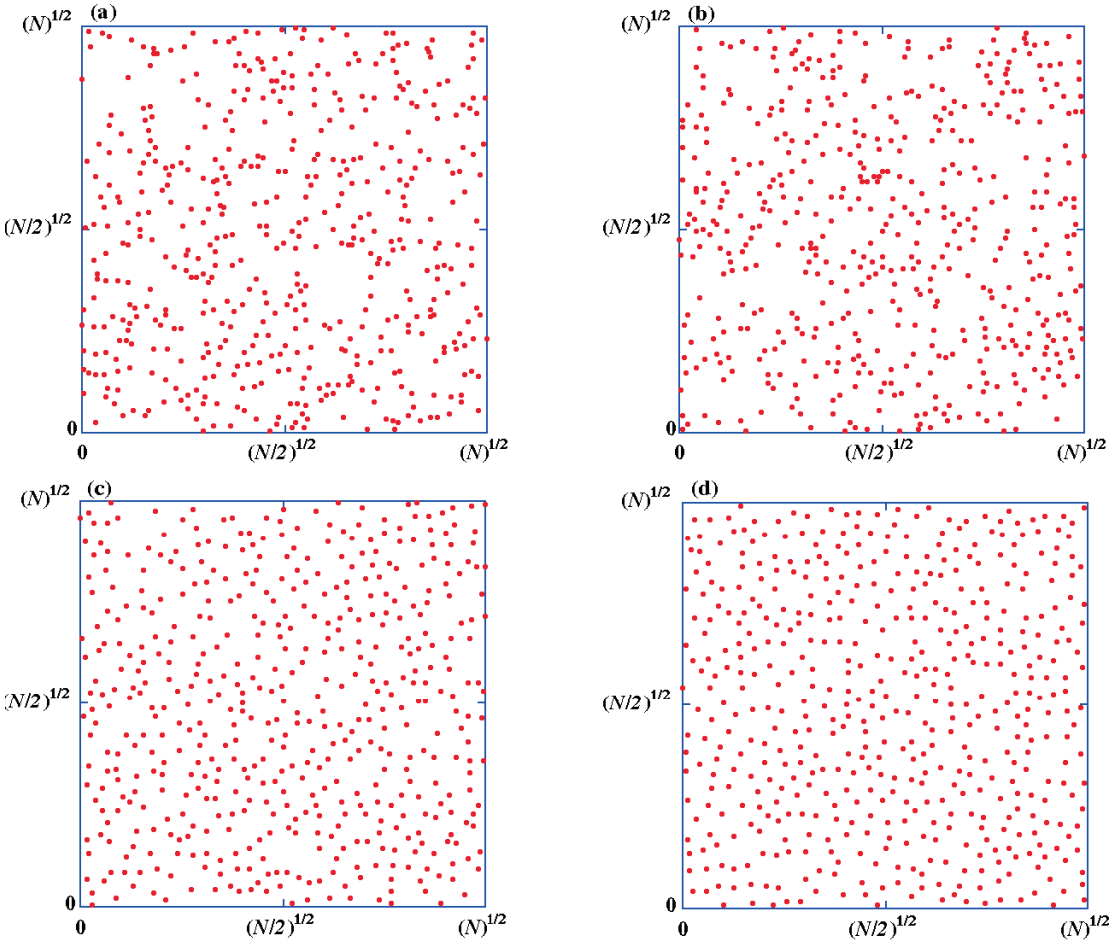

Figure 6.12: *Simulated equilibrium point patterns* ($N = 500, V = \sqrt{500} \times \sqrt{500}, \alpha = 0.4$). *(a):* $\tau = 0.05$ *(DL0.4-0.05); (b):* $\tau = 0.1$ *(DL0.4-0.1); (c):* $\tau = 0.3$ *(DL0.4-0.3) ; (d):* $\tau = 0.5$ *(DL0.4-0.5).*

Table 6.10: *Type (i) prior with uniform jumping specification: the values of the adjusting parameters* $(\delta_\alpha, \delta_\sigma)$ *in Eqs. (5.5)-(5.6) for the uniform jumpings.*

Data	$(\delta_\alpha, \delta_\sigma)$
$DL0.2-0.05$	(0.050, 0.050)
$DL0.2-0.1$	(0.023, 0.030)
$DL0.2-0.3$	(0.018, 0.026)
$DL0.2-0.5$	(0.017, 0.026)
$DL0.3-0.05$	(0.040, 0.040)
$DL0.3-0.1$	(0.021, 0.035)
$DL0.3-0.3$	(0.019, 0.034)
$DL0.3-0.5$	(0.017, 0.031)
$DL0.4-0.05$	(0.030, 0.030)
$DL0.4-0.1$	(0.025, 0.029)
$DL0.4-0.3$	(0.019, 0.029)
$DL0.4-0.5$	(0.013, 0.027)

Table 6.11: *Type (i) prior with normal jumping specification: the values of the adjusting parameters* $(\delta_\alpha^2, \delta_\sigma^2)$ *in Eqs. (5.7)-(5.8) for the normal jumpings.*

Data	$(\delta_\alpha^2, \delta_\sigma^2)$
$DL0.2-0.05$	(0.050, 0.050)
$DL0.2-0.1$	(0.011, 0.021)
$DL0.2-0.3$	(0.009, 0.018)
$DL0.2-0.5$	(0.009, 0.018)
$DL0.3-0.05$	(0.027, 0.027)
$DL0.3-0.1$	(0.015, 0.023)
$DL0.3-0.3$	(0.013, 0.022)
$DL0.3-0.5$	(0.013, 0.021)
$DL0.4-0.05$	(0.021, 0.028)
$DL0.4-0.1$	(0.019, 0.028)
$DL0.4-0.3$	(0.019, 0.027)
$DL0.4-0.5$	(0.016, 0.024)

Table 6.12: *Type (ii) prior with uniform jumping specification: the values of the adjusting parameters* $(\delta_\alpha, \delta_\sigma)$ *in Eqs. (5.5)-(5.6) for the uniform jumpings.*

Data	$(\delta_\alpha, \delta_\sigma)$
$DL0.2-0.05$	(0.030, 0.030)
$DL0.2-0.1$	(0.023, 0.030)
$DL0.2-0.3$	(0.018, 0.026)
$DL0.2-0.5$	(0.017, 0.026)
$DL0.3-0.05$	(0.036, 0.036)
$DL0.3-0.1$	(0.019, 0.033)
$DL0.3-0.3$	(0.017, 0.030)
$DL0.3-0.5$	(0.017, 0.030)
$DL0.4-0.05$	(0.025, 0.025)
$DL0.4-0.1$	(0.031, 0.038)
$DL0.4-0.3$	(0.028, 0.038)
$DL0.4-0.5$	(0.027, 0.037)

Table 6.13: *Type (ii) prior with normal jumping specification: the values of the adjusting parameters* $(\delta_\alpha^2, \delta_\sigma^2)$ *in Eqs. (5.7)-(5.8) for the normal jumpings.*

Data	$(\delta_\alpha^2, \delta_\sigma^2)$
$DL0.2-0.05$	(0.018, 0.020)
$DL0.2-0.1$	(0.011, 0.021)
$DL0.2-0.3$	(0.009, 0.018)
$DL0.2-0.5$	(0.009, 0.018)
$DL0.3-0.05$	(0.025, 0.025)
$DL0.3-0.1$	(0.015, 0.023)
$DL0.3-0.3$	(0.013, 0.022)
$DL0.3-0.5$	(0.013, 0.020)
$DL0.4-0.05$	(0.020, 0.026)
$DL0.4-0.1$	(0.014, 0.025)
$DL0.4-0.3$	(0.019, 0.027)
$DL0.4-0.5$	(0.016, 0.024)

Then, to estimate (α, τ) simultaneously, we performed a simulation of 26000 steps in the Metropolis-Hastings algorithm for all data sets. For the investigation of the convergence, we assessed the burn-in time T^* and the stopping time T using similar methods described as previous subsection. As a result, the values of $\hat{R}$ for both parameters were well below 1.1 for all cases. Then we estimated the burn-in and the stopping time $(T^*, T) = (1000, 26000)$. Therefore, we have used 25000 samples for all data sets as samples from the joint posterior densities $p(\alpha, \tau \mid X)$. We will explain the assessment of convergence in detail in §8.2.

In Tables 6.14, 6.15 and 6.16, the values of simulated posterior mean of respective parameters under the both types (i) and (ii) of prior with each of three types of the jumping density are shown for all data sets, respectively. In these Tables, the maximum likelihood estimates (MLE) of α and τ are also given. From Tables 6.14-6.16, we can see that each posterior mean and each MLE of τ are very close in every data set DL0.2, DL0.3 and DL0.4. From Table 6.14, it can be seen that each posterior mean of α is slightly greater than each MLE. There, it seems that, for the data whose τ is large, the difference of the mean and MLE tends to be small. In Table 6.15, we can see that, under the type (i) prior, each posterior mean of α is slightly greater than each MLE. Then, we also see that, under the type (ii) prior, for the data whose τ is large, each mean is close to each MLE. From Table 6.16, it is found that, under the type (i) prior, each posterior mean of α is close to each MLE except the case of DL0.4-0.05. Then, we can see that, under the type (ii) prior, for the data whose τ is large, each mean is smaller than each MLE.

Table 6.14: *Posterior means of* (α, τ)*, indicated by '*$\overline{\alpha}$*' and '*$\overline{\tau}$*', under both types (i) and (ii) of prior with three jumpings for data sets (DL0.2-0.05, DL0.2-0.1, DL0.2-0.3, DL0.2-0.5); 'prior-jump' stands for combinations of the prior and jumping for respective parameters. In the last column, the maximum likelihood estimates of* (α, τ) *are also given.*

Data	$prior - jump$	$\overline{\alpha}$	$\overline{\tau}$	$(\hat{\alpha}, \hat{\tau})$
DL0.2-0.05	uni-uni	0.19166	0.052369	(0.18285, 0.052707)
	uni-nor	0.19333	0.052136	
	nor-uni	0.21444	0.052878	
	nor-nor	0.20251	0.052607	
	uni-ind	0.18528	0.052064	
	nor-ind	0.20643	0.052585	
DL0.2-0.1	uni-uni	0.23406	0.099200	(0.19823, 0.099319)
	uni-nor	0.23412	0.099037	
	nor-uni	0.23378	0.099308	
	nor-nor	0.23603	0.099339	
	uni-ind	0.21835	0.099127	
	nor-ind	0.22660	0.098952	
DL0.2-0.3	uni-uni	0.21377	0.30403	(0.20417, 0.30301)
	uni-nor	0.21426	0.30385	
	nor-uni	0.21678	0.30428	
	nor-nor	0.21965	0.30482	
	uni-ind	0.21234	0.30385	
	nor-ind	0.21801	0.30387	
DL0.2-0.5	uni-uni	0.24185	0.53467	(0.22425, 0.53496)
	uni-nor	0.24847	0.53847	
	nor-uni	0.24505	0.53615	
	nor-nor	0.24227	0.53440	
	uni-ind	0.23852	0.53381	
	nor-ind	0.24045	0.53486	

Table 6.15: *Posterior means of* (α, τ)*, indicated by '*$\overline{\alpha}$*' and '*$\overline{\tau}$*', under both types (i) and (ii) of prior with three jumpings for data sets (DL0.3-0.05, DL0.3-0.1, DL0.3-0.3, DL0.3-0.5); 'prior-jump' stands for combinations of the prior and jumping for respective parameters. In the last column, the maximum likelihood estimates of* (α, τ) *are also given.*

Data	$prior - jump$	$\overline{\alpha}$	$\overline{\tau}$	$(\hat{\alpha}, \hat{\tau})$
DL0.3-0.05	uni-uni	0.33957	0.052857	(0.32021, 0.053772)
	uni-nor	0.34449	0.052886	
	nor-uni	0.32034	0.053121	
	nor-nor	0.31493	0.052961	
	uni-ind	0.32998	0.053202	
	nor-ind	0.30868	0.052979	
DL0.3-0.1	uni-uni	0.28268	0.10609	(0.26578, 0.10650)
	uni-nor	0.27960	0.10575	
	nor-uni	0.26952	0.10581	
	nor-nor	0.26539	0.10559	
	uni-ind	0.26735	0.10491	
	nor-ind	0.27078	0.10495	
DL0.3-0.3	uni-uni	0.30985	0.29714	(0.30814, 0.29729)
	uni-nor	0.30999	0.29694	
	nor-uni	0.30248	0.29579	
	nor-nor	0.30679	0.29664	
	uni-ind	0.31508	0.29788	
	nor-ind	0.29787	0.29471	
DL0.3-0.5	uni-uni	0.29067	0.50413	(0.28779, 0.50228)
	uni-nor	0.29203	0.50535	
	nor-uni	0.28786	0.50265	
	nor-nor	0.28360	0.50058	
	uni-ind	0.29484	0.50810	
	nor-ind	0.28155	0.50113	

Table 6.16: *Posterior means of* (α, τ)*, indicated by '*$\overline{\alpha}$*' and '*$\overline{\tau}$*', under both types (i) and (ii) of prior with three jumpings for data sets (DL0.4-0.05, DL0.4-0.1, DL0.4-0.3, DL0.4-0.5); 'prior-jump' stands for combinations of the prior and jumping for respective parameters. In the last column, the maximum likelihood estimates of* (α, τ) *are also given.*

Data	$prior - jump$	$\overline{\alpha}$	$\overline{\tau}$	$(\hat{\alpha}, \hat{\tau})$
DL0.4-0.05	uni-uni	0.41098	0.059886	(0.37402, 0.060521)
	uni-nor	0.40758	0.059831	
	nor-uni	0.36051	0.059620	
	nor-nor	0.36660	0.059789	
	uni-ind	0.38238	0.059593	
	nor-ind	0.35914	0.059728	
DL0.4-0.1	uni-uni	0.38346	0.091055	(0.38955, 0.091530)
	uni-nor	0.38909	0.091168	
	nor-uni	0.35688	0.090693	
	nor-nor	0.35851	0.090811	
	uni-ind	0.37372	0.091354	
	nor-ind	0.35429	0.090954	
DL0.4-0.3	uni-uni	0.41678	0.32409	(0.41113, 0.32269)
	uni-nor	0.41948	0.32442	
	nor-uni	0.40065	0.31955	
	nor-nor	0.40057	0.31979	
	uni-ind	0.41491	0.32244	
	nor-ind	0.39633	0.31800	
DL0.4-0.5	uni-uni	0.39236	0.52426	(0.38955, 0.52237)
	uni-nor	0.39100	0.52330	
	nor-uni	0.38179	0.51757	
	nor-nor	0.38090	0.51712	
	uni-ind	0.39154	0.52537	
	nor-ind	0.38394	0.51999	

Figs. 6.13, 6.14 and 6.15 display the comparison of the marginal posterior densities of α and τ ($p(\alpha \mid X)$ and $p(\tau \mid X)$) under both types (i) and (ii) of prior with three types of the jumping densities for all data sets, respectively. From Figs. 6.13, 6.14 and 6.15, we can see that, for all data sets, the coincidence of the shape of the marginal posterior densities of τ between the type (i) and (ii) priors is good for four kinds of combinations of the two priors and the uniform and normal jumpings. From Figs. 6.13(c), (e) and (g), it seems that the marginal posterior densities of α between the type (i) and (ii) priors is nearly similar for DL0.2-0.1, DL0.2-0.3 and DL0.2-0.5. For the DL0.2-0.05, we can see that the marginal posterior density of α under the type (ii) prior shifts to the right slightly as indicated in Fig. 6.13(a). It can be seen from Figs. 6.14(e) and (g) and 6.15(e) and (g) that, for the data sets (DL0.3-0.3, DL0.3-0.5) and (DL0.4-0.3, DL0.4-0.5), the marginal posteriors of α under the type (ii) prior shift to the left slightly. And from Figs. 6.14(a) and (c) and 6.15(a) and (c), it seems that, for the (DL0.3-0.05, DL0.3-0.1) and (DL0.4-0.05, DL0.4-0.1), the marginal posteriors of α under the type (ii) prior show the slightly larger peak and the slightly narrow than that under the type (i) prior. The slightly different spread of marginal posteriors of α between the type (i) and (ii) priors might be the influence of the choice of priors. For the cases of the independence samplers, because each acceptance ratio was low, the use of the independence samplers is not suitable for inference.

In this subsection, we have verified our Bayesian procedure for the two-parameter Soft-Core models for various point patterns for the case of $N = 500$. In the next section, we will apply our procedure to various small sample point patterns.

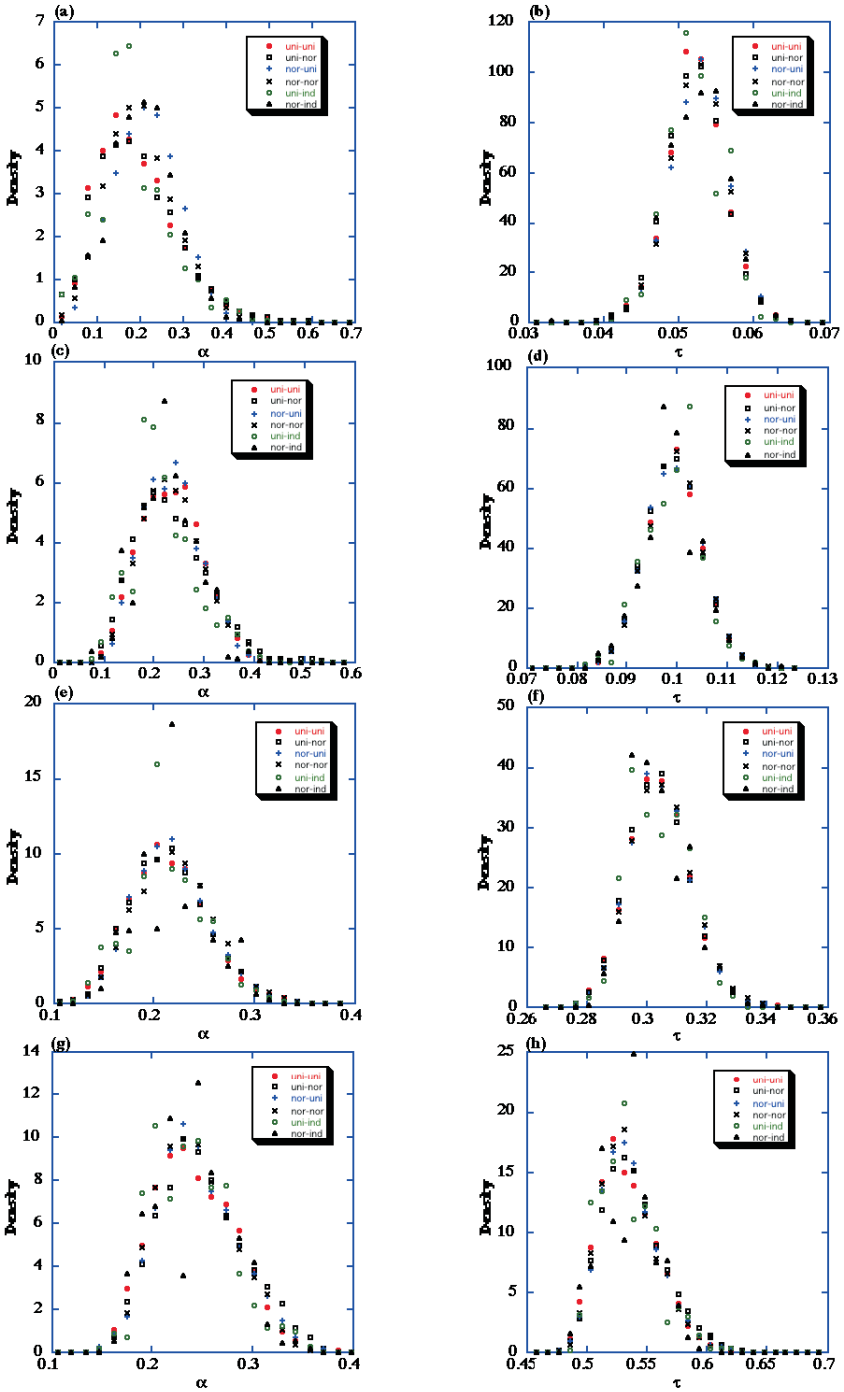

Figure 6.13: *Comparison of marginal posterior densities of α and τ under the type (i) prior with uniform jumping (uni-uni:●), normal jumping (uni-nor:□) and independence sampler (uni-ind:○), respectively, and under the type (ii) prior with uniform jumping (nor-uni:+), normal jumping (nor-nor:×) and independence sampler (nor-ind:△), respectively. (a),(b): DL0.2-0.05; (c),(d): DL0.2-0.1; (e),(f): DL0.2-0.3; (g),(h): DL0.2-0.5.*

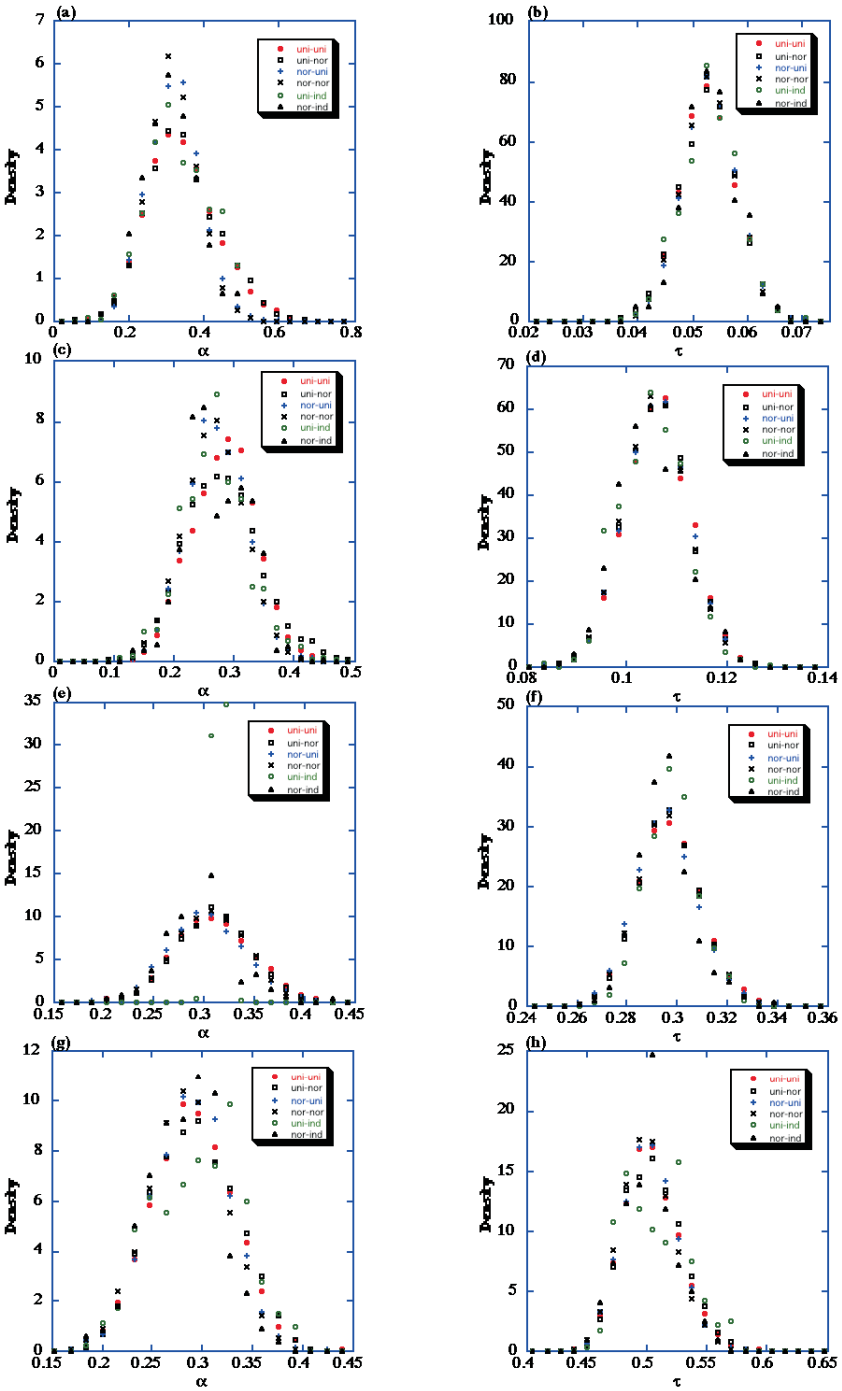

Figure 6.14: *Comparison of marginal posterior densities of α and τ under the type (i) prior with uniform jumping (uni-uni:●), normal jumping (uni-nor:□) and independence sampler (uni-ind:○), respectively, and under the type (ii) prior with uniform jumping (nor-uni:+), normal jumping (nor-nor:×) and independence sampler (nor-ind:△), respectively. (a),(b): DL0.3-0.05; (c),(d): DL0.3-0.1; (e),(f): DL0.3-0.3; (g),(h): DL0.3-0.5.*

