Chapter 2
Performance Evaluation of Gibbs Sampling for Bayesian Extracting Sinusoids

M. Cevri and D. Üstündag

Abstract This chapter involves problems of estimating parameters of sinusoids from white noisy data by using Gibbs sampling (GS) in a Bayesian inferential framework which allows us to incorporate prior knowledge about the nature of sinusoidal data into the model. Modifications of its algorithm is tested on data generated from synthetic signals and its performance is compared with conventional estimators such as Maximum Likelihood (ML) and Discrete Fourier Transform (DFT) under a variety of signal to noise ratio (SNR) conditions and different lengths of data sampling (N), regarding to Cramér–Rao lower bound (CRLB) that is a limit on the best possible performance achievable by an unbiased estimator given a dataset. All simulation results show its effectiveness in frequency and amplitude estimation of noisy sinusoids.

Keywords Bayesian inference · Parameter estimation · Gibbs sampling · Cramér–Rao lower bound and Power spectral density

2.1 Introduction

The sinusoidal frequency model embedded in noise is extensively important because of its wide applicability in many areas of science and engineering such as, modeling and manipulation of time-series from speech, audio to radar, seismology, nuclear magnetic resonance, communication problems and underwater acoustics [28].

We therefore address here a problem of estimating parameters of noisy sinusoids within a Bayesian inferential framework that provides a rigorous mathematical foundation for making inferences about them and a basis for quantifying uncertainties in their estimates. Under an assumption that a number of sinusoids is known a priori,

M. Cevri (✉)
Faculty of Science, Department of Mathematics, Istanbul University, Istanbul, Turkey
e-mail: cevri@istanbul.edu.tr

D. Üstündag
Faculty of Science and Letters, Department of Mathematics, Marmara University, Istanbul, Turkey
e-mail: dustundag@marmara.edu.tr

N. Mastorakis, V. Mladenov (eds.), *Computational Problems in Engineering,* Lecture Notes in Electrical Engineering 307, DOI 10.1007/978-3-319-03967-1_2,

several algorithms have already been applied to spectral analysis and parameter estimation problems, such as least-square fitting [33], maximum likelihood(ML) [25], discrete Fourier transform (DFT) [29, 8], and periodogram [27]. After Jayness' work [21], researchers in different fields of science have given much attention to the relationship between Bayesian inference and parameter estimation. Bretthorst and the others [4, 16, 6, 11, 12, 1, 38, 36, 37, 39] have done excellent works in this area for the last 16 years.

In Bayesian framework, it is necessary to evaluate high dimensional integrals that can be difficult and complex to tackle with. In order to solve these problems, different stochastic sampling algorithms have already been suggested and implemented by the different researches. Therefore, we introduce here one of the stochastic algorithms called Gibbs sampling [11, 12, 7] for recovering sinusoids from noisy data and compare its performance with classical estimators, regarding to Cramér–Rao lower bound (CRLB), that is widely used in statistical signal processing as a benchmark to evaluate unbiased estimators given a dataset [30]. For this purpose, a series of simulation studies with a variation in levels of noise and length of data sampling for a single sinusoid is set up.

The outline of this chapter is as follows. In Sect. 2.2, the harmonic signal models are introduced. In Sect. 2.3, we briefly outline Bayesian data analysis and summarize Gibbs sampling estimator in Sect. 2.4. Cramer–Rao lower bound (CRLB) is introduced in Sect. 2.5. Computer simulation results are given in Sect. 2.6 to evaluate the performance of the Gibbs sampling estimator by comparing with that of classical estimators in different conditions. Finally, conclusions from these simulations are drawn.

2.2 Harmonic Signal Model

In many experiments, a discrete data set $\mathbf{D} = \{d_1, d_2, ..., d_N\}^T$ denoted as an output of a physical system that we want to be modeled is sampled from an unknown function $y(t)$ at discrete times $\{t_1, ..., t_N\}^T$:

$$\begin{aligned} d_i &= y(t_i) \\ &= f(t_i; \theta) + e_i, \quad (i = 1, \ldots, N), \end{aligned} \tag{2.1}$$

where θ is a vector containing parameters that characterize behavior of physical system $f(t;\theta)$ and that are usually unknown. The term e_i is assumed to be drawn from a known random process. The choice of the model function $f(t,\theta)$ depends on the specific application, but we will consider here a superposition of k sinusoids:

$$f(t,\theta) = \sum_{j=1}^{k} a_{c_j} \cos(t\omega_j) + a_{s_j} \sin(t\omega_j), \tag{2.2}$$

where $\{a_{c_j}, a_{s_j}\} \in \mathbb{R}^{2k}$ and $\omega_j \in (0, \pi)$ are amplitudes and angular frequencies, respectively. Hence, Eq. (2.1) can be written in the matrix-vector form:

$$\mathbf{D} = \mathbf{G}\mathbf{a} + \mathbf{e}, \tag{2.3}$$

where $\mathbf{D}$ is $(N \times 1)$ matrix of data points and $\mathbf{e}$ is $(N \times 1)$ matrix of independent identically distributed Gaussian noise samples. $\mathbf{G}$ is $(N \times 2k)$ matrix whose each column is a basis function evaluated at each point of time series. The linear coefficient $\mathbf{a}$ is a $(2k \times 1)$ matrix whose components are arranged in order of coefficients of cosine and sine terms $\{a_{c_1}, a_{s_1}, ..., a_{c_k}, a_{s_k}\}$. Then, the goal of data analysis is usually to infer $\theta = \{(a_{c_j}, a_{s_j}, \omega_j)\}_{j=1}^{k}$ from $\mathbf{D}$ and it is a non-linear optimization, due to frequencies. In signal processing literature, numerous approaches are based on frequentists statistics whereas only a few of them based on Bayesian statistics.

2.3 Bayesian Data Analysis

Bayesian inference can be provided from the product rule of probability calculus which can be originated rigorously starting with the formulation of a small number of desiderata required to define a rational theory of inference as first enunciated by Cox [9], with a more complete treatment given by Jaynes [22]. This formulation directs to the ordinary rules of probability calculus and indicates that every allowed (consistent) theory for inference must be mathematically equivalent to probability theory, or else inconsistent.

By using Bayes' rule [2, 3, 18], the context of the current problem can be expressed as follows:

$$p(\boldsymbol{\theta} \mid \mathbf{D}, I) = \frac{p(\boldsymbol{\theta})\, p(\mathbf{D} \mid \boldsymbol{\theta}, I)}{p(\mathbf{D})}, \tag{2.4}$$

where $p(\boldsymbol{\theta})$ is the prior probability density function (PDF) of the parameter vector $\boldsymbol{\theta}$ that encapsulates our state of knowledge of the parameters before observing $\mathbf{D}$; $p(\mathbf{D} \mid \boldsymbol{\theta}, I)$ is called the likelihood function when considered as a function of $\boldsymbol{\theta}$, but it is known as the sampling distribution when considered as a function of $\mathbf{D}$. $p(\mathbf{D})$ is denoted as an evidence or the marginal likelihood and $p(\boldsymbol{\theta} \mid \mathbf{D}, I)$ is the posterior PDF of the parameters $\boldsymbol{\theta}$ of interest, which summarizes the last information about it:

$$p(\boldsymbol{\theta} \mid \mathbf{D}, I) \propto p(\boldsymbol{\theta})\, p(\mathbf{D} \mid \boldsymbol{\theta}, I). \tag{2.5}$$

It is noted that for parameter estimation, the evidence $p(\mathbf{D})$ is $\boldsymbol{\theta}$-independent because of constant and simply plays role of a normalization factor. To proceed further in the specification of the posterior PDF, we now need to assign functional forms

for $p(\boldsymbol{\theta})$ and $p(\mathbf{D}\,|\,\boldsymbol{\theta}, I)$. After computing $p(\boldsymbol{\theta}\,|\mathbf{D}, I)$, the problem turns out to search a vector $\boldsymbol{\theta}$ that satisfies

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{ p(\boldsymbol{\theta}\,|\mathbf{D}, I) \right\}, \tag{2.6}$$

where Θ is a parameter space.

2.4 Gibbs Sampling

In order to avoid computing the multivariate maximization problem described in Sect. 2.3, an alternative way is the one, proposed by Dou and Hogdson [11, 12], which combines Gibbs sampling (GS) with Bayesian inference theory. Gibbs sampling is an iterative Monte Carlo sampling process [14, 26, 20] and a special case of Metropolis–Hastings sampling [27, 19] wherein the random value is always accepted. It was also used by Geman and Geman [15] in image restoration. Statisticians [35, 13] began to utilize the method for Bayesian computations. It is based on supposing that the target distribution is a posterior probability distribution but, it can be applied to any target distribution, when their full conditional probability distributions are available. We extend here its derivation for multiple frequency signals and briefly summarize it below, but refer to the papers [12, 1] for more detail information.