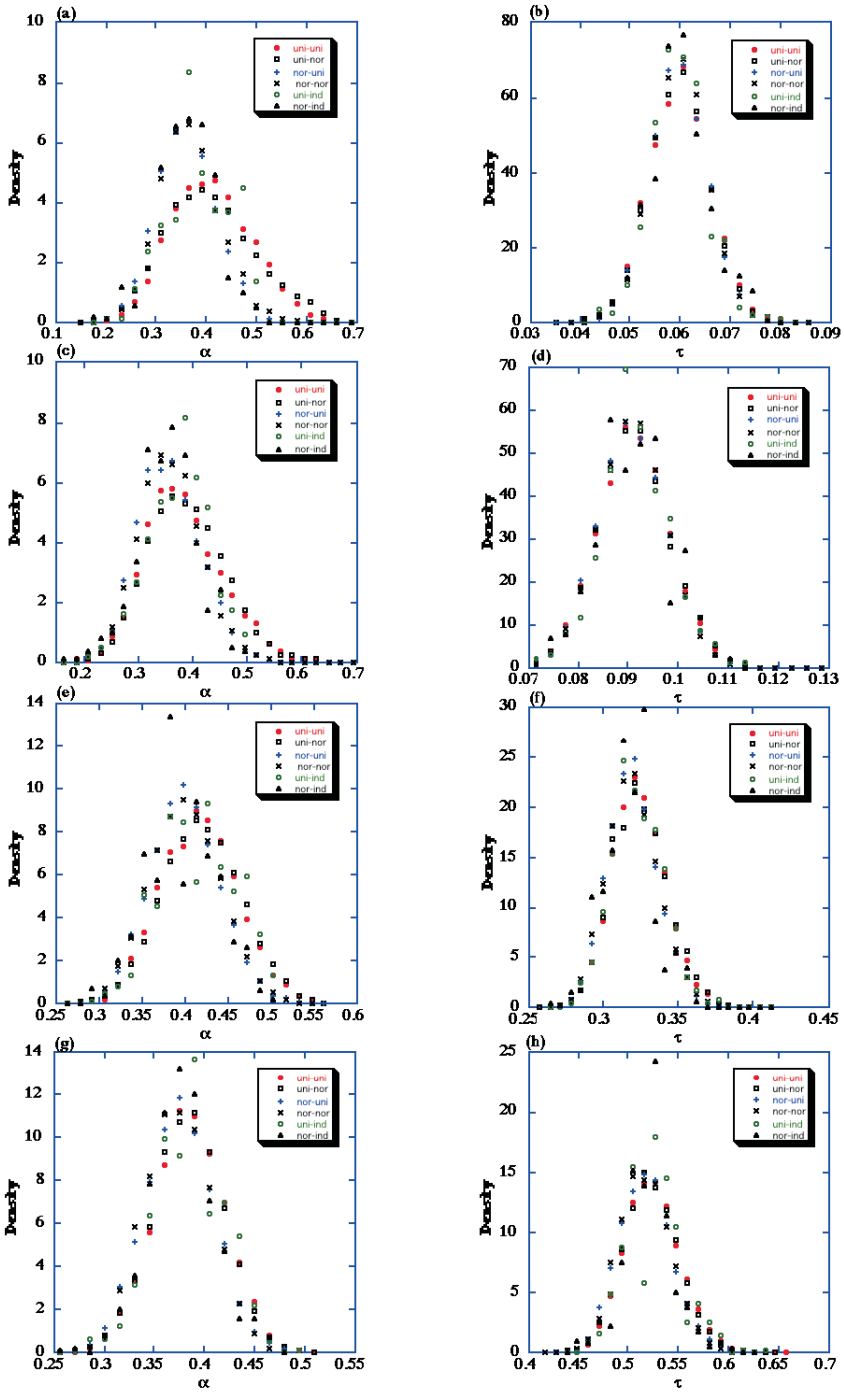

Figure 6.15: *Comparison of marginal posterior densities of α and τ under the type (i) prior with uniform jumping (uni-uni:●), normal jumping (uni-nor:□) and independence sampler (uni-ind:○), respectively, and under the type (ii) prior with uniform jumping (nor-uni:+), normal jumping (nor-nor:×) and independence sampler (nor-ind:△), respectively. (a),(b): DL0.4-0.05; (c),(d): DL0.4-0.1; (e),(f): DL0.4-0.3; (g),(h): DL0.4-0.5.*

6.2 Small sample point patterns; example: the two-parameter Soft-Core models for various small sample point patterns

In the previous section, we have verified that our Bayesian procedure can be applied to the simulated point patterns for the case of $N = 500$. In this section, to investigate the influence of the choice of priors on posterior, we apply our procedure to relatively small point patterns. Then we performed our Bayesian procedure of the two-parameter Soft-Core models for the simulated equilibrium point patterns generated by MCMC for the cases of $N = 100$, $V = \sqrt{100} \times \sqrt{100}$, $\alpha = 0.2, 0.3, 0.4$, $\tau = 0.05, 0.1, 0.3, 0.5$. These data sets are named as follows; for $\alpha = 0.2$ corresponding to τ : (DS0.2-0.05, DS0.2-0.1, DS0.2-0.3, DS0.2-0.5), for $\alpha = 0.3$: (DS0.3-0.05, DS0.3-0.1, DS0.3-0.3, DS0.3-0.5) and for $\alpha = 0.4$: (DS0.4-0.05, DS0.4-0.1, DS0.4-0.3, DS0.4-0.5), respectively. Figs. 6.16, 6.17 and 6.18 illustrate these simulated point patterns. We regard these point patterns as relatively small samples in comparison with the case of $N = 500$. From Figs. 6.16, 6.17 and 6.18, it is clear that as the reduced density τ is getting larger, the degree of regularity increases. This tendency is similar to the case of the large sample point patterns as remarked in §6.1.3. We consider that it will be interesting, through our procedure, to estimate the two-parameter Soft-Core models for the case of small sample point patterns.

By similar methods as described in §6.1.3, we applied two types of the prior again. To specify the type (i) prior densities of α and σ for all data sets, we put the values of hyperparameters $(a, b) = (0, 0.5)$ in Eq. (4.4). And we set the values of (c, d) in Eq. (4.5): $(c, d) = (0, 0.224)$ for $\tau = 0.05$; $(c, d) = (0, 0.316)$ for $\tau = 0.1$; $(c, d) = (0, 0.548)$ for $\tau = 0.3$; $(c, d) = (0, 0.707)$ for $\tau = 0.5$, respectively. For the type (ii) priors, we put the values of hyperparameters $(\mu_\alpha, s_\alpha) = (0.25, 0.125)$ and $(\mu_\sigma, s_\sigma) = (0.612, 0.306)$ in Eqs. (4.6) and (4.7) for all data sets, respectively.

For both types (i) and (ii) of prior, we applied again the three jumping densities: uniform, normal and independence sampler. For uniform and normal jumping densities, the adjusting parameters of Eqs. (5.5)-(5.8) were chosen such that Markov chain simulation should have the acceptance rate about $0.40 \sim 0.44$, mentioned in §5.2.2. Tables 6.17-6.20 provide the values of the adjusting parameters. For the independence samplers, we set the values of the hyperparameters $(a, b) = (0, 0.5)$ in Eq. (5.9). Then corresponding to the value of τ for each data set, we put the values of (c, d) in Eq. (5.10) as the same values of prior as described above.

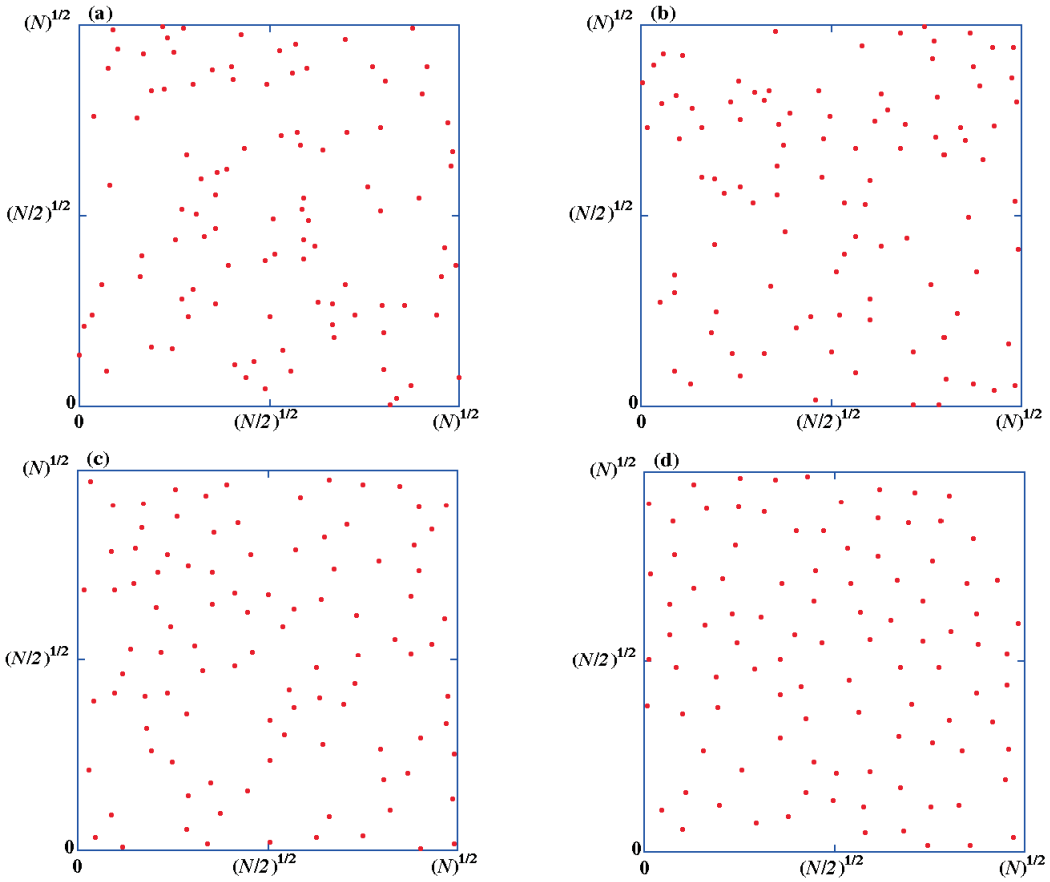

Figure 6.16: *Simulated equilibrium point patterns* ($N = 100, V = \sqrt{100} \times \sqrt{100}, \alpha = 0.2$). *(a):* $\tau = 0.05$ *(DS0.2-0.05); (b):* $\tau = 0.1$ *(DS0.2-0.1); (c):* $\tau = 0.3$ *(DS0.2-0.3) ; (d):* $\tau = 0.5$ *(DS0.2-0.5).*

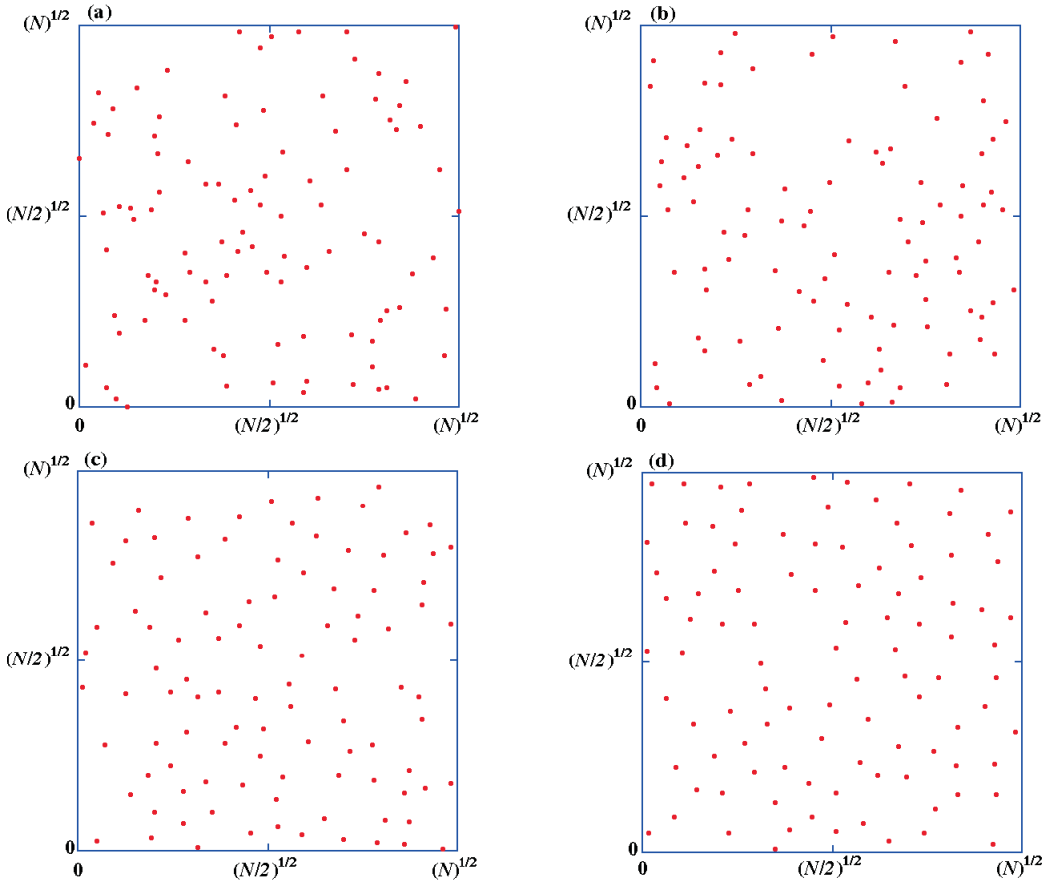

Figure 6.17: *Simulated equilibrium point patterns* ($N = 100, V = \sqrt{100} \times \sqrt{100}, \alpha = 0.3$). *(a):* $\tau = 0.05$ *(DS0.3-0.05); (b):* $\tau = 0.1$ *(DS0.3-0.1); (c):* $\tau = 0.3$ *(DS0.3-0.3) ; (d):* $\tau = 0.5$ *(DS0.3-0.5).*

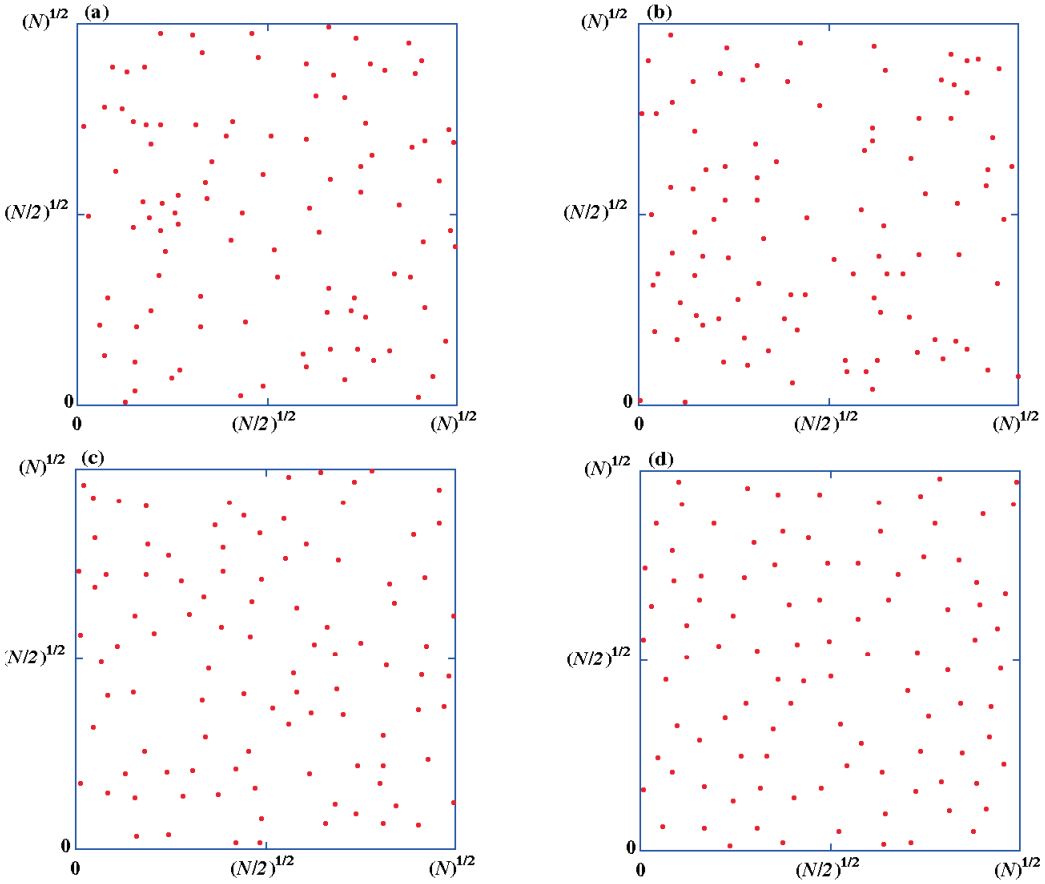

Figure 6.18: *Simulated equilibrium point patterns* ($N = 100, V = \sqrt{100} \times \sqrt{100}, \alpha = 0.4$). *(a):* $\tau = 0.05$ *(DS0.4-0.05); (b):* $\tau = 0.1$ *(DS0.4-0.1); (c):* $\tau = 0.3$ *(DS0.4-0.3) ; (d):* $\tau = 0.5$ *(DS0.4-0.5).*

Table 6.17: *Type (i) prior with uniform jumping specification: the values of the adjusting parameters* $(\delta_\alpha, \delta_\sigma)$ *in Eqs. (5.5)-(5.6) for the uniform jumpings.*

Data	$(\delta_\alpha, \delta_\sigma)$
$DS0.2-0.05$	(0.085, 0.080)
$DS0.2-0.1$	(0.060, 0.072)
$DS0.2-0.3$	(0.050, 0.060)
$DS0.2-0.5$	(0.020, 0.050)
$DS0.3-0.05$	(0.085, 0.075)
$DS0.3-0.1$	(0.090, 0.080)
$DS0.3-0.3$	(0.006, 0.040)
$DS0.3-0.5$	(0.003, 0.010)
$DS0.4-0.05$	(0.100, 0.095)
$DS0.4-0.1$	(0.080, 0.082)
$DS0.4-0.3$	(0.003, 0.040)
$DS0.4-0.5$	(0.003, 0.040)

Table 6.18: *Type (i) prior with normal jumping specification: the values of the adjusting parameters* $(\delta_\alpha^2, \delta_\sigma^2)$ *in Eqs. (5.7)-(5.8) for the normal jumpings.*

Data	$(\delta_\alpha^2, \delta_\sigma^2)$
$DS0.2-0.05$	(0.030, 0.060)
$DS0.2-0.1$	(0.013, 0.043)
$DS0.2-0.3$	(0.010, 0.040)
$DS0.2-0.5$	(0.007, 0.037)
$DS0.3-0.05$	(0.040, 0.040)
$DS0.3-0.1$	(0.065, 0.055)
$DS0.3-0.3$	(0.003, 0.050)
$DS0.3-0.5$	(0.002, 0.050)
$DS0.4-0.05$	(0.050, 0.040)
$DS0.4-0.1$	(0.040, 0.050)
$DS0.4-0.3$	(0.002, 0.035)
$DS0.4-0.5$	(0.002, 0.039)

Table 6.19: *Type (ii) prior with uniform jumping specification: the values of the adjusting parameters* $(\delta_\alpha, \delta_\sigma)$ *in Eqs. (5.5)-(5.6) for the uniform jumpings.*

Data	$(\delta_\alpha, \delta_\sigma)$
$DS0.2-0.05$	(0.075, 0.070)
$DS0.2-0.1$	(0.060, 0.070)
$DS0.2-0.3$	(0.053, 0.060)
$DS0.2-0.5$	(0.026, 0.053)
$DS0.3-0.05$	(0.090, 0.080)
$DS0.3-0.1$	(0.075, 0.075)
$DS0.3-0.3$	(0.025, 0.035)
$DS0.3-0.5$	(0.003, 0.010)
$DS0.4-0.05$	(0.080, 0.080)
$DS0.4-0.1$	(0.008, 0.080)
$DS0.4-0.3$	(0.004, 0.045)
$DS0.4-0.5$	(0.003, 0.044)

Table 6.20: *Type (ii) prior with normal jumping specification: the values of the adjusting parameters* $(\delta_\alpha^2, \delta_\sigma^2)$ *in Eqs. (5.7)-(5.8) for the normal jumpings.*

Data	$(\delta_\alpha^2, \delta_\sigma^2)$
$DS0.2-0.05$	(0.020, 0.050)
$DS0.2-0.1$	(0.018, 0.005)
$DS0.2-0.3$	(0.015, 0.045)
$DS0.2-0.5$	(0.013, 0.040)
$DS0.3-0.05$	(0.055, 0.050)
$DS0.3-0.1$	(0.050, 0.050)
$DS0.3-0.3$	(0.007, 0.005)
$DS0.3-0.5$	(0.005, 0.005)
$DS0.4-0.05$	(0.060, 0.050)
$DS0.4-0.1$	(0.050, 0.050)
$DS0.4-0.3$	(0.005, 0.030)
$DS0.4-0.5$	(0.002, 0.030)

Then, to estimate (α, τ) simultaneously, we performed a simulation of 60000 steps in the Metropolis-Hastings algorithm for all data sets. For the assessment of convergence, we evaluated the burn-in time T^* and the stopping time T for our single long run by using similar methods as described in §6.1.3. Then we estimated $(T^*, T) = (10000, 60000)$. Therefore, we have used 50000 samples for all data sets as samples from the joint posterior densities $p(\alpha, \tau \mid X)$. We will explain the assessment of convergence in §8.2.

Tables 6.21, 6.22 and 6.23 provide the values of simulated posterior mean of respective parameters under the both types (i) and (ii) of prior with three jumping densities for all data sets, respectively. In the last column of the Tables, the maximum likelihood estimates (MLE) of α and τ are also given. Then we can see that each posterior mean and each MLE of τ are nearly close in every data set DS0.2, DS0.3 and DS0.4.

On the other hand, it is clear that each posterior mean of α is different between type (i) and (ii) priors. From Table 6.21, we can see that each posterior mean of α is greater than each MLE. From Table 6.22, we can see that, for the (DS0.3-0.05, DS0.3-0.1), each posterior mean of α under the type (i) prior is greater than each MLE, and each posterior mean of α under the type (ii) prior is less than MLE. It can be seen from Table 6.23 that each posterior mean of α under the type (i) prior is larger than MLE for the DS0.4-0.05. Then we can see that each posterior mean of α under the type (ii) prior is less than MLE for all data sets DS0.4. It can be found from three Tables 6.21-6.23 that, for the data whose τ is large, the difference of the mean of α between the type (i) and (ii) prior tends to be small.

Table 6.21: *Posterior means of* (α, τ)*, indicated by '*$\overline{\alpha}$*' and '*$\overline{\tau}$*', under both types (i) and (ii) of prior with each jumping density for data sets (DS0.2-0.05, DS0.2-0.1, DS0.2-0.3, DS0.2-0.5); 'prior-jump' stands for combinations of the prior and jumping for respective parameters. In the last column, the maximum likelihood estimates of* (α, τ) *are also given.*

Data	*prior* − *jump*	$\overline{\alpha}$	$\overline{\tau}$	$(\hat{\alpha}, \hat{\tau})$
DS0.2-0.05	uni-uni	0.36721	0.059111	(0.21194, 0.062325)
	uni-nor	0.34042	0.059010	
	nor-uni	0.26890	0.060581	
	nor-nor	0.25841	0.060311	
	uni-ind	0.26085	0.059522	
	nor-ind	0.25996	0.060615	
DS0.2-0.1	uni-uni	0.33297	0.090118	(0.23717, 0.092476)
	uni-nor	0.25128	0.089021	
	nor-uni	0.25930	0.090943	
	nor-nor	0.25849	0.090722	
	uni-ind	0.25325	0.089728	
	nor-ind	0.26048	0.090992	
DS0.2-0.3	uni-uni	0.25335	0.28354	(0.21294, 0.27452)
	uni-nor	0.23845	0.28117	
	nor-uni	0.25529	0.28349	
	nor-nor	0.23847	0.28204	
	uni-ind	0.24699	0.28212	
	nor-ind	0.24323	0.28271	
DS0.2-0.5	uni-uni	0.22945	0.51986	(0.19923, 0.50421)
	uni-nor	0.24116	0.52692	
	nor-uni	0.23213	0.52098	
	nor-nor	0.23246	0.52130	
	uni-ind	0.23569	0.52401	
	nor-ind	0.23485	0.52120	

Table 6.22: *Posterior means of* (α, τ)*, indicated by '*$\overline{\alpha}$*' and '*$\overline{\tau}$*', under both types (i) and (ii) of prior with each jumping density for data sets (DS0.3-0.05, DS0.3-0.1, DS0.3-0.3, DS0.3-0.5); 'prior-jump' stands for combinations of the prior and jumping for respective parameters. In the last column, the maximum likelihood estimates of* (α, τ) *are also given.*

Data	$prior - jump$	$\overline{\alpha}$	$\overline{\tau}$	$(\hat{\alpha}, \hat{\tau})$
DS0.3-0.05	uni-uni	0.44074	0.043206	(0.31299, 0.046066)
	uni-nor	0.44429	0.043194	
	nor-uni	0.28373	0.043420	
	nor-nor	0.28631	0.043596	
	uni-ind	0.28251	0.042379	
	nor-ind	0.27138	0.043586	
DS0.3-0.1	uni-uni	0.43445	0.091305	(0.31047, 0.093013)
	uni-nor	0.42272	0.090864	
	nor-uni	0.29818	0.090318	
	nor-nor	0.30210	0.090407	
	uni-ind	0.31662	0.088631	
	nor-ind	0.29224	0.089812	
DS0.3-0.3	uni-uni	0.29559	0.29958	(0.29022, 0.32114)
	uni-nor	0.22323	0.30569	
	nor-uni	0.29305	0.32089	
	nor-nor	0.28140	0.31789	
	uni-ind	0.31176	0.32377	
	nor-ind	0.29034	0.32002	
DS0.3-0.5	uni-uni	0.30967	0.51393	(0.29977, 0.51055)
	uni-nor	0.35856	0.54476	
	nor-uni	0.29157	0.50432	
	nor-nor	0.31131	0.51556	
	uni-ind	0.33432	0.53194	
	nor-ind	0.30257	0.51169	

Table 6.23: *Posterior means of* (α, τ)*, indicated by '*$\overline{\alpha}$*' and '*$\overline{\tau}$*', under both types (i) and (ii) of prior with each jumping density for data sets (DS0.4-0.05, DS0.4-0.1, DS0.4-0.3, DS0.4-0.5); 'prior-jump' stands for combinations of the prior and jumping for respective parameters. In the last column, the maximum likelihood estimates of* (α, τ) *are also given.*

Data	*prior − jump*	$\overline{\alpha}$	$\overline{\tau}$	$(\hat{\alpha}, \hat{\tau})$
DS0.4-0.05	uni-uni	0.48746	0.052066	(0.37988, 0.054667)
	uni-nor	0.48283	0.052017	
	nor-uni	0.31259	0.051132	
	nor-nor	0.31439	0.051016	
	uni-ind	0.32413	0.050137	
	nor-ind	0.29962	0.050839	
DS0.4-0.1	uni-uni	0.39569	0.12415	(0.40367, 0.12808)
	uni-nor	0.35535	0.12232	
	nor-uni	0.33579	0.12170	
	nor-nor	0.35353	0.12313	
	uni-ind	0.37854	0.12311	
	nor-ind	0.34026	0.12192	
DS0.4-0.3	uni-uni	0.37713	0.26653	(0.39491, 0.27248)
	uni-nor	0.34305	0.26061	
	nor-uni	0.35757	0.26325	
	nor-nor	0.35620	0.26364	
	uni-ind	0.38813	0.26939	
	nor-ind	0.35378	0.26293	
DS0.4-0.5	uni-uni	0.35272	0.47149	(0.37883, 0.51718)
	uni-nor	0.38443	0.51068	
	nor-uni	0.35494	0.47546	
	nor-nor	0.34858	0.49819	
	uni-ind	0.39828	0.53138	
	nor-ind	0.36674	0.50888	

The comparison of the posterior densities of α and τ under both types (i) and (ii) of prior with three jumping densities are plotted in Figs. 6.19, 6.20 and 6.21 for all data sets, respectively. From Figs. 6.19-6.21(b), (d), (f) and (h), we can see that, for all data sets, the coincidence of the shape of the marginal posterior densities of τ between the type (i) and (ii) priors is good for four kinds of combinations of the two priors and the uniform and normal jumpings. On the other hand, for the parameter α, the results of posterior are quite different. From Figs. 6.19-6.21(e) and (g), it seems that, for the cases of $\tau = 0.3, 0.5$, the marginal posteriors of α under both types (i) and (ii) of prior are said to be similar in all data sets. On the contrary, for the cases of relatively small reduced density $\tau = 0.05, 0.1$, the marginal posteriors of α between type (i) and (ii) priors show different spread regardless of the jumping density (see Figs. 6.19-6.21(a) and (c)). From the results, it was found that priors have strong influence on the posterior for small sample point patterns whose reduced density τ is relatively small.

In this section, we have verified our Bayesian procedure for the two-parameter Soft-Core models for various point patterns for the case of $N = 100$. In the next section, we will discuss our procedure applied to various point patterns.

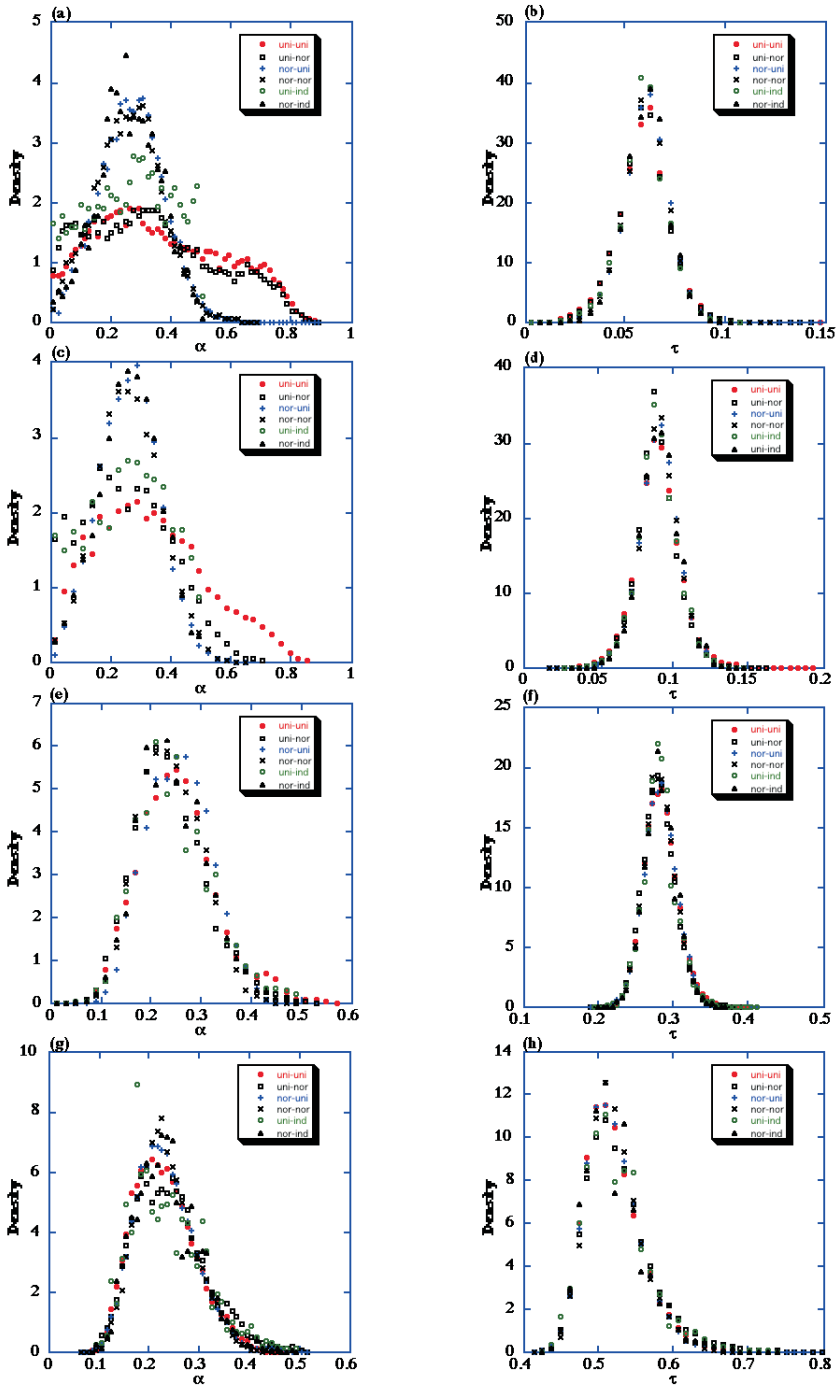

Figure 6.19: *Comparison of marginal posterior densities of α and τ under the type (i) prior with uniform jumping (uni-uni:●), normal jumping (uni-nor:□) and independence sampler (uni-ind:○), respectively, and under the type (ii) prior with uniform jumping (nor-uni:+), normal jumping (nor-nor:×) and independence sampler (nor-ind:△), respectively. (a),(b): DS0.2-0.05; (c),(d): DS0.2-0.1; (e),(f): DS0.2-0.3; (g),(h): DS0.2-0.5.*

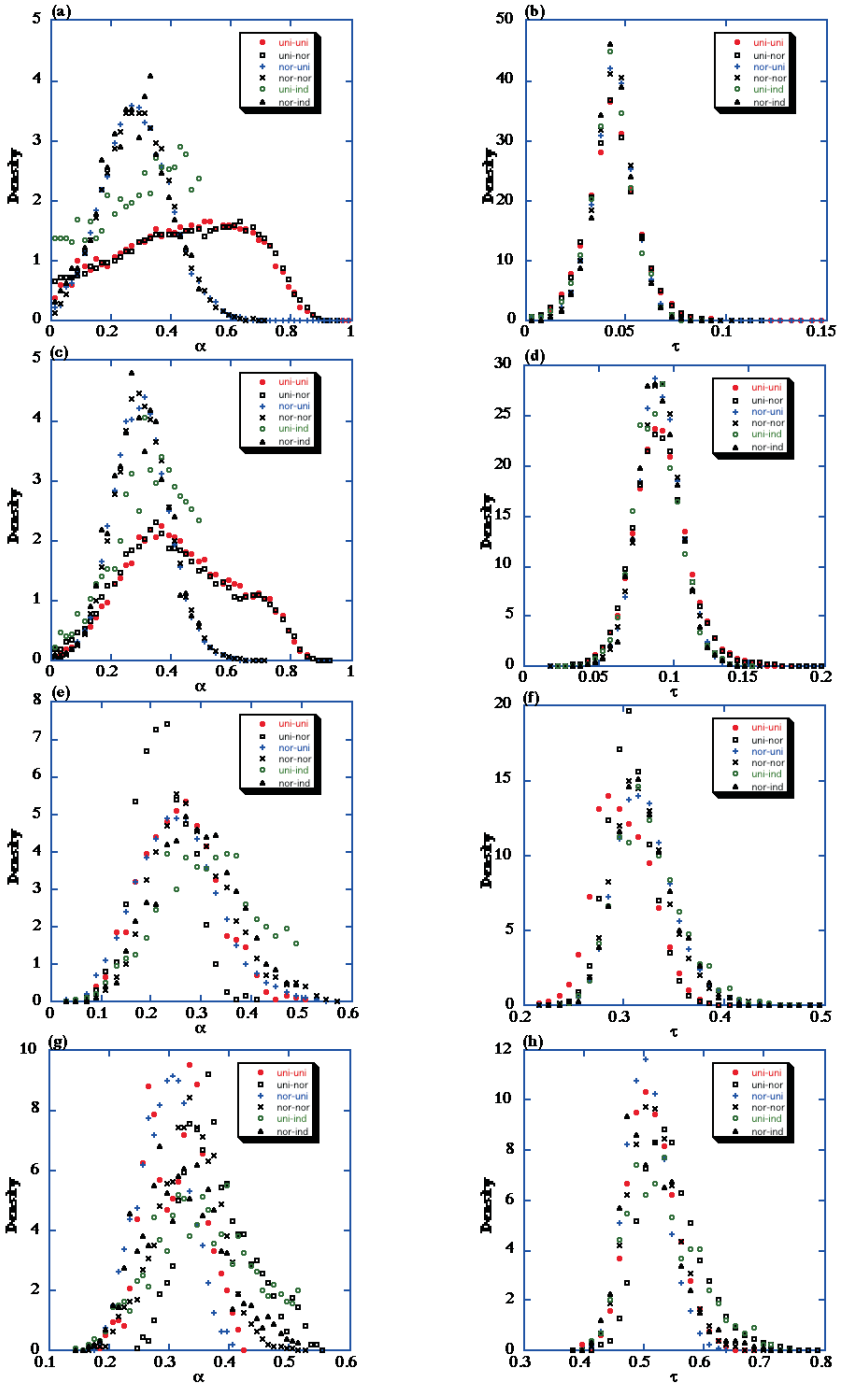

Figure 6.20: *Comparison of marginal posterior densities of α and τ under the type (i) prior with uniform jumping (uni-uni:●), normal jumping (uni-nor:□) and independence sampler (uni-ind:○), respectively, and under the type (ii) prior with uniform jumping (nor-uni:+), normal jumping (nor-nor:×) and independence sampler (nor-ind:△), respectively. (a),(b): DS0.3-0.05; (c),(d): DS0.3-0.1; (e),(f): DS0.3-0.3; (g),(h): DS0.3-0.5.*

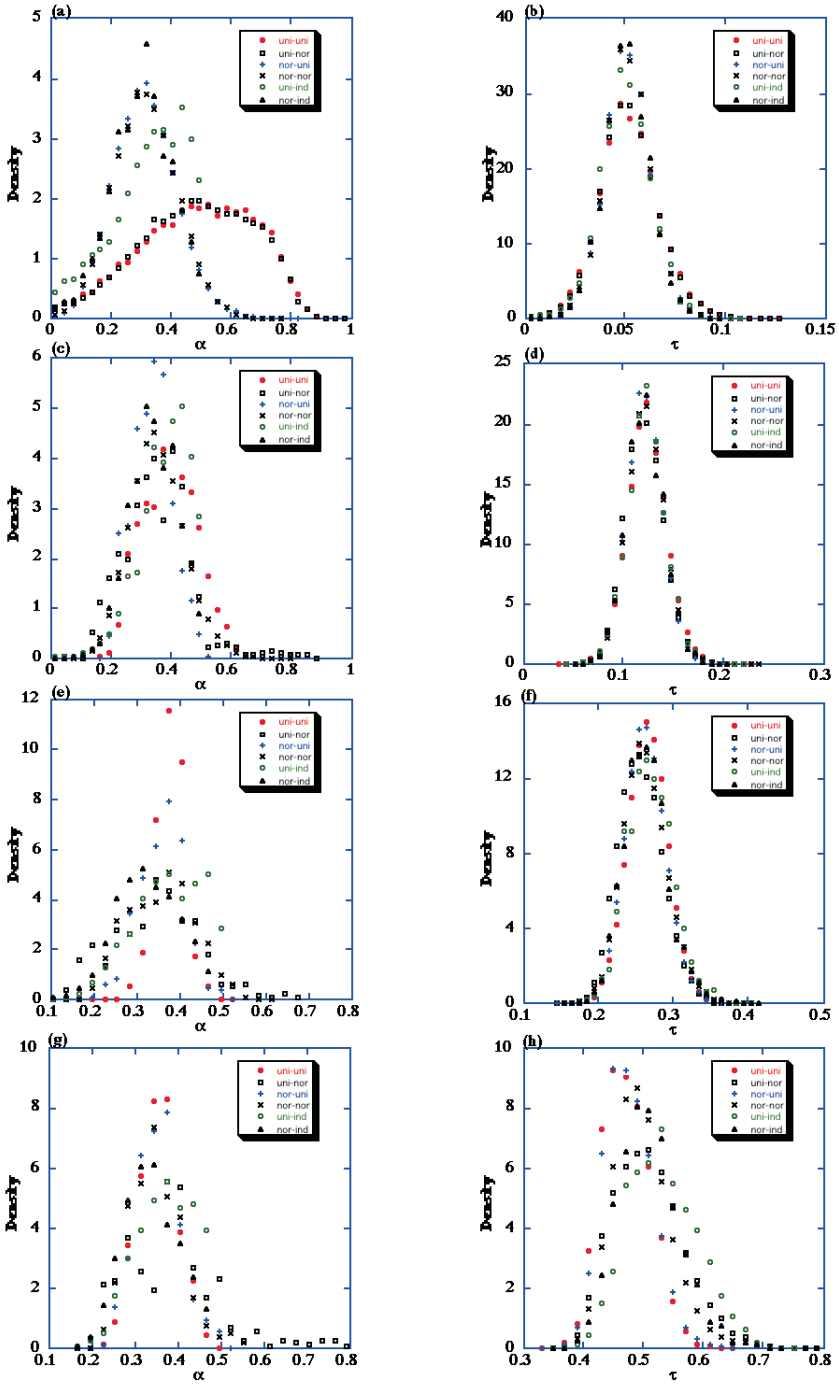

Figure 6.21: *Comparison of marginal posterior densities of α and τ under the type (i) prior with uniform jumping (uni-uni:●), normal jumping (uni-nor:□) and independence sampler (uni-ind:○), respectively, and under the type (ii) prior with uniform jumping (nor-uni:+), normal jumping (nor-nor:×) and independence sampler (nor-ind:△), respectively. (a),(b): DS0.4-0.05; (c),(d): DS0.4-0.1; (e),(f): DS0.4-0.3; (g),(h): DS0.4-0.5.*

6.3 Discussions

In this chapter, we have presented our Bayesian procedure for the two-parameter Soft-Core interaction potential models for the simulated equilibrium point patterns generated by MCMC for the cases of large and relatively small sample point patterns. Then our procedure is proved to be applicable to a wide class of various regular point patterns.

When large sample point patterns ($N = 500$) are given, from Figs 6.13-6.15, the coincidence between the shape of each marginal posterior density of α and τ is good regardless of combinations of the prior and jumping densities. Then it seems that the influence of the choice of prior densities upon posterior density is not too strong.

For relatively small sample point patterns ($N = 100$), from Figs 6.19-6.21, the coincidence between the shape of marginal posterior densities of τ is said to be good for all combinations of the prior and jumping. On the other hand, for the parameter α, the behavior of posterior are different. For the cases of relatively large reduced density (*i.e.* the degree of regularity is relatively high), the marginal posteriors of α under both types (i) and (ii) of prior with uniform and normal jumping densities seem to be similar in all data sets. On the contrary, for the cases of relatively small reduced density (*i.e.* the degree of regularity is relatively low), the marginal posteriors of α between the two priors show different spread regardless of the jumping densities. When the number of points N is small, Bayesian inferences are more sensitive to choice of prior density than inferences with large samples (Little and Rubin (2002)). Then it will be interesting to apply our Bayesian procedure to small sample real data. These results will be discussed in §8.4 and 8.6.

We here refer to the case of the application of the independence samplers. In each simulation, convergence of respective parameters seems to be good, but the acceptance rate were below 0.1 for all cases. Since the samplers have no adjusting parameters, these values were very low. Since the optimal jumping rule has acceptance rate about $0.40 \sim 0.44$ in two-dimension, as stated in §5.2.2, the use of the independence samplers is not appropriate for our Bayesian inference.

Chapter 7

Real data

7.1 Preliminary analysis

As the illustrative examples, we applied our method to four observed data, namely *charged steel balls* (Ogata and Tanemura (1989), Mase *et al.* (1994)), two *blue cones in a macaque retina* (Shapiro *et al.* (1985)), and *nesting pattern of the Gray Gulls* (Howell *et al.* (1974)), which are illustrated in Fig. 7.1, Figs. 7.2(a)-(b) and Fig. 7.3, respectively. Configurations of these data all show regular point pattern. Intuitively speaking, the degree of regularity of the first three data seems to be higher than the data of Gray Gulls. Table 7.1 gives the number of individuals N and the rectangular region V for each data set.

Table 7.1: *The number of individuals N and the rectangular region V for each data set.*

	Balls	P6T13	M6T10	Gulls
N	271	398	427	110
V	$2.81 \times 2.79[cm^2]$	$125 \times 85[mm^2]$	$125 \times 85[mm^2]$	$100 \times 100[m^2]$

To categorize the type of distribution of our point patterns as a preliminary analysis (Cressie (1993)), we obtained the following two *indices of clumping*; the *Morisita's index I_δ* based on quadrat counts (Morisita (1959)) and the *Hopkins-Skellam index A* based on nearest-neighbor distance measures (Hopkins and Skellam (1954)). The Morisita's index I_δ is expressed by

$$I_\delta = q \sum_{i=1}^{q} c_i(c_i - 1)/N(N-1), \tag{7.1}$$

where c_i is the number of individuals in the ith quadrat and q is the number of contiguous quadrats which divide the rectangular region V. In the special case when the spatial point pattern is a homogeneous Poisson point process, I_δ is equal to 1. And if the point pattern is a clustered type, I_δ is greater than 1, else if the point pattern is a regular type, I_δ is less than 1. On the other hand, the Hopkins-Skellam index A is represented as

$$A = \sum_{j=1}^{M} r_{1j}^2 \Big/ \sum_{j=1}^{M} r_{2j}^2, \tag{7.2}$$

where r_{1j} is the distance between randomly sampled point and its nearest individual, r_{2j} is the distance between a randomly chosen individual and its nearest individual in the rectangular V, and where M is the total number of samples. When the point pattern is considered as a Poisson pattern, A is equal to 1. If the point pattern is a clustered type, or a regular type, then A is greater than 1, or less than 1, respectively. Table 7.1 gives the values of the Morisita's index I_δ and the Hopkins-Skellam index A for all data. In the first column of Table 7.1, $I_\delta(i \times i)$ represents the value of Eq. (7.1) for $q = i \times i$ grid of quadrats, $\chi_0^2 = I_\delta(N-1) + q - N$ represents χ^2 test statistic with $(q-1)$ degrees of freedom for the Poisson null hypothesis and $Pr\{X \leq \chi^2_{0(q-1)}\}$ is the probability below the $\chi^2_{0(q-1)}$ point. The values of $Pr\{X \leq \chi^2_{0(q-1)}\}$ are calculated numerically. Each χ^2 statistic obtained for suitably large q leads to the rejection of the null hypothesis of the Poisson model in the two-sided test for the significance level 0.05/2 for all data sets. In the last row of Table 7.1, the values of A are the mean of independent 1000 trial values of A in the case of $M = 1000$ in Eq. (7.2), and Φ_0 represents test statistic for the Poisson model based on the distance methods, where $\Phi_0 = 2\{A/(1+A) - 1/2\}\sqrt{2N+1}$ approximately obeys a standard normal distribution for not too small N. We have used a periodic boundary condition, that is to say, the rectangular region V is regarded as a torus, in calculating the Hopkins-Skellam index A (Ripley (1977, 2004), see §8.5). From Table 7.1, the null hypothesis of Poisson model is rejected in cases of both quadrat methods and distance methods. All data sets appear to be regular from the values of test statistics.

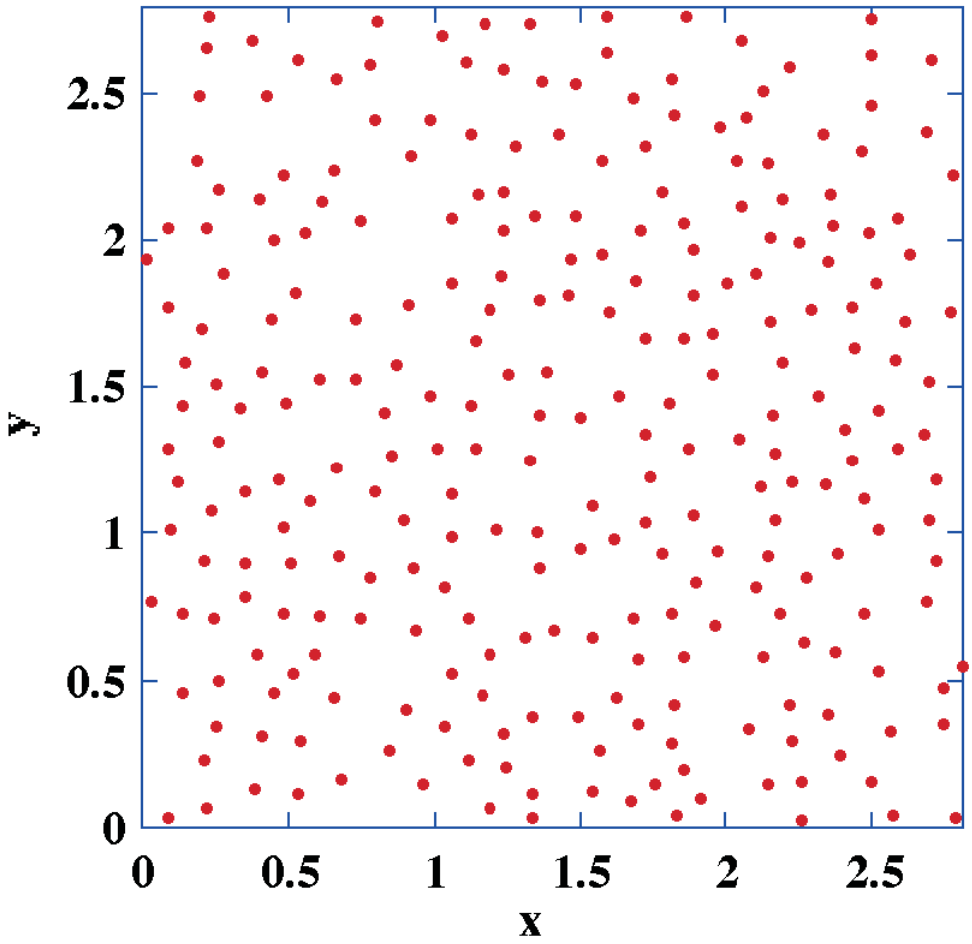

Figure 7.1: *Map of the charged steel balls (abbreviated to Balls) ($N = 271, V = 2.81 \times 2.79[cm^2]$).*

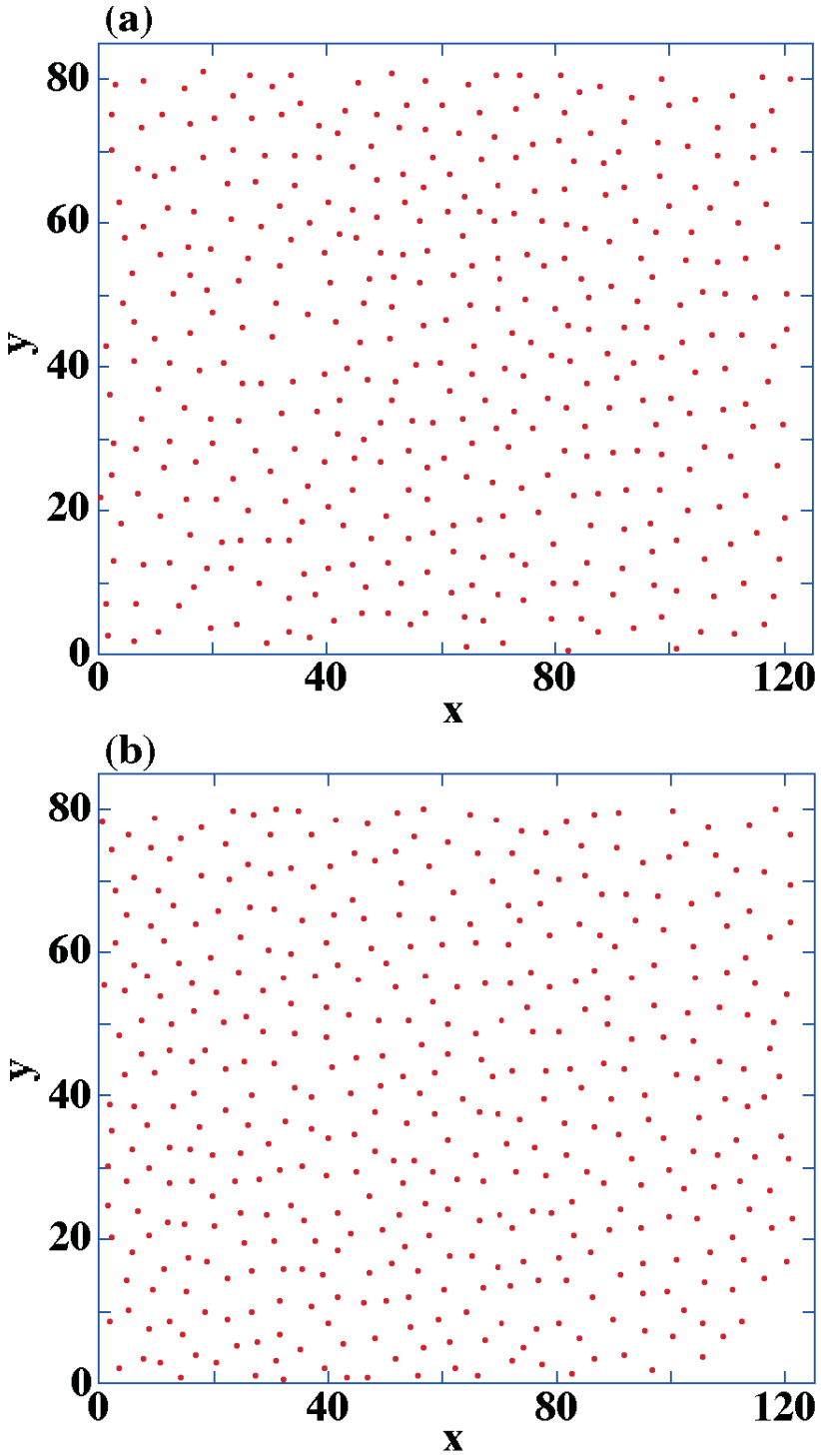

Figure 7.2: *Two point patterns of blue cones of macaque retina in the region* $V = 125 \times 85$ $[mm^2]$ *in the scale of the photomicrograph. (a) P6T13 (*$N = 398$*; 6 degrees above the horizontal meridian and 13 degrees to the right of the vertical); (b) M6T10 (*$N = 427$*; 6 degrees below and 10 degrees to the right).*

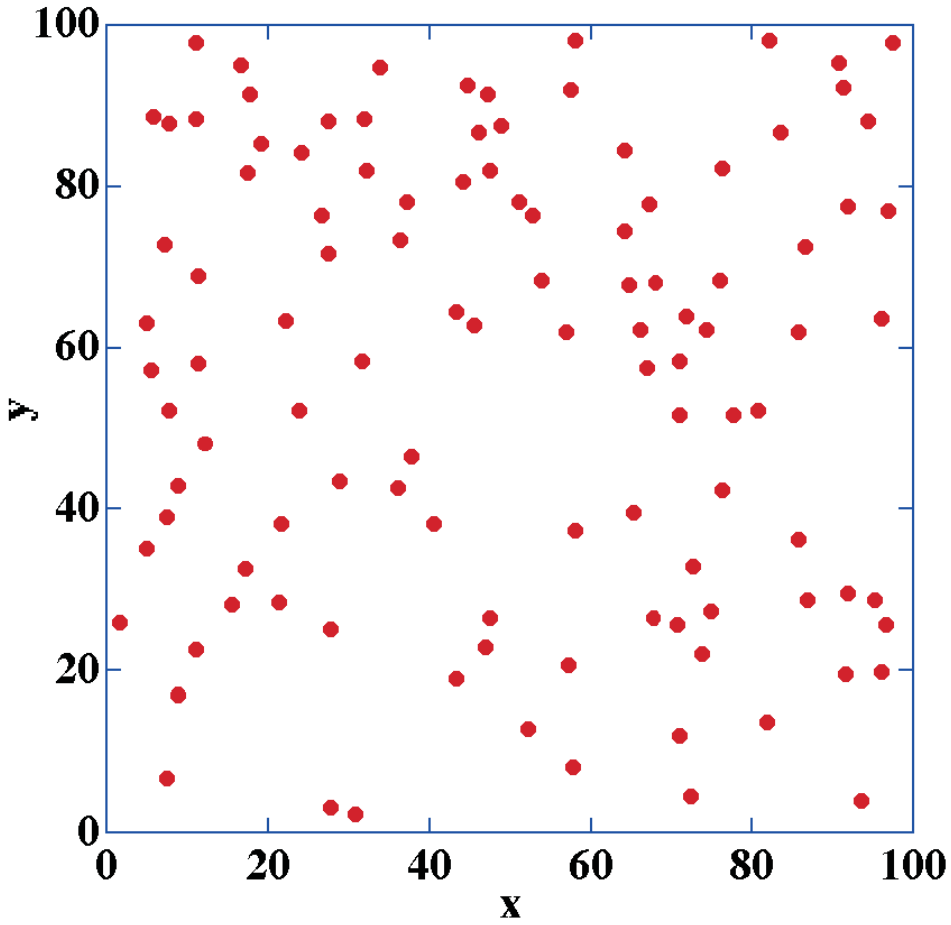

Figure 7.3: *The map of nests of Gray Gulls, Larus modestus (abbreviated to Gulls)* ($N = 110,\ V = 100 \times 100\ [m^2]$).

Table 7.2: *Results of the values of the Morisita's index I_δ of Eq. (7.1) and the Hopkins-Skellam index A of Eq. (7.2), and their tests of the Poisson model based on the quadrat and distance methods.*

	Balls	P6T13	M6T10	Gulls
$I_\delta(4\times 4), \chi^2_0$	$0.962, 4.74$	$0.981, 7.46$	$0.997, 13.72$	$0.993, 14.24$
$Pr\{X \le \chi^2_{0(4\times 4-1)}\}$	0.00589	0.0564	0.453	0.493
$I_\delta(8\times 8), \chi^2_0$	$0.855, 23.85$	$0.897, 22.11$	$0.922, 29.77$	$0.907, 52.86$
$Pr\{X \le \chi^2_{0(8\times 8-1)}\}$	1.81×10^{-6}	3.82×10^{-7}	1.17×10^{-4}	0.185
$I_\delta(12\times 12), \chi^2_0$	$0.665, 49.85$	$0.778, 54.87$	$0.800, 57.80$	$0.889, 130.90$
$Pr\{X \le \chi^2_{0(12\times 12-1)}\}$	~ 0	~ 0	~ 0	0.243
$I_\delta(16\times 16), \chi^2_0$	$0.497, 119.19$	$0.578, 87.47$	$0.630, 97.38$	$0.641, 215.87$
$Pr\{X \le \chi^2_{0(16\times 16-1)}\}$	~ 0	~ 0	~ 0	0.0359
$I_\delta(20\times 20), \chi^2_0$	$0.284, 205.68$	$0.435, 174.70$	$0.449, 164.27$	$0.400, 333.60$
$Pr\{X \le \chi^2_{0(20\times 20-1)}\}$	~ 0	~ 0	~ 0	0.00752
A, Φ_0	$0.362, -10.92$	$0.313, -14.77$	$0.334, -14.60$	$0.792, -1.73$
$Pr\{X \le \Phi_0\}$	~ 0	~ 0	~ 0	0.0422

7.2 Charged steel balls

The mapped data of charged steel balls (abbreviated to Balls) is shown in Fig. 7.1, and it was obtained by the following way (Ogata and Tanemura (1989), Mase *et al.* (1994)). Many steel balls of diameter 0.5 mm were put in a transparent plastic box. When the box was shaken violently by hand, all particles charged up with electricity and are supposed to form a regular point pattern. The values of the Morisita's index I_δ and the Hopkins-Skellam index A, in Table 7.1, indicate clearly that this data is classified as a regular point pattern. The interaction between particles may not be described simply by the Coulomb repulsive force, because of the existence of the wall of the plastic box, which is also charged with electricity. So it will be interesting to estimate a repulsive Soft-Core interaction potential between steel balls by our procedure.

7.3 Blue cones in a macaque retina

Figs. 7.2(a)-(b) exhibit the two spatial point patterns of blue cones in a macaque retina (abbreviated to P6T13 and M6T10, respectively). The retina of the primates consists of two photoreceptors of rod and cone cells. The cones of the retina in primates are classified into three types: blue-sensitive, green-sensitive and red-sensitive. The two data sets, P6T13 and M6T10, were collected from different areas of the retina at about the same distance, 13 degrees, from the fovea (the center of the retina). The data in Figs. 7.2(a)-(b) were read from the photomicrographs in Shapiro *et al.* (1985) (see Ogata and Tanemura (1989)). Therefore the scale of length in the data does not represent that of real objects. The blue cones (blue-sensitive) represent a comparatively small fraction of the total cone population, and their configuration forms a regular point pattern of loosely packed cones among other cone types. From the values of two indices of clumping, I_δ and A, in Table 7.1, both P6T13 and M6T10 are classified as regular point patterns. So the blue cones are considered to distribute evenly in the fovea to work for photoreceptors, and it will be interesting to analyze the pattern of blue cones by our procedure.

7.4 Nesting pattern of the Gray Gulls

Fig. 7.3 shows a nesting pattern of the Gray Gulls, *Larus modestus* (abbreviated to Gulls), near the Pacific coast of South America. It is reported that the habitat of these gulls has no plants and is rather uniform in this region V. In Ogata and Tanemura (1989), the data was analyzed, and a small reduced density was obtained. In our preliminary analysis, I_δ and A were obtained in Table 7.1, too. The values of the Morisita's index I_δ and Hopkins-Skellam index A indicate that this data is interpreted as a regular type in short range. Because this data shows a different character from other three data in long range, it will be interesting to analyze it by our procedure.

Chapter 8

Application to the real data

8.1 Performing MCMC experiments

In Chapter 6, we have verified our Bayesian procedure for the two-parameter Soft-Core potential models to various simulated regular point patterns. Then, in this chapter, we would like to apply our procedure to four real data sets explained in Chapter 7. To perform Bayesian estimation of the two-parameter Soft-Core models for each of our data sets, we considered two types of prior density (with jumping density), as described in §4.2: (i) uniform prior densities (with uniform jumping densities) (type (i) prior) and (ii) normal prior densities (with normal jumping densities) (type (ii) prior). Here, we make reference to the literature of Okabe and Tanemura (2006).

To specify prior densities, we can set the values of hyperparameters (a, b, c, d) in Eqs. (4.4)-(4.5) for the type (i) prior densities. Since the log-likelihood in Eq. (3.14) is effective in the range $0 < \alpha \leq 0.5$ and $0 < \tau \leq 0.75$ as stated in §3.2, and since σ is related to τ through Eq. (3.4), these values are given in Table 8.1 for all data sets. For the type (ii) priors, the hyperparameters of normal densities are specified as follows. In Eqs. (4.6)-(4.7), their means (μ_α, μ_σ) are chosen as the center of the range of each parameter, and their standard deviations (s_α, s_σ) for α and σ are respectively chosen, through trial and error, as the values of half of the mean (see Table 8.1). These normal priors are supposed to cover the above effective range.

For jumping densities for both types (i) and (ii) of prior, their adjusting parameters are chosen such that Markov chain simulation should have the acceptance rate around $0.40 \sim 0.44$ of the Metropolis-Hastings steps, as described in §5.2.2. For the normal jumping densities in Eqs. (5.7)-(5.8), the method of generating random values from a truncated normal distribution $\mathrm{TN}_{[0,\infty)}(\cdot \mid \theta^{t-1}, \delta_\theta^2)$ proceeds as follows. Given the current point θ^{t-1}, sample a proposal point θ^* from a normal density $\mathrm{N}(\theta^{t-1}, \delta_\theta^2)$ recurrently until θ^* is sampled in the range $[0, \infty)$ (*e.g.* Geweke (1991)), as stated in §5.2.1. Table 8.2 shows the values

of the optimal adjusting parameters $(\delta_\alpha, \delta_\sigma)$ in Eqs. (5.5)-(5.6) of the uniform jumping densities for the type (i) prior, and $(\delta_\alpha^2, \delta_\sigma^2)$ in Eqs. (5.7)-(5.8) of the normal jumping densities for the type (ii) prior.

Then we performed a single long run simulation of 26000 steps in the Metropolis-Hastings algorithm for three data sets (Balls, P6T13, M6T10), and of 60000 steps for Gulls data. These simulations were performed for both types (i) and (ii) of prior, and the periodic boundary condition was used for calculating the total potential energy in Eq. (3.14), as described in §6.1.1 (see §8.5). The above choice of the number of steps will be discussed later. For comparison, the case of the independence samplers will be also examined in §8.2.4.

Table 8.1: *Prior specification: the values of hyperparameters* (a, b, c, d) *in Eqs. (4.4)-(4.5) for the type (i) prior, and the means* (μ_α, μ_σ) *and the variances* (s_α^2, s_σ^2) *in Eqs. (4.6)-(4.7) for the type (ii) prior. These hyperparameters* (a, b, c, d) *are also used in the case of the independence sampler as described in §8.2.3.*

Data	(a, b, c, d)	(μ_α, μ_σ, s_α^2, s_σ^2)
Balls	(0, 0.50, 0, 0.147)	(0.25, 0.0736, 0.125^2, 0.0368^2)
P6T13	(0, 0.50, 0, 4.48)	(0.25, 2.24, 0.125^2, 1.12^2)
M6T10	(0, 0.50, 0, 4.32)	(0.25, 2.16, 0.125^2, 1.08^2)
Gulls	(0, 0.50, 0, 8.26)	(0.25, 4.13, 0.125^2, 2.06^2)

Table 8.2: *Jumping specification: the values of the adjusting parameters* $(\delta_\alpha, \delta_\sigma)$ *in Eqs. (5.5)-(5.6) for the type (i) prior, and* $(\delta_\alpha^2, \delta_\sigma^2)$ *in Eqs. (5.7)-(5.8) for the type (ii) prior.*

Data	(δ_α, δ_σ)	(δ_α^2, δ_σ^2)
Balls	(0.035, 0.008)	(0.017^2, 0.0055^2)
P6T13	(0.035, 0.20)	(0.018^2, 0.13^2)
M6T10	(0.035, 0.15)	(0.012^2, 0.12^2)
Gulls	(0.058, 0.68)	(0.031^2, 0.49^2)

8.2 MCMC convergence

8.2.1 Assessing stopping time of the Metropolis-Hastings algorithm

In this thesis, single long run approach is adopted, as described in §5.3.1. To assess the convergence of the Metropolis-Hastings algorithm, we used the diagnostics quantity (potential scale reduction factor) $\hat{R}$ of Gelman and Rubin's method as described in §5.3.1. To estimate the stopping time T of a single long run, we performed $k = 5$ independent simulations whose number of iterations is $(l^* + l)$ steps each (l^*: the length of the sequence discarded as burn-in (the first part); l: the length of the sequence after burn-in (the second part)) for all data sets Balls, P6T13, M6T10 and Gulls. Then we fixed the length (l^*, l) to be $(1000, 5000)$ of each simulation for three data sets (Balls, P6T13, M6T10). For the data set Gulls, we put (l^*, l) to be $(10000, 10000)$. Starting point $(\alpha^0_{(1)}, \sigma^0_{(1)})$ of the first sequence among five simulations was chosen as the mean of the Monte Carlo samples generated by a test long run of 30000 steps for each data. And starting points $(\alpha^0_{(i)}, \sigma^0_{(i)})$ $(i = 2, 3, 4, 5)$ of the other four simulations were independently drawn from the normal densities whose mean is $(\alpha^0_{(1)}, \sigma^0_{(1)})$ and whose standard deviation is the half of the means. For each data set, we calculated the potential scale reduction factor $\hat{R}$ in Eq. (5.13) for α and σ separately. Table 8.3 shows the values of $\hat{R}$ for respective parameters for all data sets. In this table, the last column represents the values of $\hat{R}$ in the case of prescribed $k = 5$ parallel sequences. Since each value of $\hat{R}$ was well below 1.1 for both parameters, as the suitable set of the values (l^*, l), we evaluated $(l^*, l) = (1000, 5000)$ for Balls, P6T13 and M6T10, and $(l^*, l) = (10000, 10000)$ for Gulls. We will discuss the burn-in time $T^*(= l^*)$ in detail in §8.2.3.

Thus, we estimated $T^*(= l^*) = 1000$ and $T = (5 \times 5000) + T^* = 26000$ for (Balls, P6T13, M6T10), and $T^* = 10000$ and $T = (5 \times 10000) + T^* = 60000$ for Gulls. Then, we performed the MCMC simulations of the stopping time T for each data.

To monitor the convergence graphically, time series plots are given in Figs. 8.1(a)-(d) for $k = 5$ independent sequences of respective parameters for Balls under the type (i) prior. Figs. 8.1(a) and (c) show initial $l^* = 1000$ steps (first part) of five independent sequences for α and σ, respectively, and Figs. 8.1(b) and (d) show correspondingly the subsequent $l = 5000$ steps (second parts) where five independent sequences are overlapping (*e.g.* Cowles and Carlin (1996), Bray and Wright (1998)). From these figures, we can see the behavior of five sequences for the case of Balls. We see here that a convergence had been reached by $l^* = 1000$ steps. Figs. 8.2(a)-(d) show the case of the type (ii) prior. The results of $k = 5$ independent simulations for P6T13 and M6T10 were similar to the case of Balls and their time series plots are given in Figs.

8.3-8.6 for $k = 5$ independent sequences of respective parameters for P6T13 and M6T10.

On the other hand, for the time series plots of $k = 5$ independent sequences by fixing $(l^*, l) = (10000, 10000)$ for Gulls, the convergence had been achieved by $l^* = 10000$ steps, and their subsequent $l = 10000$ steps sufficiently converged for respective parameters. These behaviors are given in Figs. 8.7-8.8. As described above, in the case of $(l^*, l) = (1000, 5000)$ for Gulls under the type (ii) prior, the values of $\hat{R}_\alpha$ and $\hat{R}_\sigma$ are respectively less than 1.1. We here put the convergence time to be 10000 steps for both types (i) and (ii) of prior.

Table 8.3: *Results of the values of the potential scale reduction factor $\hat{R}$ as the monitoring convergence diagnostics in the case of k=5 parallel sequences for all data sets; 'param.' and 'type' stand for the kind of parameter and the type of prior, respectively.*

Data	*param.*	*type*	$\hat{R}$ (5 *sequences*)
Balls	α	(i)	1.0033
		(ii)	1.0028
	σ	(i)	1.0016
		(ii)	1.0012
P6T13	α	(i)	1.0016
		(ii)	1.0026
	σ	(i)	1.0011
		(ii)	1.0013
M6T10	α	(i)	1.0010
		(ii)	1.0026
	σ	(i)	1.0007
		(ii)	1.0023
Gulls	α	(i)	1.0056
		(ii)	1.0032
	σ	(i)	1.0001
		(ii)	1.0008

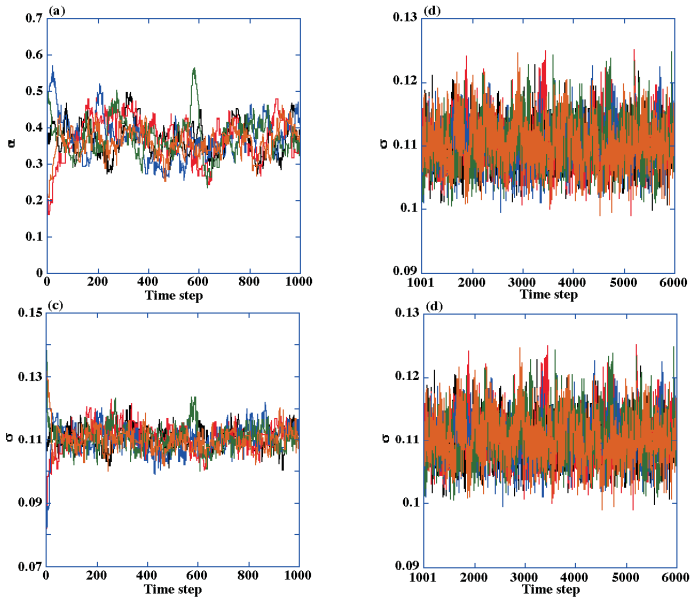

Figure 8.1: *Convergence monitoring plots of the parameters α [(a),(b)] and σ [(c),(d)] for Balls under the type (i) prior. (a),(c): First part of five sequences (t=0-1000); (b),(d): Second part of five sequences (t=1001-6000).*

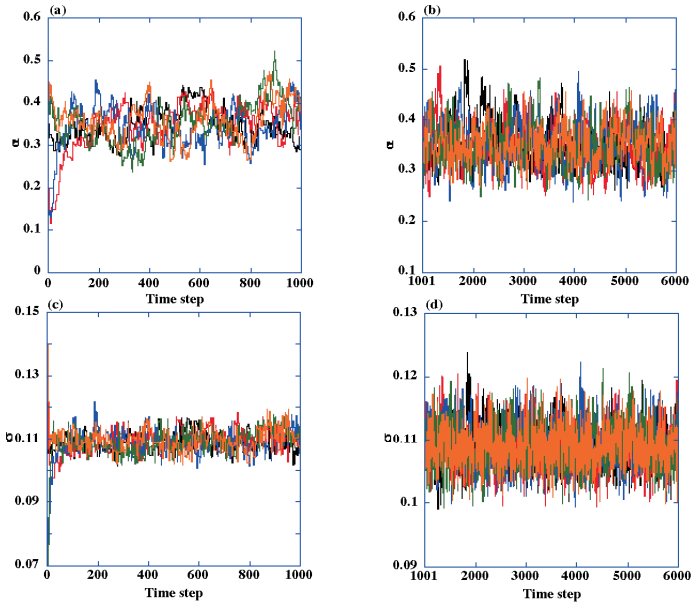

Figure 8.2: *Convergence monitoring plots of the parameters α [(a),(b)] and σ [(c),(d)] for Balls under the type (ii) prior. (a),(c): First part of five sequences (t=0-1000); (b),(d): Second part of five sequences (t=1001-6000).*

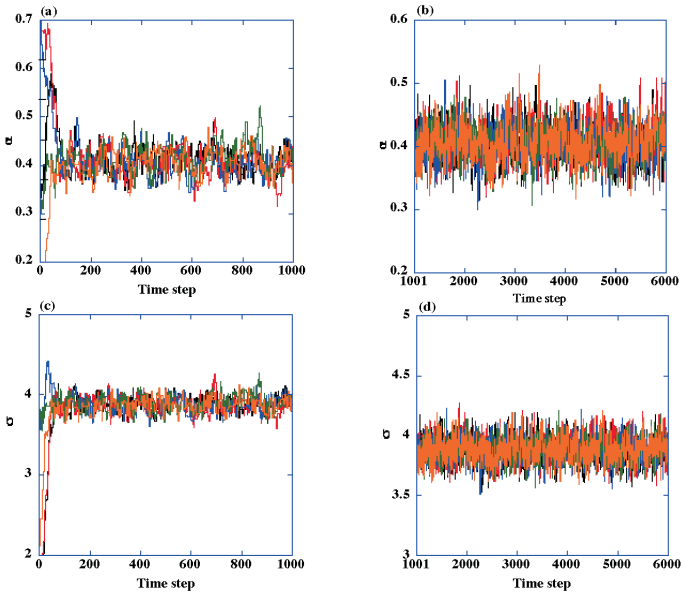

Figure 8.3: *Convergence monitoring plots of the parameters α [(a),(b)] and σ [(c),(d)] for P6T13 under the type (i) prior. (a),(c): First part of five sequences (t=0-1000); (b),(d): Second part of five sequences (t=1001-6000).*

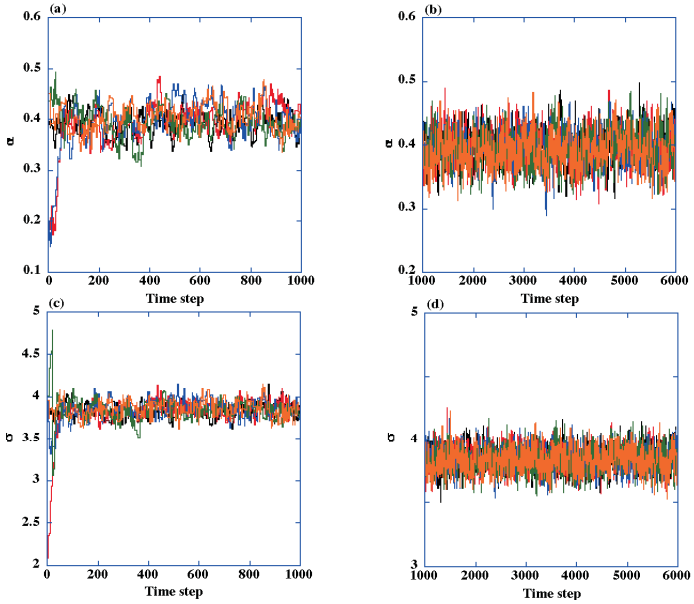

Figure 8.4: *Convergence monitoring plots of the parameters α [(a),(b)] and σ [(c),(d)] for P6T13 under the type (ii) prior. (a),(c): First part of five sequences (t=0-1000); (b),(d): Second part of five sequences (t=1001-6000).*

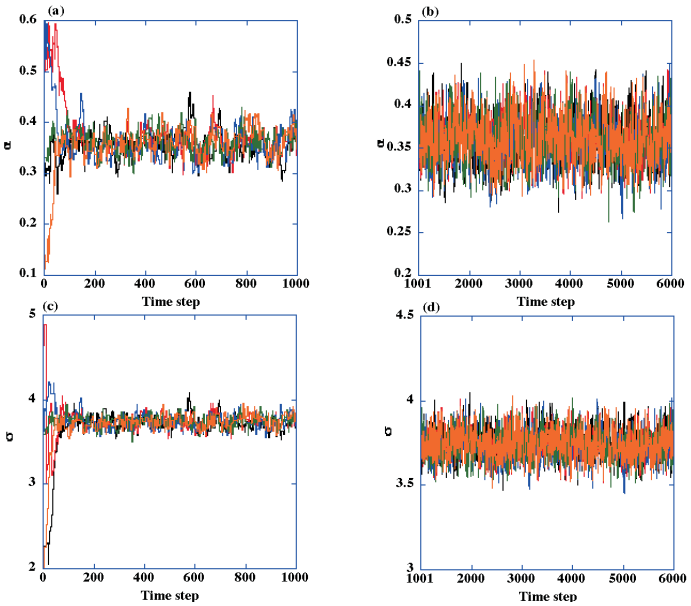

Figure 8.5: *Convergence monitoring plots of the parameters α [(a),(b)] and σ [(c),(d)] for M6T10 under the type (i) prior. (a),(c): First part of five sequences (t=0-1000); (b),(d): Second part of five sequences (t=1001-6000).*

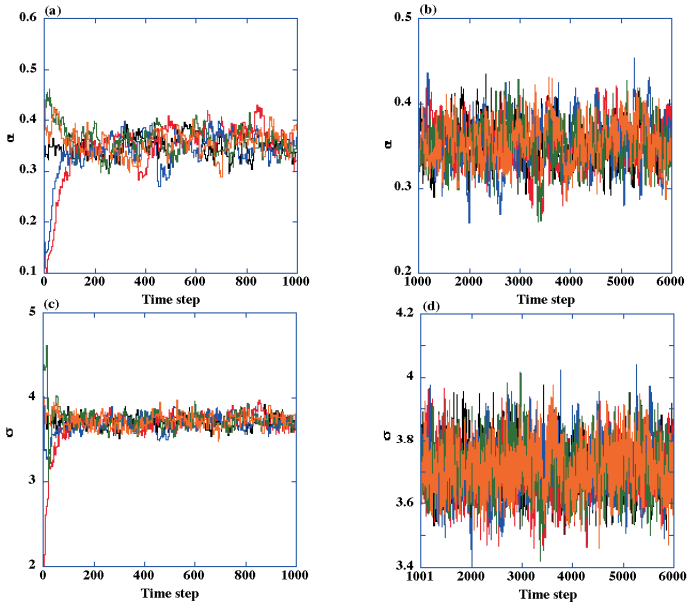

Figure 8.6: *Convergence monitoring plots of the parameters α [(a),(b)] and σ [(c),(d)] for M6T10 under the type (ii) prior. (a),(c): First part of five sequences (t=0-1000); (b),(d): Second part of five sequences (t=1001-6000).*

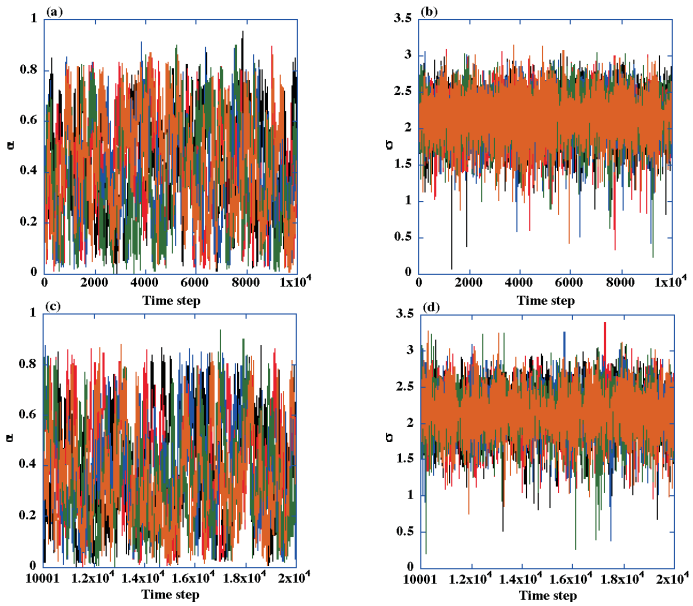

Figure 8.7: *Convergence monitoring plots of the parameters α [(a),(b)] and σ [(c),(d)] for Gulls under the type (i) prior. (a),(c): First part of five sequences (t=0-10000); (b),(d): Second part of five sequences (t=10001-20000).*

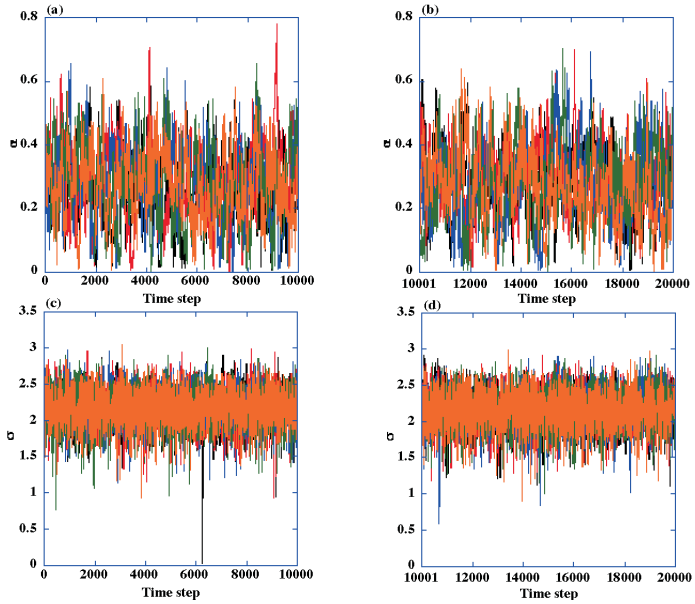

Figure 8.8: *Convergence monitoring plots of the parameters α [(a),(b)] and σ [(c),(d)] for Gulls under the type (ii) prior. (a),(c): First part of five sequences (t=0-10000); (b),(d): Second part of five sequences (t=10001-20000).*

8.2.2 Checking convergence of the single long run

In the previous subsection, we have evaluated the convergence of our single long run based on independent simulated multiple short runs with various starting points. In this subsection, we check the convergence of our single long run. The methods are as follows. We divided the single long runs (second part of 26000 steps for three data sets (Balls, P6T13, M6T10) and of 60000 steps for Gulls, respectively) into prescribed $k = 5$ sequences of equal length, and calculated the potential scale reduction factor $\hat{R}$ for α and σ separately. Table 8.4 summarizes the results. In the Table, we can see that the values of $\hat{R}$ here are well below 1.1 for all cases. This indicates that our single long runs are sufficient for sampling from the posterior density. Therefore, for these single long runs, we have used 25000 samples for three data sets (Balls, P6T13, M6T10), and 50000 samples for Gulls as samples from the joint posterior densities $p(\alpha, \sigma \mid X)$.

Table 8.4: *Results of the values of the potential scale reduction factor $\hat{R}$ as the monitoring convergence diagnostics in the case of single long run for all data sets.*

Data	*param.*	*type*	$\hat{R}$ (*single long run*)
Balls	α	(i)	1.0011
		(ii)	1.0044
	σ	(i)	1.0005
		(ii)	1.0020
P6T13	α	(i)	1.0010
		(ii)	1.0027
	σ	(i)	1.0002
		(ii)	1.0014
M6T10	α	(i)	1.0010
		(ii)	1.0069
	σ	(i)	1.0010
		(ii)	1.0028
Gulls	α	(i)	1.0069
		(ii)	1.0016
	σ	(i)	1.0003
		(ii)	1.0001

8.2.3 Assessing burn-in time of the Metropolis-Hastings algorithm

In §8.2.1-8.2.2, we have evaluated the convergence of our single long run. In this subsection, we investigate the burn-in time T^* for the single long run in detail. The methods proceed as follows. We estimated the burn-in time T^* for each single long run using $k = 5$ independent simulations for all data sets. Then, we calculated the potential scale reduction factor $\hat{R}$ for α and σ separately for initial time step which had not yet got close to stationary.

Figs. 8.9(a)-(d) exhibit, for all data sets, the values of $\hat{R}$ of respective parameters in the case of $k = 5$ sequences under both types (i) and (ii) of prior as time t increases. Figs. 8.9(a)-(c) show that each value of $\hat{R}$ is well below 1.1 for respective parameters until $t = 1000$ under both types (i) and (ii) of prior. Then we evaluated $l^* = T^* = 1000$ for Balls, P6T13 and M6T10.

On the other hand, for Gulls data, the different results were obtained. From Fig. 8.9(d), we can see that the value of $\hat{R}$ for σ is well below 1.1 under both types (i) and (ii) of prior until $t = 1000$. On the contrary, for α, the value of $\hat{R}$ does not show below 1.1 until $t = 1000$. In the case of $(k, l^*, l) = (5, 1000, 5000)$, the values of $\hat{R}$ for α and σ were as follows: $(\hat{R}_\alpha, \hat{R}_\sigma) = (1.3994, 1.0218)$ under the type (i) prior and $(\hat{R}_\alpha, \hat{R}_\sigma) = (1.0530, 1.0052)$ under the type (ii) prior, respectively. These results showed that a bigger number of iteration was needed for the Gulls. As the simulated results, we found that a few thousand iterations were needed as burn-in under both types (i) and (ii) of prior (in particular, the type (i) prior). Then $(l^*, l) = (10000, 10000)$ was chosen, that is, the burn-in time $T^* = 10000$ was estimated for Gulls.

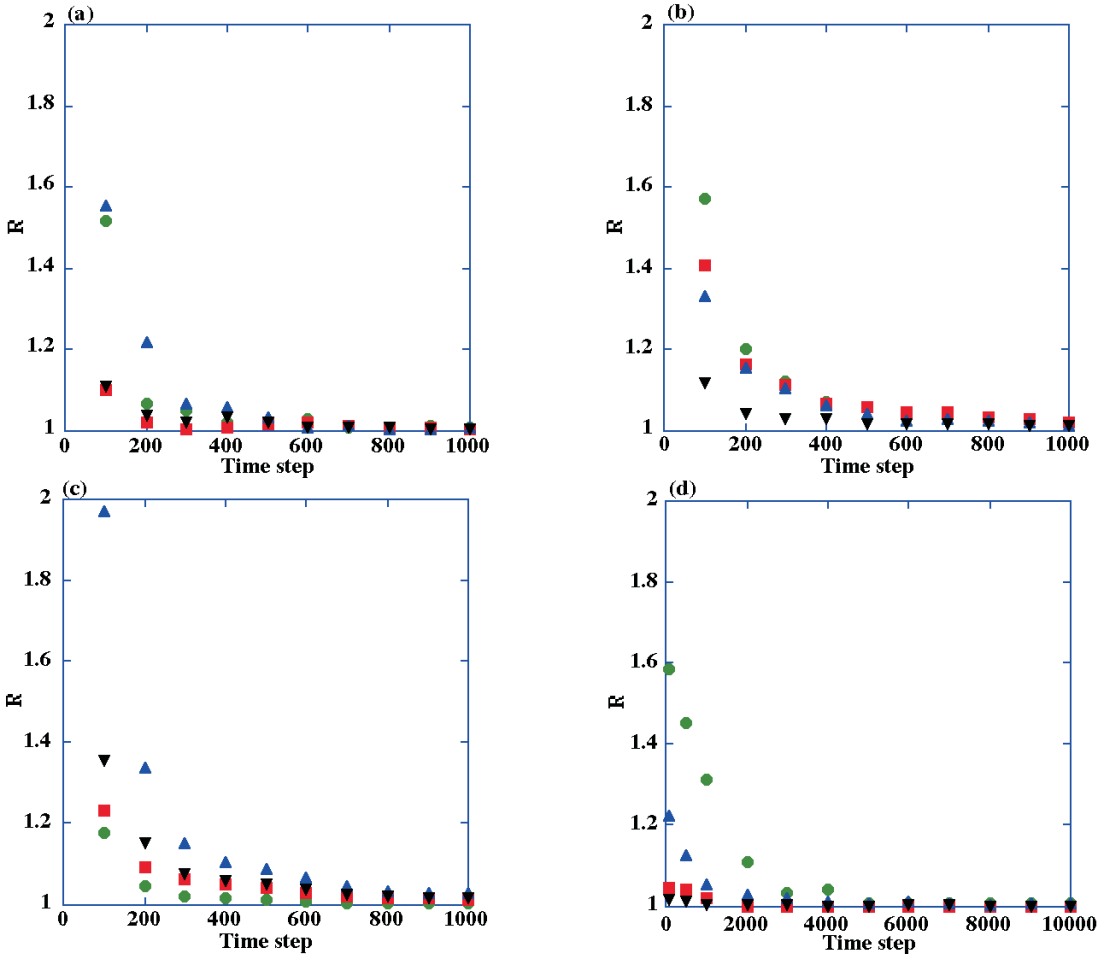

Figure 8.9: *Assessing burn-in time for single long run: the potential scale reduction factor $\hat{R}$ as the convergence diagnostics vs. the initial iteration time t in the case of $k = 5$ parallel sequences for all data sets. The symbol ● and ■ stand for $\hat{R}$ of α and σ under the type (i) prior, and the symbol ▲ and ▼ represent $\hat{R}$ of α and σ under the type (ii) prior, respectively: (a) Balls; (b) P6T13; (c) M6T10; (d) Gulls.*

Figs. 8.10(a)-(d) show the time series plots for the single long runs of the Monte Carlo output of α and σ for Balls data under both types (i) and (ii) of prior. The marginal histograms based on 25000 successive draws with a burn-in of 1000 iterations are displayed in Figs. 8.11(a)-(d) which correspond to Figs. 8.10(a)-(d), respectively. Figs. 8.12(a)-(d) and 8.14(a)-(d) show the time series plots for two data sets P6T13 and M6T10 under the two priors. Figs. 8.13(a)-(d) and 8.15(a)-(d) are the marginal histograms based on 25000 successive draws with a burn-in of 1000 iterations corresponding to Figs. 8.12(a)-(d) and 8.14(a)-(d), respectively. And the time series plots for Gulls data under the two priors are given in Figs. 8.16(a)-(d). Figs. 8.17(a)-(d) are the marginal histograms based on 50000 successive draws with a burn-in of 10000 iterations corresponding to Figs. 8.16(a)-(d), respectively.

Figs. 8.18 and 8.19 illustrate scatter plots of 25000 or 50000 draws from the joint posterior densities $p(\alpha, \tau \mid X)(\tau = \rho\sigma^2)$, defined in Eq. (4.8), for all data sets under both types (i) and (ii) of prior, respectively. Table 8.5 gives the values of the correlation coefficient between α and τ. We see from Table 8.5 that, for three data sets (Balls, P6T13, M6T10), the correlation between α and τ is relatively high and each of the values for both types (i) and (ii) of prior is very close. On the other hand, for the Gulls data set, the correlation become small for both types (i) and (ii) of prior, and the correlation coefficient of the type (ii) prior is greater than that for the type (i) prior. It seems that the prior affects the correlation for Gulls.

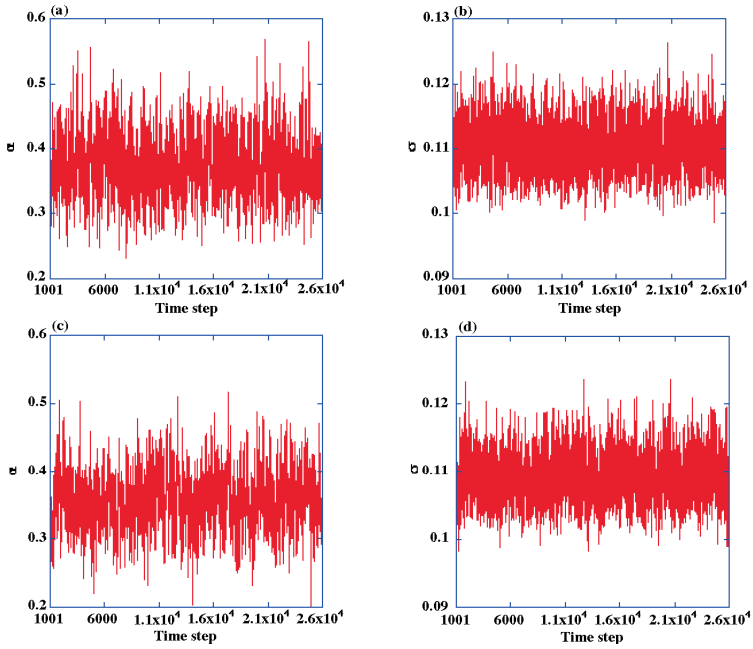

Figure 8.10: *Time series plots for α and σ of the second parts of the Markov chain simulation for Balls under both types (i) and (ii) of prior. (a),(b): the type (i) prior; (c),(d): the type (ii) prior.*

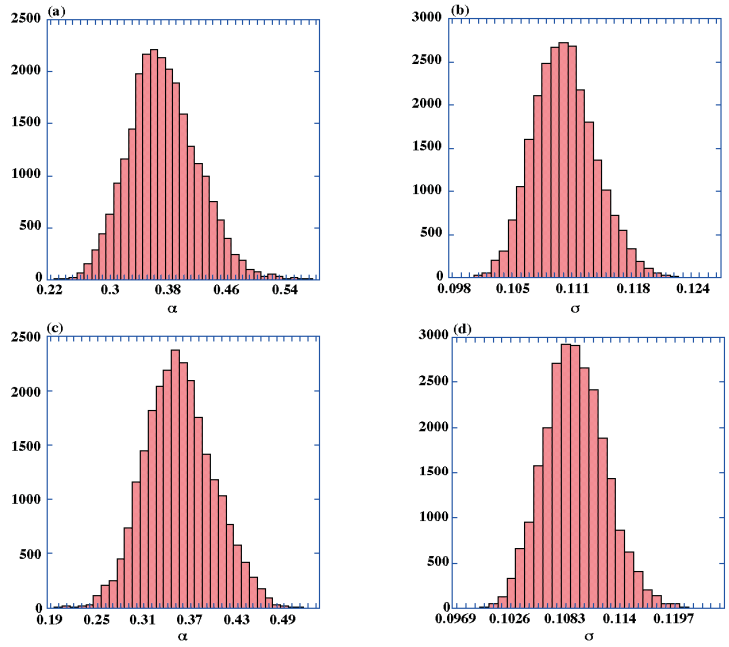

Figure 8.11: *Marginal histograms of α and σ of 25000 simulation draws for Balls under both types (i) and (ii) of prior. (a),(b): the type (i) prior; (c),(d): the type (ii) prior.*

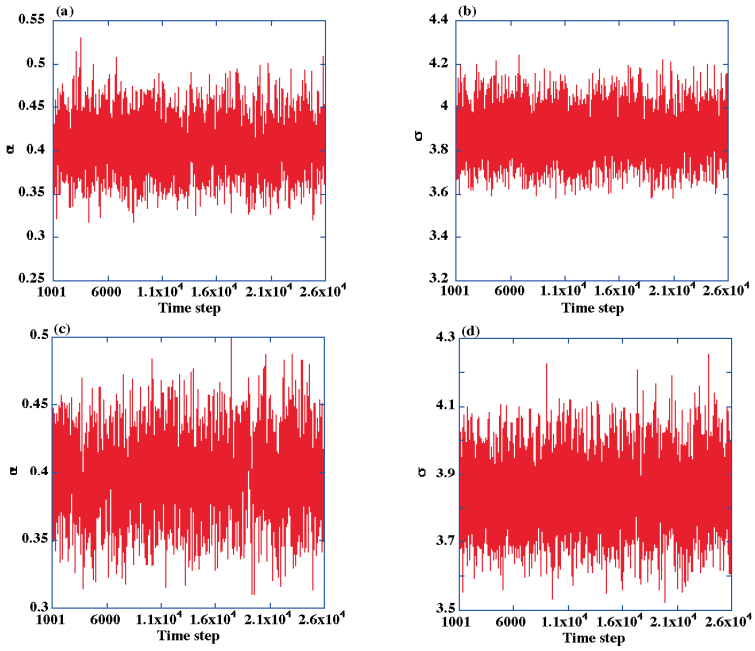

Figure 8.12: *Time series plots for α and σ of the second parts of the Markov chain simulation for P6T13 under both types (i) and (ii) of prior. (a),(b): the type (i) prior; (c),(d): the type (ii) prior.*

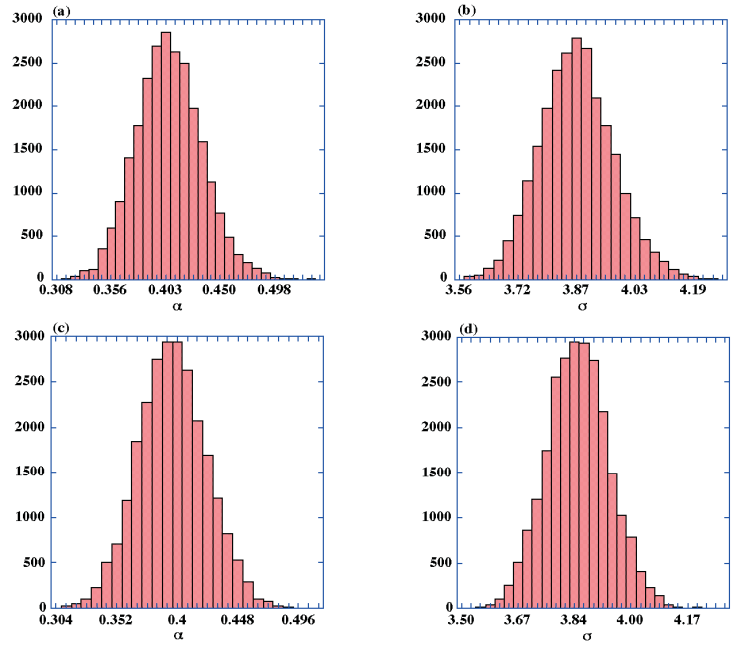

Figure 8.13: *Marginal histograms of α and σ of 25000 simulation draws for P6T13 under both types (i) and (ii) of prior. (a),(b): the type (i) prior; (c),(d): the type (ii) prior.*

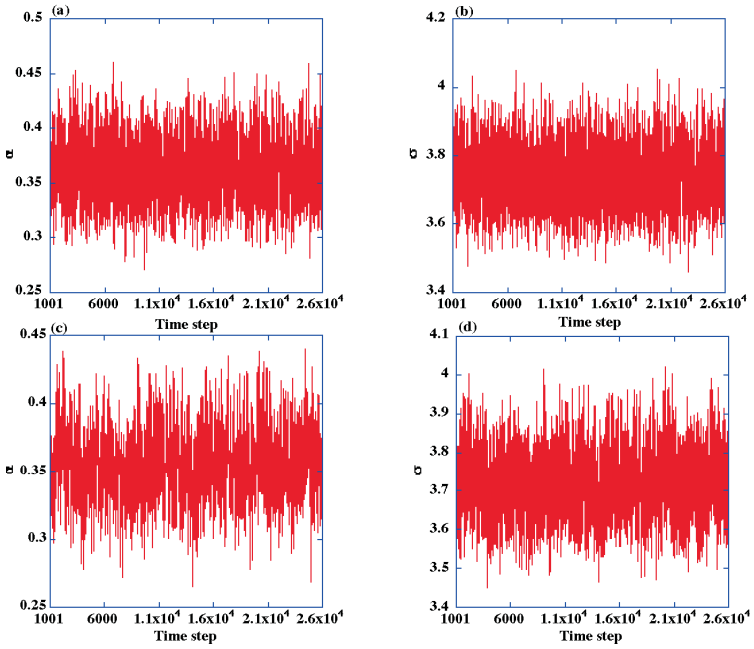

Figure 8.14: *Time series plots for α and σ of the second parts of the Markov chain simulation for M6T10 under both types (i) and (ii) of prior. (a),(b): the type (i) prior; (c),(d): the type (ii) prior.*

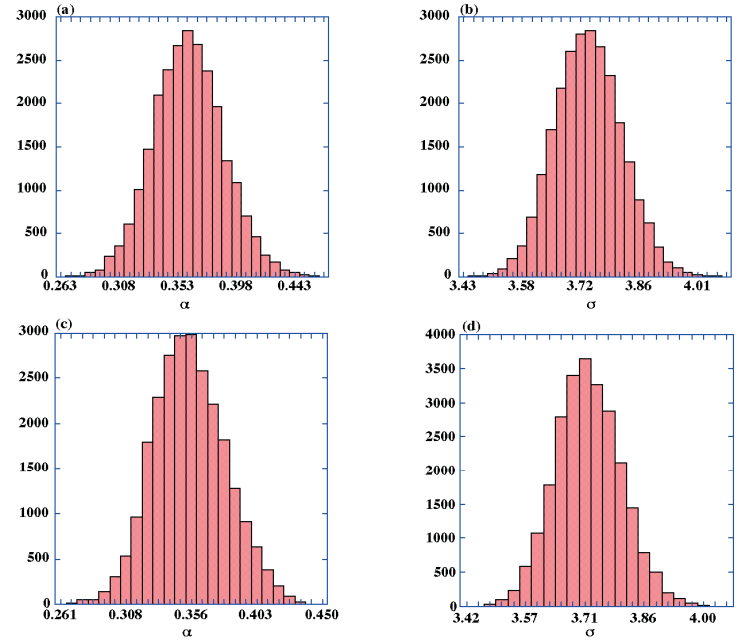

Figure 8.15: *Marginal histograms of α and σ of 25000 simulation draws for M6T10 under both types (i) and (ii) of prior. (a),(b): the type (i) prior; (c),(d): the type (ii) prior.*

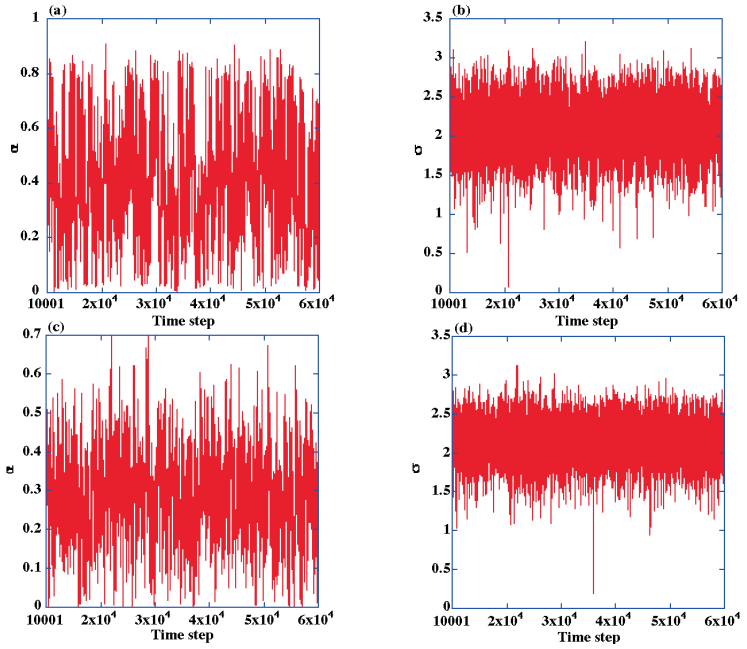

Figure 8.16: *Time series plots for α and σ of the second parts of the Markov chain simulation for Gulls under both types (i) and (ii) of prior. (a),(b): the type (i) prior; (c),(d): the type (ii) prior.*

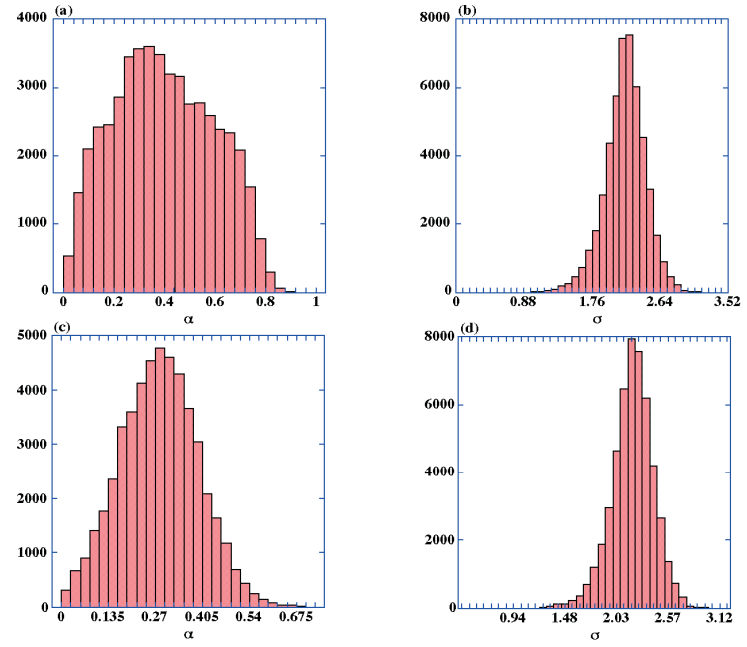

Figure 8.17: *Marginal histograms of α and σ of 50000 simulation draws for Gulls under both types (i) and (ii) of prior. (a),(b): the type (i) prior; (c),(d): the type (ii) prior.*

Table 8.5: *The values of correlation coefficient (C.C.) between α and τ for both types (i) and (ii) of prior.*

Data	*C.C. of type(i)*	*C.C. of type(ii)*
Balls	0.68834	0.66286
P6T13	0.72794	0.70926
M6T10	0.73414	0.73176
Gulls	0.10410	0.19054

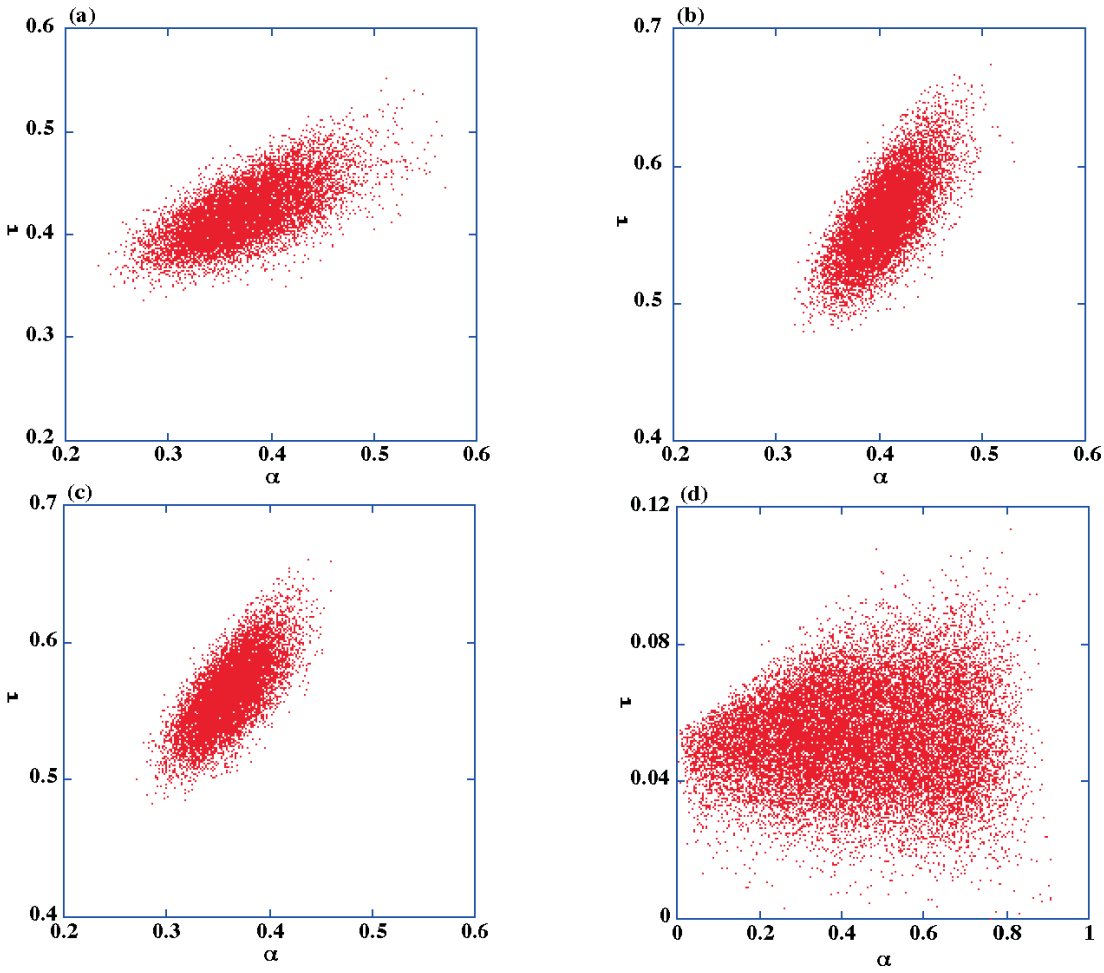

Figure 8.18: *Scatter plots of the joint posterior density of α and τ under the type (i) prior: (a)-(c) 25000 simulation draws for three data sets (Balls, P6T13, M6T10); (d) 50000 simulation draws for Gulls.*

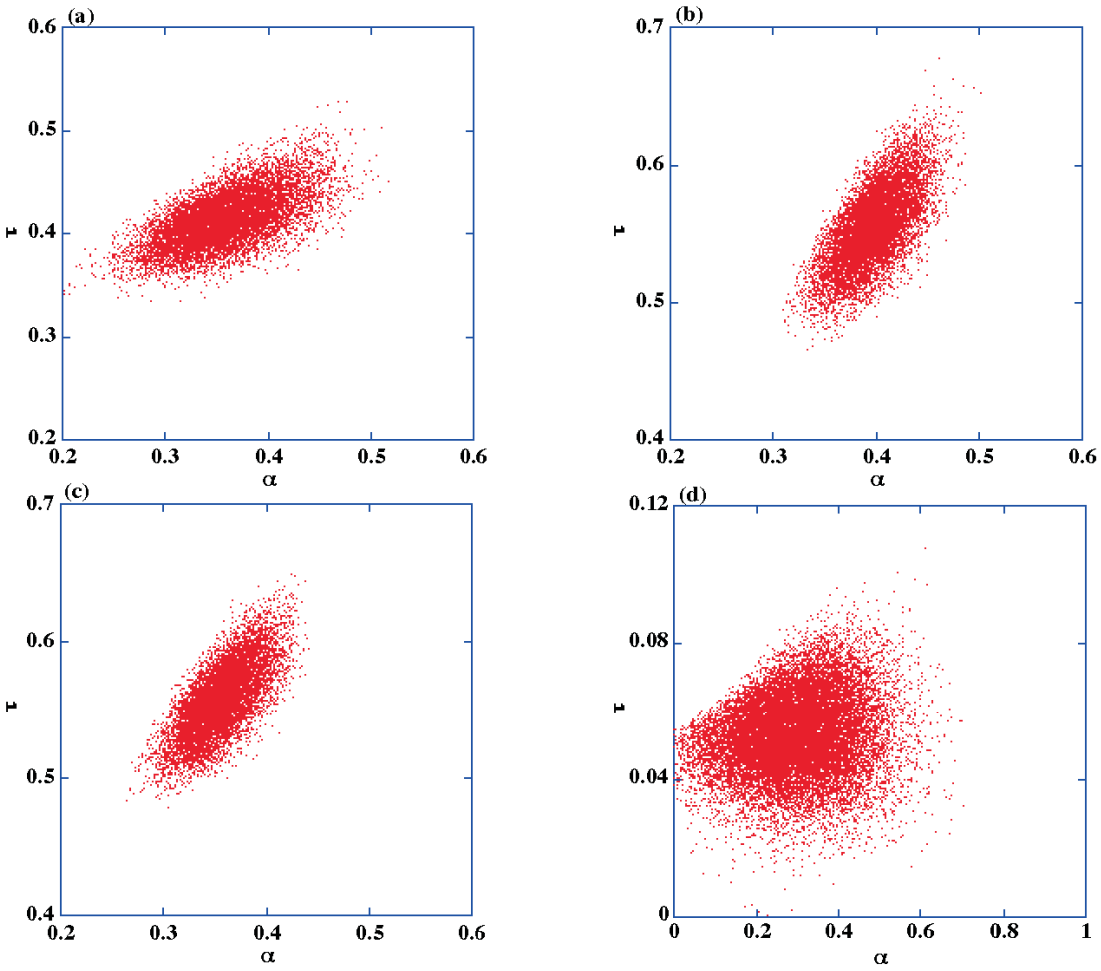

Figure 8.19: *Scatter plots of the joint posterior density of α and τ under the type (ii) prior: (a)-(c) 25000 simulation draws for three data sets (Balls, P6T13, M6T10); (d) 50000 simulation draws for Gulls.*

8.2.4 Case of the independence sampler

For comparison, we applied the two independence samplers to the jumping densities of (α, σ) under the type (i) priors, as described in §5.2.1. The values of hyperparameters (a, b, c, d) in Eqs. (4.4)-(4.5) for the type (i) prior densities and for their jumping densities in Eqs. (5.9)-(5.10) are given in Table 8.1 for all data sets. Then we performed a single long run simulation for all data sets. The burn-in time T^* and the stopping time T were estimated by using Gelman and Rubins's method, as remarked in §8.2. Table 8.6 provides the values of $\hat{R}$ for respective parameters in the case of $(k, l^*, l) = (5, 1000, 5000)$ for all data sets. Since each value of $\hat{R}$ was well below 1.1 for both parameters from Table 8.6, we estimated $(T^*, T) = (1000, (5 \times 5000 + 1000 = 26000))$ for all data sets. Then we carried out the MCMC simulation of the stopping time $T = 26000$ for each data.

Table 8.6: *Results of the values of the potential scale reduction factor $\hat{R}$ as the monitoring convergence diagnostics for the independence samplers of (α, σ) in the case of k=5 parallel sequences for all data sets.*

Data	*param.*	$\hat{R}$ (5 *sequences*)
Balls	α	1.0051
	σ	1.0048
P6T13	α	1.0101
	σ	1.0148
M6T10	α	1.0038
	σ	1.0051
Gulls	α	1.0038
	σ	1.0051

Furthermore, to check the convergence, we divided the single long runs (second part of 26000 steps for all data sets) into five sequences of equal length, and calculated the potential scale reduction factor $\hat{R}$ for α and σ separately. Table 8.7 summarizes the results for all data sets. We can see that the values of $\hat{R}$ here are well below 1.1 for all cases. Then it seems that our single long runs are sufficient for sampling from the target posterior density.

Table 8.7: *Results of the values of the potential scale reduction factor* $\hat{R}$ *as the monitoring convergence diagnostics for the independence samplers of* (α, σ) *in the case of single long run for all data sets.*

Data	*param.*	$\hat{R}$ (*single long run*)
Balls	α	1.0055
	σ	1.0016
P6T13	α	1.0103
	σ	1.0104
M6T10	α	1.0159
	σ	1.0088
Gulls	α	1.0023
	σ	1.0013

Figs. 8.20-8.23 show the time series plots for the single long runs of the Monte Carlo output of α and σ for all data sets under the type (i) prior with the independence samplers, and their marginal histograms which correspond to Figs. 8.20-8.23, respectively. From Figs. 8.20-8.22, we can see that each run of respective parameters does not seem to be stationary. In these simulations, each of the acceptance rate was around 0.01. From Figs. 8.23 (a) and (b), we can see that each run of respective parameters looks stationary, but the acceptance rate was about 0.1. As the simulated results, these values of the acceptance rate were very low. From Figs. 8.20-8.23, it seems that most of the shape of the marginal posteriors are distorted. Although the values of $\hat{R}$ are well below 1.1 for the independence samplers (from Tables 8.6-8.7), we found that the use of the independence samplers is not suitable for inference (see §6.3).

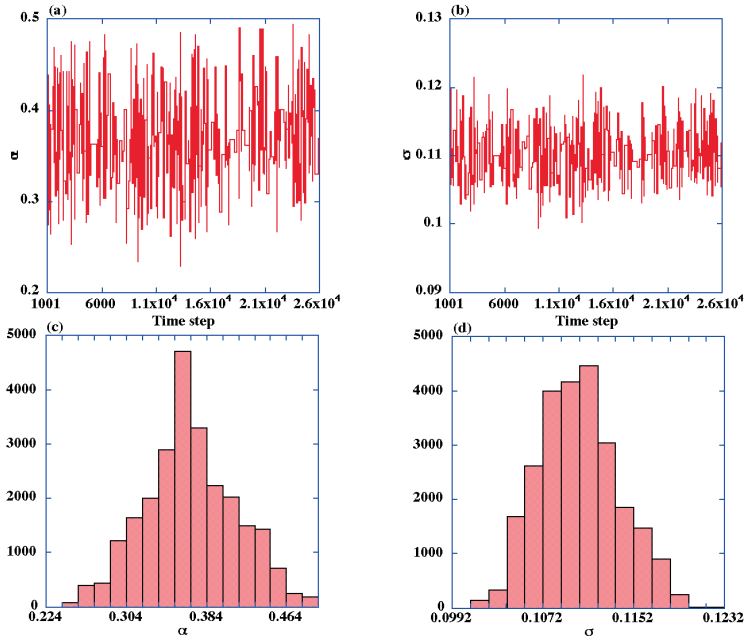

Figure 8.20: *Time series plots for* (α, σ) *of the second parts of the simulation for Balls under the (i) uniform priors with the independence samplers and their marginal histograms of* (α, σ) *of 25000 simulation draws.*

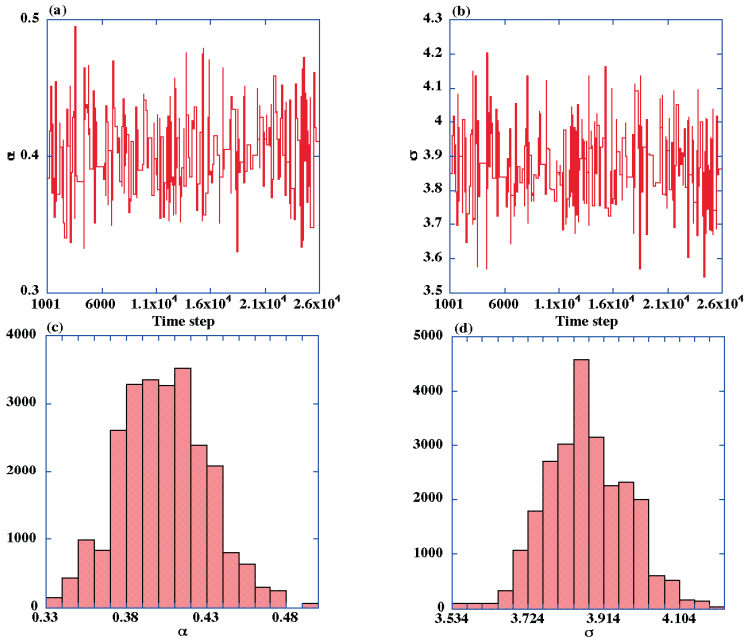

Figure 8.21: *Time series plots for* (α, σ) *of the second parts of the simulation for P6T13 under the (i) uniform priors with the independence samplers and their marginal histograms of* (α, σ) *of 25000 simulation draws.*

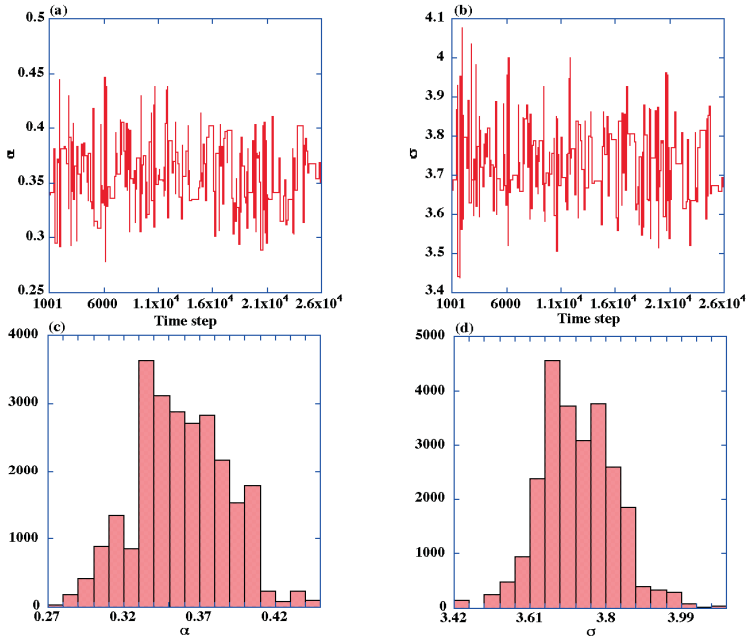

Figure 8.22: *Time series plots for* (α, σ) *of the second parts of the simulation for M6T10 under the (i) uniform priors with the independence samplers and their marginal histograms of* (α, σ) *of 25000 simulation draws.*

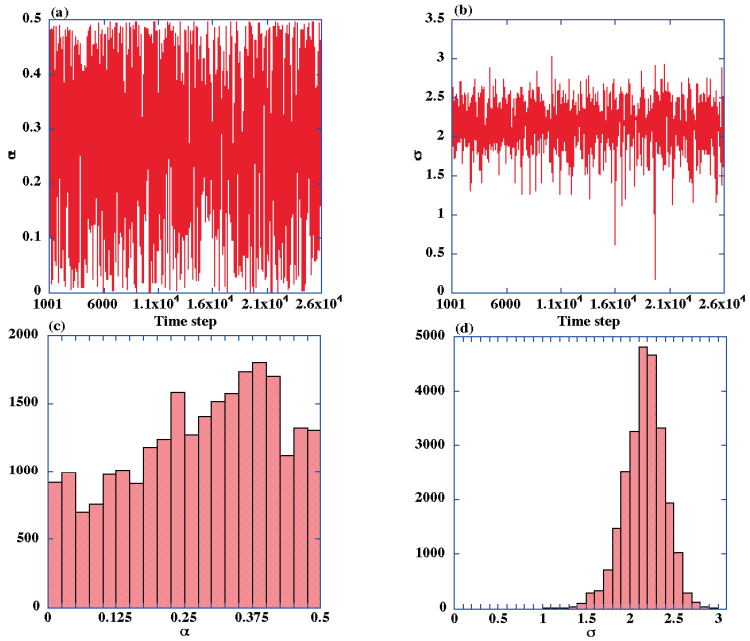

Figure 8.23: *Time series plots for* (α, σ) *of the second parts of the simulation for Gulls under the (i) uniform priors with the independence samplers and their marginal histograms of* (α, σ) *of 25000 simulation draws.*

8.3 Parametric fitting of the generalized gamma distribution to the marginal posterior densities

We fitted the *generalized gamma distribution* to the simulated marginal posterior densities $p(\alpha \mid X)$ and $p(\sigma \mid X)$ in Eqs. (4.9)-(4.10) for the parameters α and σ, respectively. The generalized gamma distribution with three parameters λ, ω and ζ is defined as:

$$g(x \mid \lambda, \omega, \zeta) = \frac{\lambda \omega^{\zeta/\lambda}}{\Gamma(\zeta/\lambda)} x^{\zeta-1} \exp(-\omega x^{\lambda}); \quad x > 0, \ \lambda, \omega, \zeta > 0. \tag{8.1}$$

Note that setting the parameter $\lambda = 1$ gives the ordinary gamma distribution. The reason for our choice of the generalized gamma distribution is that it can represent, by adjusting three parameters (λ, ω, ζ), a wide range of distribution with single mode and with the range $(0, \infty)$ of its variable (Tanemura (2003)).

Furthermore, in order to obtain posterior inference from our simulations for all data sets, we considered parametric fitting of the generalized gamma distribution to our simulated marginal posterior densities, will be discussed in §8.4.

The class intervals, the number of classes and the 95% quantiles of the observed histograms of respective parameters for all data sets are given in Table 8.8. Then we obtained numerically the estimates $(\hat{\lambda}, \hat{\omega}, \hat{\zeta})$ by using the quasi-Newton method for all data sets. Table 8.9 summarizes the results. In the table, chi-square goodness-of-fit statistics together with skewness and kurtosis are also given. Figs. 8.24 and 8.25 illustrate the simulated and the estimated marginal posteriors of respective parameters for all data sets under both types (i) and (ii) of prior, respectively. We see here that the generalized gamma distribution fits sufficiently well with the simulated marginal posterior density in every case.

Table 8.8: *Class intervals, number of classes and 95% quantiles of the observed histograms of the model parameters.*

Data	*param.*	*type*	*Class intervals*	*No. of classes*	[2.5%, 97.5%]
Balls	α	(i)	0.010	32	[0.28758, 0.46964]
		(ii)	0.010	29	[0.27423, 0.44492]
	σ	(i)	0.00095	25	[0.10404, 0.11754]
		(ii)	0.00095	23	[0.10323, 0.11592]
P6T13	α	(i)	0.0079	24	[0.35183, 0.46504]
		(ii)	0.0080	24	[0.34421, 0.44932]
	σ	(i)	0.026	24	[3.6910, 4.0766]
		(ii)	0.028	24	[3.6697, 4.0313]
M6T10	α	(i)	0.0075	24	[0.30983, 0.41495]
		(ii)	0.0079	22	[0.30560, 0.40770]
	σ	(i)	0.024	23	[3.5838, 3.9035]
		(ii)	0.029	18	[3.5698, 3.8818]
Gulls	α	(i)	0.04	27	[0.062086, 0.75753]
		(ii)	0.027	26	[0.061925, 0.50070]
	σ	(i)	0.088	22	[1.5891, 2.6790]
		(ii)	0.078	26	[1.7141, 2.6086]

Table 8.9: *Estimates* $(\hat{\lambda}, \hat{\omega}, \hat{\zeta})$ *of the generalized gamma distribution (8.1) fitted to the simulated marginal posterior densities for all data sets under both types (i) and (ii) of prior. In the fifth column,* χ^2 *represents the chi-square goodness-of-fit statistics between the simulated posterior and its generalized gamma fit,* $\chi^2_{0.05}$ *is the critical value of probability 5% and* ν *is the degrees of freedom. In the last column, 'Skew.' and 'Kurt.' stand for skewness and kurtosis, respectively.*

Data	*param.*	*type*	($\hat{\lambda}$, $\hat{\omega}$, $\hat{\zeta}$)	(χ^2, $\chi^2_{0.05}$, ν)	$(Skew., Kurt.)$
Balls	α	(i)	(1.5811, 126.30, 41.701)	(1.211, 44.99, 31)	(0.564, −1.22)
		(ii)	(2.7171, 161.560, 27.030)	(1.491, 41.34, 28)	(0.536, −1.25)
	σ	(i)	(1.0029, 9175.5, 1009.3)	(8.732, 36.42, 24)	(0.533, −1.25)
		(ii)	(1.0026, 10491.8, 1142.3)	(9.227, 33.92, 22)	(0.493, −1.29)
P6T13	α	(i)	(3.4057, 390.55, 61.576)	(2.345, 35.17, 23)	(0.497, −1.23)
		(ii)	(3.2953, 426.46, 67.782)	(0.9013, 35.17, 23)	(0.577, −1.21)
	σ	(i)	(4.7359, 0.12513, 364.00)	(0.6166, 35.17, 23)	(0.482, −1.30)
		(ii)	(4.6805, 0.14490, 373.00)	(0.1847, 35.17, 23)	(0.597, −1.19)
M6T10	α	(i)	(2.7200, 403.90, 68.970)	(0.8819, 35.17, 23)	(0.510, −1.28)
		(ii)	(2.6455, 419.42, 72.108)	(1.346, 32.67, 21)	(0.485, −1.30)
	σ	(i)	(5.8190, 0.02789, 351.00)	(0.2140, 33.92, 22)	(0.475, −1.31)
		(ii)	(5.8755, 0.02868, 379.60)	(0.3897, 27.59, 17)	(0.446, −1.33)
Gulls	α	(i)	(3.7595, 5.9889, 1.5802)	(0.8297, 32.67, 21)	(−0.427, −1.16)
		(ii)	(3.6795, 41.590, 2.5302)	(1.459, 37.65, 25)	(0.303, −1.45)
	σ	(i)	(6.0909, 0.01916, 16.023)	(5.215, 38.89, 26)	(1.08, −0.365)
		(ii)	(7.2422, 0.00760, 19.279)	(25.27, 37.65, 25)	(1.11, −0.301)

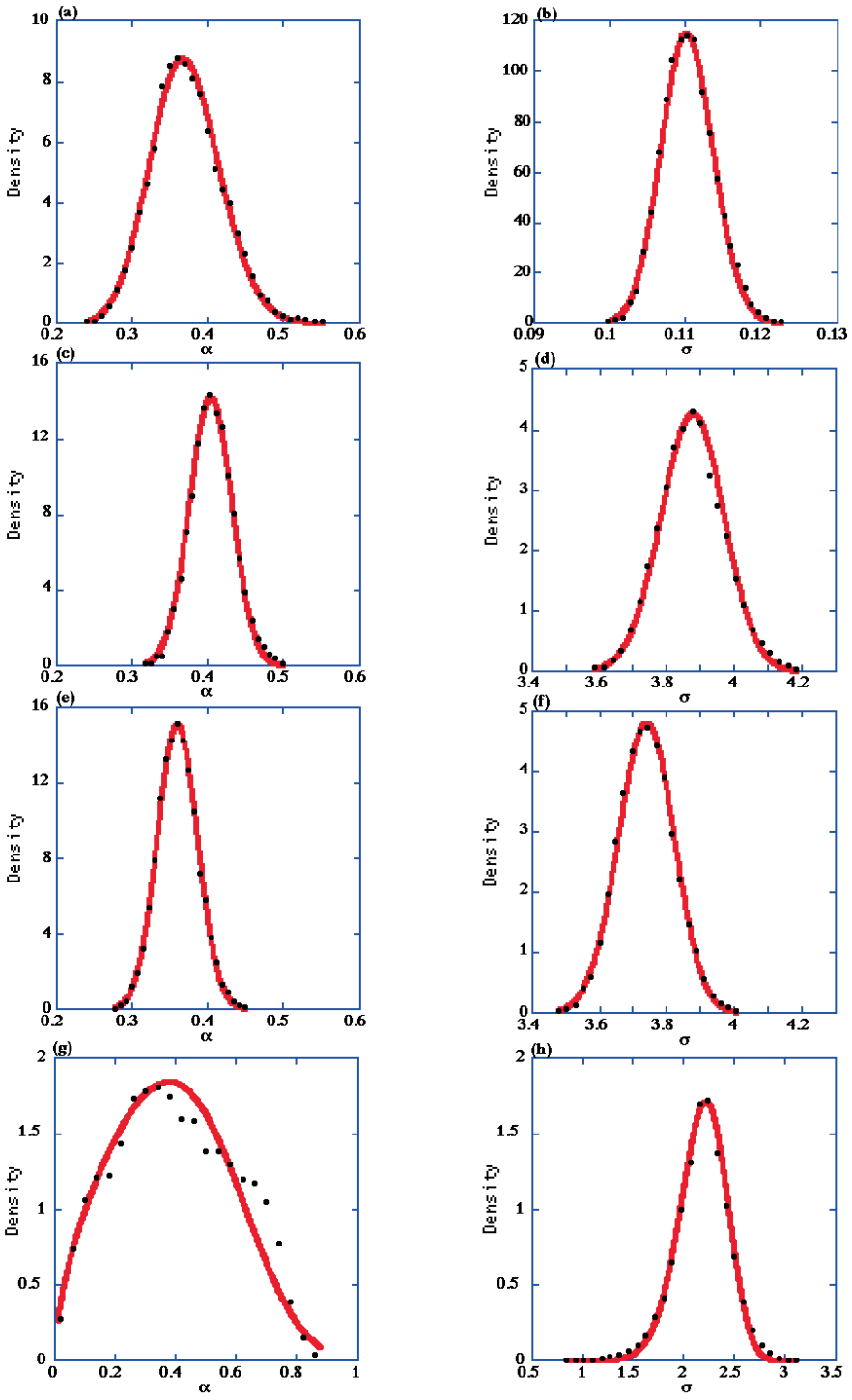

Figure 8.24: *Fitting the generalized gamma distribution to the simulated marginal posterior density, which is indicated by the symbol •, of respective parameters under the type (i) prior. (a),(b): Balls; (c),(d): P6T13; (e),(f): M6T10; (g),(h): Gulls.*

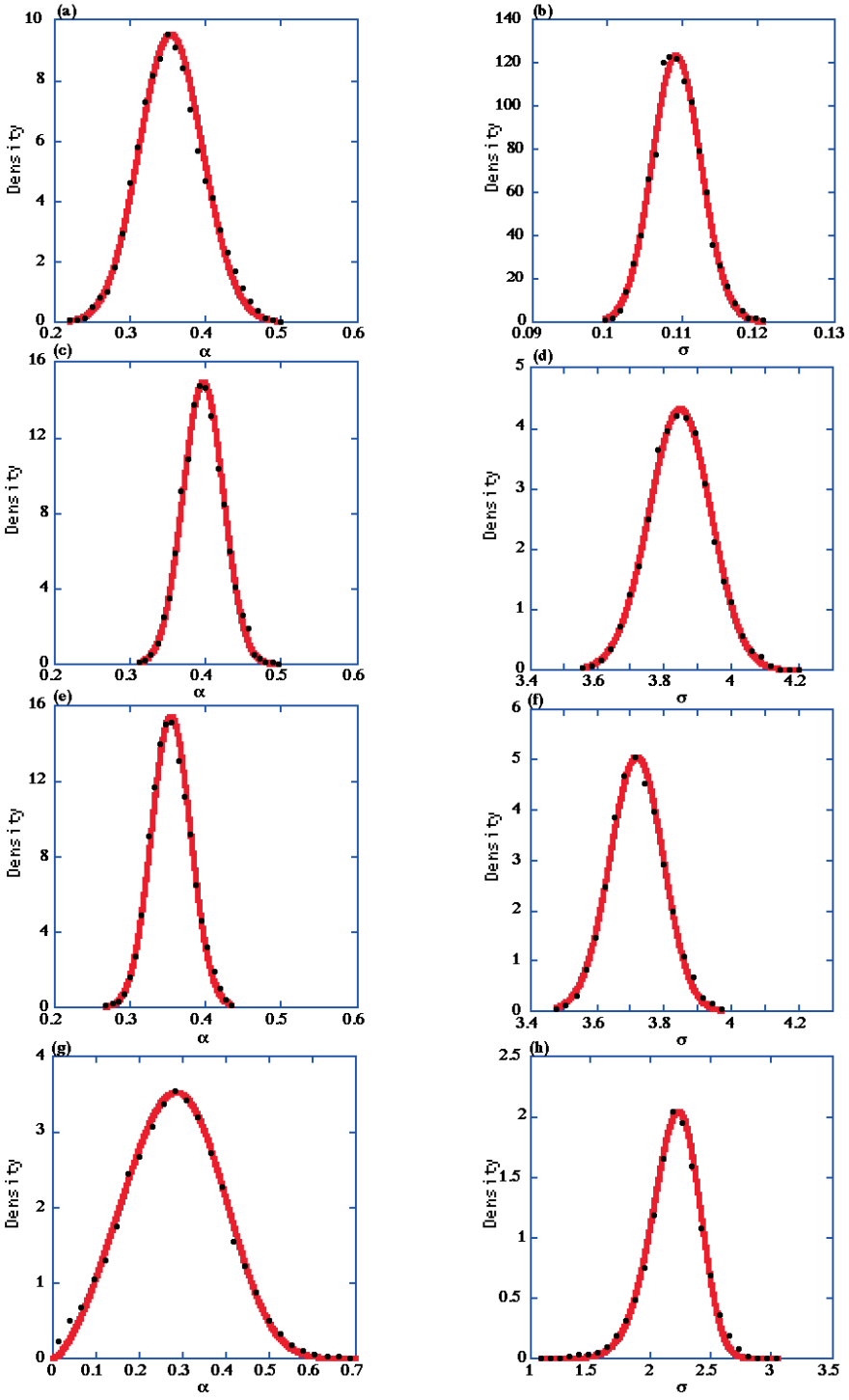

Figure 8.25: *Fitting the generalized gamma distribution to the simulated marginal posterior density, which is indicated by the symbol •, of respective parameters under the type (ii) prior. (a),(b): Balls; (c),(d): P6T13; (e),(f): M6T10; (g),(h): Gulls.*

8.4 Posterior inference

Table 8.10 gives simulated marginal posterior summary statistics for the model parameters under both types (i) and (ii) of prior for all data sets. For finding posterior modes of respective parameters, the *Newton's method* was applied to the fitted generalized gamma distributions. Marginal posterior quantiles were computed using numerical integration of the fitted densities. Since all simulated marginal posteriors of respective parameters are unimodal and not highly skewed, the 95% highest posterior density intervals include the posterior mode and the mean. In the last column of the table, the maximum likelihood estimates $(\hat{\alpha}, \hat{\sigma})$ are also given. We calculated the values of $(\hat{\alpha}, \hat{\sigma})$ from the log-likelihood (3.14) numerically (see §6.1.1). For calculating the total potential energy in Eq. (3.14), we used again the periodic boundary condition (see §8.5). In the table, the estimates of Ogata and Tanemura (1989) are also given.

From Table 8.10, we see that the marginal posterior modes of each parameter are respectively close to their maximum likelihood estimates under the type (i) uniform prior for three data sets (Balls, P6T13, M6T10). It is because, for these three data, the marginal posteriors under the uniform prior are proportional to the likelihood. Then under the type (ii) normal prior, the results of marginal posterior mode of the respective parameters tend to be smaller than those for the type (i) prior for three data. Marginal posterior densities which were fitted to the generalized gamma distribution of α and σ for all data sets under both types (i) and (ii) of prior are given in Figs. 8.26(a)-(h).

On the other hand, for the Gulls data set, the results of σ under both types (i) and (ii) of prior are similar, but the results are different for parameter α. Although the marginal posterior density of σ for the type (ii) prior shows the slightly larger peak than that for the type (i) prior, they are both said to be similar (see Fig. 8.26(h)). On the contrary, for the parameter α, the marginal posterior shows a different spread between type (i) and (ii) priors, as indicated in Fig. 8.26(g). In order to investigate this difference, we performed our Bayesian estimation procedure by using the simulated equilibrium point patterns generated by MCMC of the Soft-Core models for the cases of $\tau = 0.05, \alpha = 0.2, 0.3, 0.4$ ($N = 500, V = \sqrt{500} \times \sqrt{500}$), as stated in §6.1.3. These values of parameters τ and α were selected, as true values, so that they are near to the maximum likelihood estimate (MLE) of corresponding parameters for Gulls data. Especially, the true values of α, namely $\alpha = 0.2, 0.3$ and 0.4 were chosen due to a rather big standard error (s.e.= 0.22) of the MLE ($\hat{\alpha} = 0.29$) (see Table 8.10). The joint posterior densities of (α, τ) were computed, for each of type (i) and (ii) priors, from the equilibrium point patterns simulated under the set of these parameter values. As a result, the marginal posterior density of α did not show a big difference between type (i) and (ii) priors for all set of the true parameter values. This result suggests that the different spread between

marginal posteriors of α from type (i) and from type (ii) priors in Gulls case should be attributed to the characteristic of the Gulls data set itself.

Furthermore, we also examined the cases of $\tau = 0.05, \alpha = 0.2, 0.3, 0.4$ ($N = 100, V = \sqrt{100} \times \sqrt{100}$), as described in §6.2. Then, we obtained the similar results to the case of the Gulls data: for the cases of relatively small reduced density, the marginal posteriors of α showed different spread between two types of priors regardless of the jumping densities. It means that the different spread of $p(\alpha \mid X)$ found in Gulls case might not be attributed only to the characteristic of the Gulls data set itself. We found that when the number of points N is small, the choice of the priors for α affects the marginal posterior for the cases of the relatively reduced density.

Table 8.10: *Posterior modes, means and the 95% highest posterior density intervals (95% HPD intervals) for the model parameters under both types (i) and (ii) of prior. The values of the reduced density τ are calculated by Eq. (3.4). In the last column, the maximum likelihood estimates (MLE) are also given; values in the parentheses are MLE together with standard errors (s.e.) of Ogata and Tanemura (1989).*

Data	*param.*	*type*	*Mode*	*Mean*	95% *HPD intervals*	*MLE*
Balls	α	(i)	0.36569	0.37268	[0.28428, 0.46277]	0.35843 $(0.35, s.e. = 0.04)$
		(ii)	0.35353	0.35531	[0.27329, 0.43676]	
	σ	(i)	0.11027	0.11033	[0.10368, 0.11728]	0.10969 $(0.11, s.e. = 0.003)$
		(ii)	0.10912	0.10928	[0.10299, 0.11563]	
	τ	(i)	0.42031	0.42077	[0.37158, 0.47545]	0.41590 $(0.42, s.e. = 0.0003)$
		(ii)	0.41162	0.41312	[0.36665, 0.46217]	
P6T13	α	(i)	0.40372	0.40580	[0.34803, 0.45792]	0.40212 $(0.41, s.e. = 0.03)$
		(ii)	0.39672	0.39616	[0.34395, 0.44852]	
	σ	(i)	3.8772	3.8746	[3.6901, 4.0567]	3.8681 $(3.88, s.e. = 0.10)$
		(ii)	3.8479	3.8462	[3.6635, 4.0251]	
	τ	(i)	0.56311	0.56233	[0.51007, 0.61645]	0.56046 $(0.56, s.e. = 0.0004)$
		(ii)	0.55463	0.55446	[0.50274, 0.60689]	
M6T10	α	(i)	0.35949	0.36086	[0.30816, 0.41171]	0.35661 $(0.36, s.e. = 0.03)$
		(ii)	0.35396	0.35434	[0.30420, 0.40506]	
	σ	(i)	3.7404	3.7391	[3.5728, 3.8979]	3.7304 $(3.73, s.e. = 0.08)$
		(ii)	3.7191	3.7200	[3.5597, 3.8691]	
	τ	(i)	0.56225	0.56186	[0.51299, 0.61060]	0.55924 $(0.56, s.e. = 0.0003)$
		(ii)	0.55586	0.55639	[0.50924, 0.60162]	
Gulls	α	(i)	0.37789	0.39980	[0.055751, 0.77396]	0.29190 $(0.29, s.e. = 0.22)$
		(ii)	0.28605	0.28068	[0.081326, 0.48899]	
	σ	(i)	2.2201	2.1809	[1.6888, 2.6115]	2.2511 $(2.26, s.e. = 0.21)$
		(ii)	2.2291	2.1952	[1.7718, 2.5499]	
	τ	(i)	0.054215	0.052322	[0.031372, 0.075019]	0.055746 $(0.06, s.e. = 0.0005)$
		(ii)	0.054660	0.053550	[0.034533, 0.071520]	

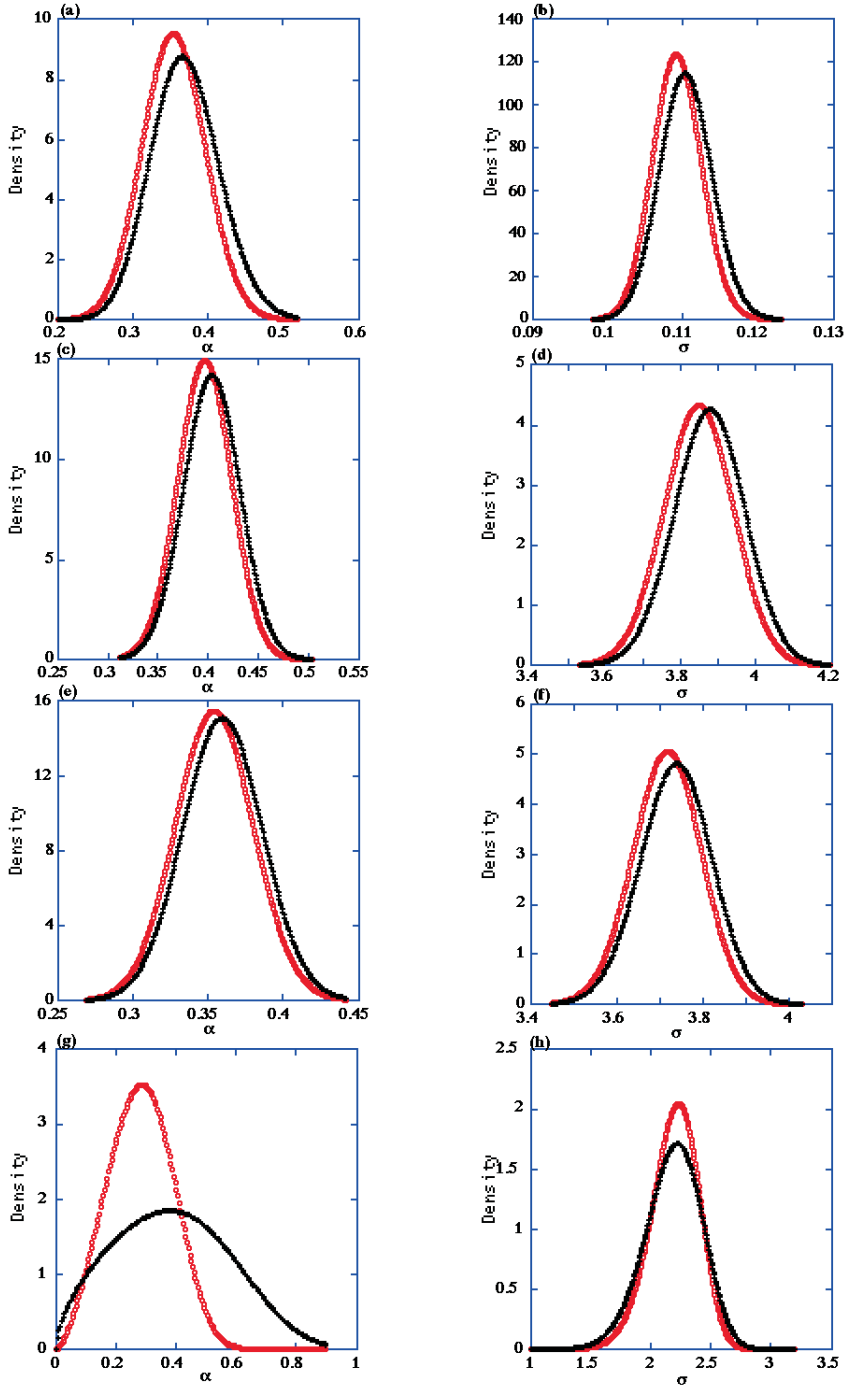

Figure 8.26: *Comparison of the generalized gamma fits of marginal posterior densities of α and σ ($p(\alpha \mid X)$ and $p(\sigma \mid X)$) under both types (i) and (ii) of prior (the symbol + indicates the densities of the type (i) prior, and the symbol $\circ$ represents the densities of the type (ii) prior). (a),(b): Balls; (c),(d): M6T10; (e),(f): P6T13; (g),(h): Gulls.*

8.5 Model evaluation: diagnostic posterior predictive checking by using L-statistics

To examine whether our results of Bayesian estimation of two parameter Soft-Core models were reasonable or not, we computed the *L-statistics* (or, *L-function*) $L(r)$ of the observed data and of the simulated data (equilibrium point patterns) predicted from the marginal posterior densities of the parameters. The L-statistics is known as the second-order diagnostic characteristic for stationary of the spatial point patterns (Ripley (1977, 1979a, 2004), Besag (1977), Stoyan *et al.* (1995), Diggle (2003)).

The L-function is the transformed function of the *K-function*. The K-function is similarly categorized into the second-order diagnostic characteristic and its definition $K(r)$ is,

$$\rho K(r) = E[\text{number of individuals within distance } r \text{ of an arbitrary individual}], \quad r \geq 0, \quad (8.2)$$

where ρ is the number density of Eq. (3.3) (*e.g.* Cressie (1993)). Note that the K-function relates to the *pair-correlation* (*or, two body-correlation*) function $g(r)$:

$$g(r) = \frac{1}{2\pi r}\frac{dK(r)}{dr}. \quad (8.3)$$

Then given the observed point pattern, the unbiased estimator $\hat{K}(r)$ of K-function is represented by

$$\hat{K}(r) = \frac{1}{\rho N}\sum_{i=1}^{N} N_i(r), \quad (8.4)$$

where $N_i(r)$ denotes the number of individuals within the circle with centre $\boldsymbol{x}_i$ and radius r. However, since individuals outside the region V are not observed, the estimator $\hat{K}(r)$ is negatively biased. When we calculate the individuals for an individual near the boundary of the region, counts will be low because individuals outside the region are not taken into account (Diggle (2003), Schabenberger and Gotway (2005)). Then to decrease the *edge effects*, in this thesis, a periodic boundary condition is applied, *i.e.* we regard the rectangular regions V as a torus, so that the individuals near the opposite edges are considered to be mutually close. Several methods have been proposed to correct the effects; see, for example, Ripley (1976, 1979b, 1988, 2004), Stein (1991), Geyer (1998) and Baddeley (1999).

When the point pattern is considered as Poisson, $K(r)$ is equal to πr^2. There $K(r) - \pi r^2$ is the deviation from the Poisson pattern. If we transform

$K(r) = \pi r^2$ into $r = \sqrt{K(r)/\pi}$, then the transformed unbiased estimator $\hat{L}(r)$ is defined as

$$\hat{L}(r) = \sqrt{\hat{K}(r)/\pi}. \tag{8.5}$$

Then we can graphically evaluate the type of distribution of a given point pattern by comparing $\hat{L}(r)$ with the line $L(r) = r$.

We compared plot of L-statistics $\hat{L}(r)$ graphically for the observed data with its envelopes of 99 simulated data which were generated by using the values of the posterior mode of the parameters α and σ given in Table 8.10. Plots of observed $\hat{L}(r)$ with its simulated envelopes for every data under both types (i) and (ii) of prior are illustrated in Figs. 8.27(a)-(d) and 8.28(a)-(d), respectively. From Figs. 8.27 and 8.28, although Gulls data indicates a regular in short range, every data set shows inhibitory property. And we find that the L-statistics for observed data are not deviated from the range of upper and lower envelopes of 99 simulated data for all data sets. It seems that the L-statistics is not sensitive to the two priors (i) and (ii). These results indicate the reasonability of our Bayesian estimation of the two parameter Soft-Core models for spatial point patterns.

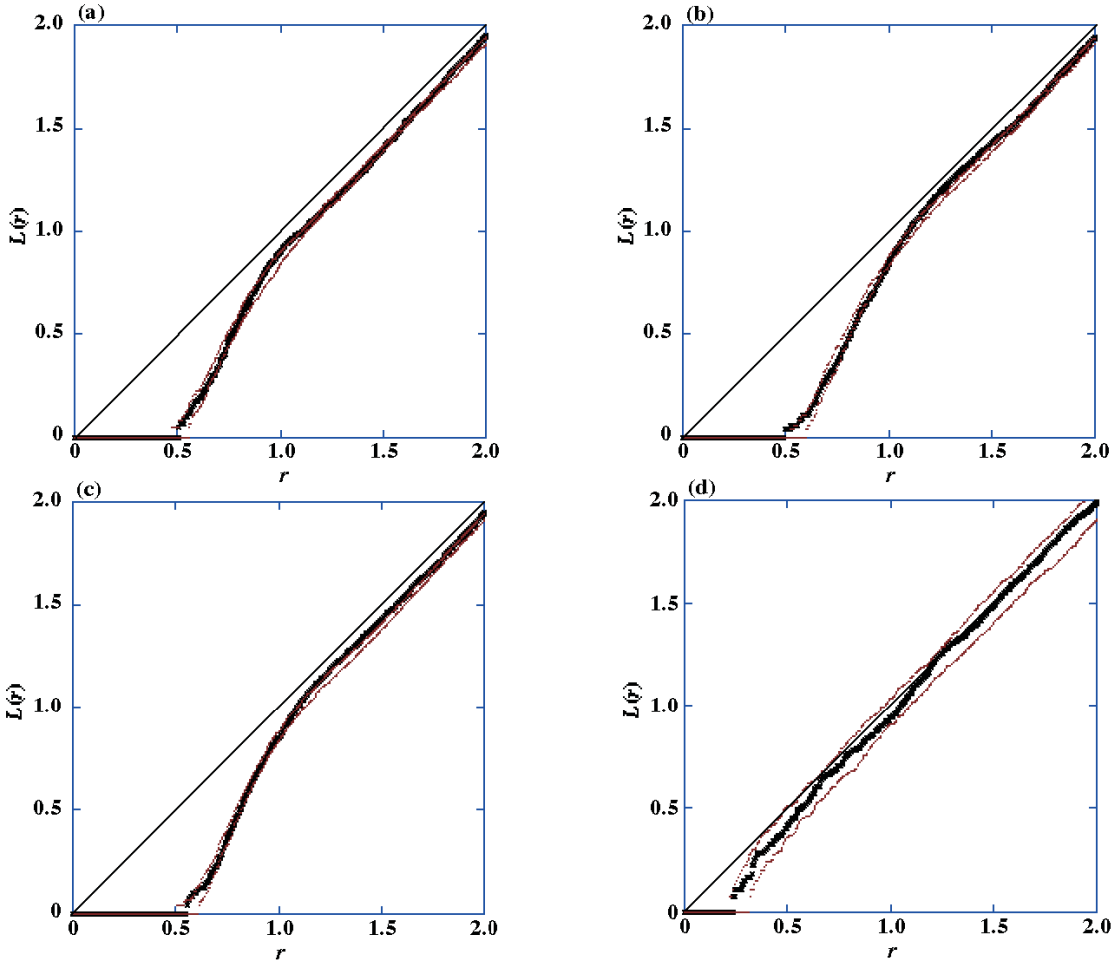

Figure 8.27: *Plots of L-statistics for the data and its upper and lower envelopes from 99 simulated point patterns of the Soft-Core models in the case of the type (i) prior. The symbol* $\times$ *indicates the values of the L-statistics of the raw data, and the symbol* $\cdot$ *represents the values of the L-statistics of the upper and lower envelopes of simulations. The 45° line corresponds to the Poisson model. The unit of length is equal to* $\sqrt{V/N}$ *in the graph, respectively: (a) Balls; (b) P6T13; (c) M6T10; (d) Gulls.*

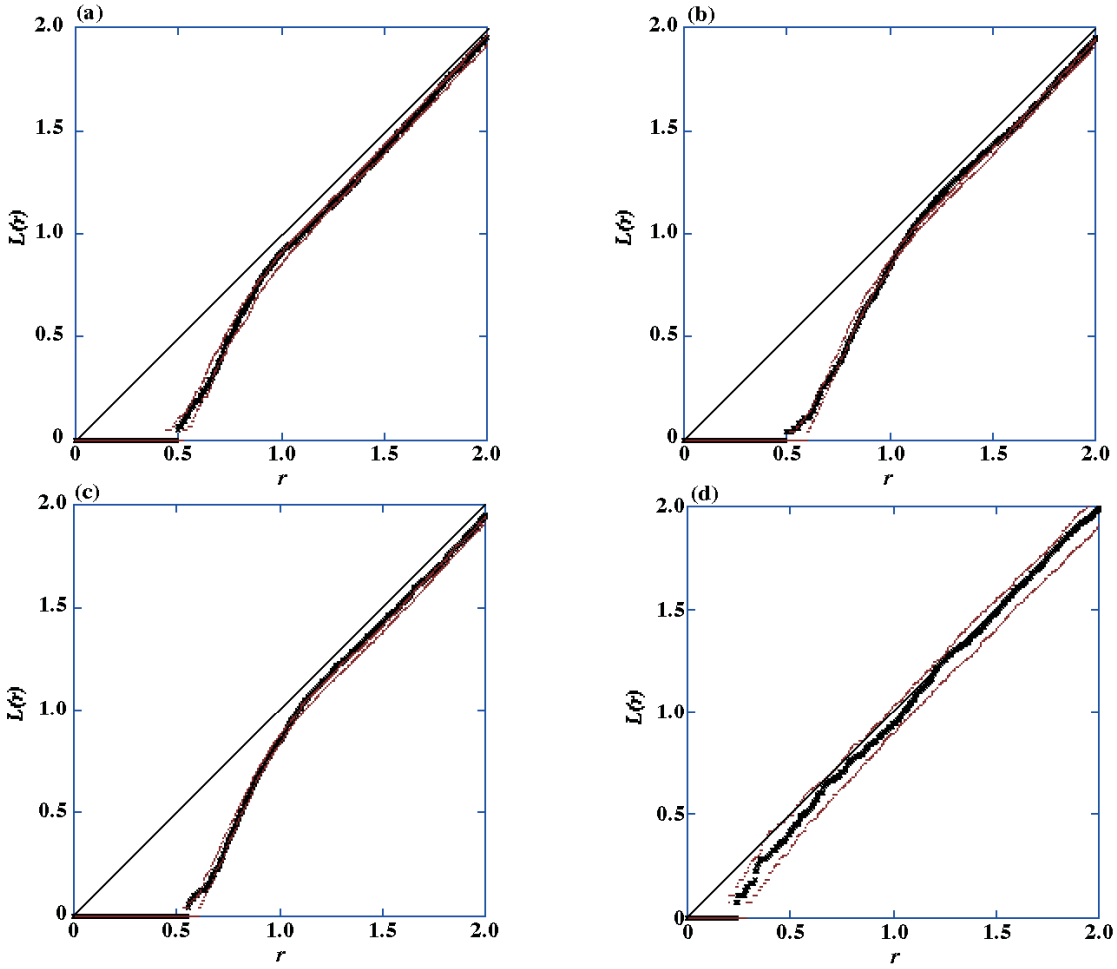

Figure 8.28: *Plots of L-statistics for the data and its upper and lower envelopes from 99 simulated point patterns of the Soft-Core models in the case of the type (ii) prior. The symbol* $\times$ *indicates the values of the L-statistics of the raw data, and the symbol* $\cdot$ *represents the values of the L-statistics of the upper and lower envelopes of simulations. The 45° line corresponds to the Poisson model. The unit of length is equal to* $\sqrt{V/N}$ *in the graph, respectively: (a) Balls; (b) P6T13; (c) M6T10; (d) Gulls.*

8.6 Discussions

Data analysis of the spatial point patterns for Balls, P6T13, M6T10, and Gulls are briefly summarized as follows. First of all, regarding the Balls data, the posterior mode values of σ and α are about 0.11 and 0.36, respectively (see Table 8.10). It means that the estimated interaction range of charged steel ball is about 1.1 mm, which is about twice as long as the diameter of steel ball, and that the estimated softness of the Soft-Core potential is about $n = 5.6$. These results can be considered to be reasonable as a consequence of complex interactions between steel balls under the existence of charged wall of the plastic box.

Next, about blue cones in a macaque retina, P6T13 and M6T10, from Table 8.10, the values of posterior mode of the parameters $(\alpha, n, \sigma, \tau)$ are about $(0.40, 5.0, 3.8, 0.56)$ and $(0.36, 5.6, 3.7, 0.56)$, respectively. These values are similar in both P6T13 and M6T10; since the two data sets were respectively sampled from near areas in the center of the retina, the above results are quite natural. These results indicate that since the blue cones represent a relatively small fraction of the total cone population, the blue cones regularly distribute in the fovea in order the photoreceptors work efficiently (see Figs 8.27 and 8.28). It will be interesting to analyze a pattern of other type of cones or a pattern of cones in other primates.

Finally, for the pattern of Gulls, the nests of gulls result in a certain spacing out one another because of territoriality of gulls. From Table 8.10, the posterior mode of σ, that is, the mean interval of neighboring nests is about 2.2 meters. Since the body size of a Gray Gull is a few tens of centimeters, their territory is not crowded with one another (the posterior mode of $\tau \sim 0.05$). This agrees with our result that the reduced density of the nesting pattern of Gulls is rather small.

In order to investigate the sensitivity of prior and jumping densities, we performed our Bayesian procedure under the uniform priors with normal jumpings and the normal priors with uniform jumpings by crossing the combination of prior and jumping densities, as remarked in §6.1.3 and 6.2. Then marginal posterior densities, modes, means and convergence of respective parameters were similar to the results of the case of the previous combinations of prior and jumping density except for the Gulls data set. As stated in §6.3, for large sample point patterns ($N = 500$) the influence of the choice of priors upon posterior is not too strong. On the other hand, for relatively small point patterns, we investigated the case of $N = 100$. Then, we obtained the similar results between the case of $N = 100$ and the Gulls data, as described in §6.3 and §8.4. For the case where the number of points N and the reduced density τ are relatively small, a bigger number of iteration is needed and the influence of the choice of priors on posterior is strong for α.

Chapter 9

Concluding remarks

In this thesis, we have considered Bayesian estimation of the two-parameter Soft-Core interaction potential models between individuals for various regular point patterns by using MCMC methods.

For a given point pattern in equilibrium under a certain repulsive interaction potential in a finite two-dimensional region, such an equilibrium point pattern is statistically represented by the Gibbs distribution of Eq. (2.2). The likelihood of parameters which characterizes the interaction potentials can be described by the Gibbs distribution. However, since the form of the normalizing factor (2.3) of the Gibbs distribution is a high multiplicity of integral, it is very difficult to obtain the likelihood in general. For this reason, Bayesian analysis for these spatial point patterns has been scarcely studied. If we make use of the effective approximate log-likelihood for the Soft-Core models (Ogata and Tanemura (1989)), then we can consider our Bayesian estimation procedure of parameterized repulsive interactions for various regular point patterns. The main contribution of the thesis is to develop, through MCMC methods, a method to obtain the marginal posterior densities of the parameters for the Soft-Core models for given point patterns. There, the useful approximate log-likelihood (Ogata and Tanemura (1989)) plays an important part in the Metropolis-Hastings algorithm. Then two types of prior densities (type (i) and (ii) prior) of the parameters are applied and jumping densities with similar type as prior density are applied in Markov chain simulations.

First, we have presented our results of Metropolis-Hastings iterative simulations for various simulated equilibrium point patterns, which were generated from MCMC of the Soft-Core models for the cases of large and small size. As the results, when the number of points N is large ($N = 500$), the coincidence between the shape of each marginal posterior density of α and τ (*i.e.* σ) seems to be good regardless of combinations of the prior and jumping densities. When the number of points N is relatively small ($N = 100$), the coincidence between the shape of marginal posterior densities of τ is said to be good for

all combinations of the prior and jumping densities. On the other hand, for the parameter α, for the cases of relatively large reduced density, the marginal posteriors of α under both types (i) and (ii) of prior with each jumping density seem to be similar in all data sets. On the contrary, for the cases of relatively small reduced density, the marginal posteriors of α between the type (i) and (ii) priors show different spread. Then, it seems that priors have strong influence on posterior for small sample point patterns (see §6.3, 8.4 and 8.6).

Next, we have shown the application of our Bayesian procedure to four real data sets. In order to obtain posterior inference for real data sets, we considered parametric fitting of the generalized gamma distribution to the simulated marginal posterior densities for the parameters. The validity of our procedure and the model evaluation were examined graphically using L-statistics. Then reasonable results were obtained, and our Bayesian procedure was shown to be applicable to a wide class of regular point patterns (see §8.4 and 8.6).

In addition, we also investigated and confirmed MCMC convergence thoroughly. Then we evaluated the stopping and the burn-in time of each single long run by calculating the diagnostics quantity based on independent multiple short runs with various starting points (Gelman and Rubin (1992), Cowles and Carlin (1996), Gelman *et al.* (2004)) (see §5.3 and 8.2).

In our Bayesian procedure, it is essential to use the approximated log-likelihood for the Soft-Core models given in Eq. (3.14). In our future work, we will consider the extension of the effective range of the approximated log-likelihood. Then we are planning to prepare a further paper of investigations of our Bayesian procedure based on simulated equilibrium point patterns and a wider class of regular point patterns. Various prior densities of the parameters will be also considered, then we will discuss as a problem of sensitivity to the prior density.

Acknowledgements

I am very grateful to Professor M. Tanemura of the Graduate University for Advanced Studies for his valuable advice and encouragement for continuing my study and research.

I also thank Professor Y. Ogata of the Graduate University for Advanced Studies for his useful advice. I learned a lot of things by attending the seminar.

I would like to thank Professor Y. Itoh of the Graduate University for Advanced Studies for his grateful comments and advice.

I wish to thank Professor Y. Tamura of the Graduate University for Advanced Studies for his kind advice and encouragement.

I feel gratitude to Professor S. Mase of the University of Tokyo Institute of Technology for his useful comments.

I would like to thank Professor Y. Iba of the Graduate University for Advanced Studies for giving me valuable comments and suggestions for the improvement of my study.

I would like to express my gratitude to Mr. K. Katsura of the Institute of Statistical Mathematics. He kindly taught me many suggestions for my computational work.

I also thank the members in the student room of the Institute of Statistical Mathematics.

Finally, I appreciate my family who supported me through the whole process of a doctor course.

Bibliography

[1] Ahlberg, J. H., Nilson, E. N. and Walsh, J. L. (1967). *The Theory of Splines and Their Applications*. Academic Press, New York.

[2] Asmussen, S., Glynn, P.W. and Thorisson, H. (1992). Stationary detection in the initial transient problem, *ACM Trans. Modelling and Computer Simulations*, 2: 130–157.

[3] Baddeley, A. J. (1999). Spatial sampling and censoring. In *Stochastic Geometry: Likelihood and Computation*, ed. O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. Van Lieshout. Chapman & Hall, London.

[4] Besag, J. (1977). Contribution to the discussion on Dr Ripley's paper, *J. Roy. Statist. Soc., Ser. B*, **39**, 193–194.

[5] Brooks, S. P. (1998). Markov chain Monte Carlo method and its application, *The Statistician*, **47**, 69–100.

[6] Brooks, S. P. and Roberts, G. O. (1998). Diagnostic convergence of Markov chain Monte Carlo algorithms, *Statist. Compt.*, **8**, 319–335.

[7] Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Wiley Classics, New York.

[8] Bray, I. and Wright, D. E. (1998). Application of Markov chain Monte Carlo methods to modelling birth prevalence of Down syndrome, *Appl. Statist.*, **47**, Part 4, 589–602.

[9] Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review, *J. Amer. Statist. Assoc., Ser. B*, **91**, 883–904.

[10] Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Willey & Sons, New York.

[11] David, F. N. and Moore, P. G. (1954). Notes on contagious distributions in plant populations, *Ann. Bot. Lond. N. S.*, **18**, 47–53.

[12] Diggle, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns, *Biometrics*, **35**, 87–101.

[13] Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes, *Internat. Statist. Rev.*, **62**, 99–117.

[14] Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd edn. Arnold, London.

[15] Gamerman, D. (1997). *Markov Chain Monte Carlo Stochastic Simulation for Baysian Inference*. Chapman & Hall, London.

[16] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85**, 398–409.

[17] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion), *Statist. Sci.*, **7**, 457–511.

[18] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd edn. Chapman & Hall, London.

[19] Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities in *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Inference*, seattle, April 22–24, (Ed. E.M. Keramidas), 571–578, Fairfax: Inference Foundation of North America, Inc.

[20] Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." in *Bayesian Statistics 4*, eds. Bernardo, J. M., Berger, J., Dawid, A. P. and Smith, A. F. M., Oxford, U.K.: Oxford University Press, pp. 169–193.

[21] Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion), *Statist. Sci.*, **7**, 473–511.

[22] Geyer, C. J. (1998). Likelihood inference for spatial point processes. In O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (eds), *Stochastic Geometry, Likelihood and Computation*, Monogr. Statist. Appl. Probab. 80, pp. 79–140. Boca Raton, FL; Chapman & Hall, London.

[23] Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. In *Markov chain Monte Carlo in practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter). Chapman & Hall, London.

[24] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97–109.

[25] Heikkinen, J. and Penttinen, A. (1999). Bayesian smoothing in the estimation of the pair potential function of Gibbs point process, *Bernoulli*, **5**(6), 1119–1136.

[26] Hopkins, B. and Skellam, J. G. (1954). A new method of determining the type of distribution of plant individuals, *Ann. Bot. Lond. N. S.*, **18**, 213–227.

[27] Howell, T. R., Araya, B. and Millie, W. R. (1974). Breeding biology of the Gray Gull, *Larus modestus, Univ. Calif. Publ. Zool.*, **104**, 1–57.

[28] Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo, *Statist. Sci.*, **16**, No. 4, 312–334.

[29] Juan, W. T., Huang, Z. H., Hsu, J. W., Lai, Y. J., and Lin, I. (1998). Observation of dust Coulomb clusters in a plasma trap, *Phys. Rev. E*, **58**, R6947.

[30] Kass, R. E., Carlin. B. P., Gelman, A. and Neal, R. M. (1997). MCMC in practice: a roundtable discussion. *Technical Report.* Carnegie Mellon University, Chicago.

[31] Lai, Y. J. and Lin, I. (1999). Packing and defects of strongly coupled two-dimensional Coulomb clusters: Numerical simulation, *Phys. Rev. E*, **60**, R4744.

[32] Lloyd, M. (1967). Mean crowding, *Jour. Animal Ecology.*, **36**, 1–30.

[33] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd edn. John Willey & Sons, New York.

[34] Mase, S., Ogata, Y. and Tanemura, M. (1994). Statistical analysis of mapped point patterns–present condition of theory and application, *Amer. Math. Soc. Transl.*, (2) **161**, 95–108.

[35] Matérn. B. (1960). Spatial variation, *Meddelanden fran Statens Skogsforskningsinstitut*, **49**, No.5, 1–144.

[36] Melzer, A., Trottenberg, T. and Piel, A. (1994). Experimental determination of the charge on dust particles forming Coulomb lattices, *Phys. Lett. A*, **191**, 301.

[37] Mengersen, K. L., Robert, C. P. and Guihenneuc-Jouyaux, C. (1999). "MCMC convergence diagnostics A review," in *Bayesian statistics 6*, eds. Bernardo, J. M., Berger, J. O. and Smith, A. F. M., Oxford, U. K.: Oxford University Press, pp. 415–440.

[38] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21**, 1087–1092.

[39] Møller, J. (2003). *Spatial Statistics and Computational Methods*. Springer-Verlag, New York.

[40] Morisita, M. (1959). Measuring of the dispersion and analysis of distribution patterns, *Mem. Fac. Sci. Kyushu. Univ. Ser. E. Biol.*, **2**, 215–235.

[41] Nitter, T. (1996). *Plasma Sources Sci. Technol.*, **5**, 93.

[42] Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns, *Ann. Inst. Statist. Math.*, **40**, 29–39.

[43] Ogata, Y. and Tanemura, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure, *Ann. Inst. Statist. Math.*, **33B**, 315–338.

[44] Ogata, Y. and Tanemura, M. (1984). Likelihood analysis of spatial point patterns, *J. Roy. Statist. Soc., Ser. B*, **46**, 496–518.

[45] Ogata, Y. and Tanemura, M. (1989). Likelihood estimation of soft-core interaction potentials for Gibbsian point patterns, *Ann. Inst. Statist. Math.*, **41**, 583–600.

[46] O'Hagan, A. and Forster, J. (2004) Kendall's Advanced Theory of Statistics, volume 2B: *Bayesian Inference*, 2nd edn. Arnold, London.

[47] Okabe, M. and Tanemura, M. (2006). Baysian estimation of Soft-Core potential models for spatial point patterns, *J. Japan Statist. Soc.*, **36**, 121–147.

[48] Raftery, A. E. and Lewis, S. (1992). "How many iterations in the Gibbs sampler?," in *Bayesian statistics 4*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., Oxford, U. K.: Oxford University Press, pp.763–773.

[49] Ripley, B. D. (1976). The second-order analysis of stationary point processes, *J. Appl. Probab.*, **13**, 255–266.

[50] Ripley, B. D. (1977). Modelling spatial patterns (with discussion), *J. Roy. Statist. Soc. Ser. B*, **39**, 172–212.

[51] Ripley, B. D. (1979a). Simulating spatial patterns: Dependent samples from a multivariate density, *Appl. Statist.*, **28**, 109–112.

[52] Ripley, B. D. (1979b). Test of "randomness" for spatial point patterns, *J. Roy. Statist. Soc. Ser. B*, **41**, 368–374.

[53] Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*, Cambridge: Cambridge University Press.

[54] Ripley, B. D. (2004). *Spatial Statistics*. John Willey & Sons, New York.

[55] Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

[56] Robert, C. P. (2004). *The Baysian Choice*, 2nd edn. Springer-Verlag, New York.

[57] Roberts, G. O. (1992). "Convergence diagnostics of the Gibbs sampler," in *Bayesian statistics 4*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., Oxford, U. K.: Oxford University Press, pp. 775–782.

[58] Roberts, G. and Tweedie, R. (2001). *Understanding MCMC*. Springer-Verlag, New York.

[59] Ruelle, D. (1977). *Statistical Mechanics*, 3rd printing with corrections, W. A. Benjamin, Reading, Massachusetts.

[60] Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall, London.

[61] Shapiro, M. B., Schein, S. J. and Monasterio, F. M. (1985). Regularity and structure of the spatial pattern of blue cones of macaque retina, *J. Amer. Statist. Assoc.*, **80**, 803–814.

[62] Stein, M. L. (1991). A new class of estimators for the reduced second moment measure of point processes, *Biometrika*, **78**, 281–286.

[63] Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic geometry and its Applications*, 2nd edn. John Willey & Sons, New York.

[64] Tanemura, M. (2003). Statistical distributions of Poisson Voronoi cells in two and three dimensions, *Forma*, **18**, 221–247.

[65] Thomas Jr., E. and Watson, M. (1999). *Phys. Plasmas*, **6**, 4111.

[66] Yu, B. and Mykland, P. A. (1998). Looking at Markov samplers through cusum path plots: a simple diagnostic idea, *Statist. Comput.*, **8**(3), 275–286.