For linear parameters $\mathbf{a}$ in Eq. (2.3), when σ^2 is known and there is no any specific information about $\{\mathbf{a}_c, \mathbf{a}_s\}$ prior to the observation $\mathbf{D}$, then Eq. (2.5) turns out to be the following form:

$$p(\mathbf{a}_c, \mathbf{a}_s \,|\mathbf{D}, \sigma^2, \omega, I) \propto p(\mathbf{D}\,|\omega, \mathbf{a}_c, \mathbf{a}_s, \sigma^2, I), \tag{2.7}$$

where $p(\mathbf{a}_c, \mathbf{a}_s) \propto$ constant as an uninformative uniform prior PDF for $\{\mathbf{a}_c, \mathbf{a}_s\}$. The marginal posterior distribution of $\mathbf{a}$ given $\boldsymbol{\omega}$ and $\mathbf{D}$ becomes a multivariate normal distribution $\mathcal{N}_m(\hat{\mathbf{a}}, \sigma^2(\mathbf{G}^T\mathbf{G})^{-1})$ [11, 12]:

$$p(\mathbf{a}\,|\,\mathbf{D}, \omega, \sigma^2) = \frac{|\,\mathbf{G}^T\mathbf{G}\,|^{1/2}}{(\sqrt{2\pi}\sigma)^m} e^{-\frac{1}{2\sigma^2}(\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{G}^T\mathbf{G}(\mathbf{a}-\hat{\mathbf{a}})} \tag{2.8}$$

where $\hat{\mathbf{a}}$ is best estimate for $\mathbf{a}$ and $p(\mathbf{a}\,|\,\mathbf{D}, \sigma^2)$ is maximized at $\hat{\mathbf{a}}$. When the variance σ^2 is unknown, by using Jeffreys prior

$$p(\sigma^2) = \frac{1}{\sigma^2} \tag{2.9}$$

and integrating this joint posterior PDF in (2.7) with respect to σ^2,

$$p(\mathbf{a}_c, \mathbf{a}_s \,|\mathbf{D}, \omega, I) \propto \int_0^\infty p(\mathbf{D}|\omega, \mathbf{a}_c, \mathbf{a}_s, \sigma^2)\, p(\sigma^2) d\sigma^2$$
$$\propto \int_0^\infty \frac{p(\mathbf{D}|\omega, \mathbf{a}_c, \mathbf{a}_s, \sigma^2)}{\sigma^2} d\sigma^2, \tag{2.10}$$

the marginal posterior distribution of $\mathbf{a}$ given $\boldsymbol{\omega}$ and $\mathbf{D}$ in Eq. (2.8) turns into the multivariate Student's t distribution $t_m(\hat{\mathbf{a}}, s^2(\mathbf{G}^T\mathbf{G})^{-1}, \nu)$ [11, 12]:

$$p(\mathbf{a} \mid \mathbf{D}, \omega) = \frac{\Gamma[(\nu+m)/2]\,|\,\mathbf{G}^T\mathbf{G}\,|^{1/2}\; s^{-m}}{\left[\Gamma(1/2)^m \Gamma(\nu/2)(\sqrt{\nu})^m\right.} \left[1 + \frac{(\mathbf{a}-\hat{\mathbf{a}})^T \mathbf{G}^T\mathbf{G}(\mathbf{a}-\hat{\mathbf{a}})}{\nu s^2}\right]^{-(\nu+m)/2} \tag{2.11}$$

where $\nu = N - m$ is degrees of freedom and $s^2 = \frac{1}{\nu}(\mathbf{D}-\hat{\mathbf{D}})^T(\mathbf{D}-\hat{\mathbf{D}})$ is sampling variance.

Suppose that a_{c_j} is the only unknown parameter among the others $\{\mathbf{a}_{c_{-j}}, \mathbf{a}_s, \boldsymbol{\omega}\}$ where $\mathbf{a}_{c_{-j}} = \{a_{c_1}, ..., a_{c_{j-1}}, a_{c_{j+1}}, ..., a_{c_k}\}$. Under the assumption of known distribution of the noise, Eq. (2.8) for the conditional PDF of a_{c_j} given that $\mathbf{a}_{c_{-j}}, \mathbf{a}_s, \boldsymbol{\omega}, \mathbf{D}$ and σ^2 have already been known becomes a univariate Gaussian distribution:

$$p(a_{c_j} \left| \mathbf{a}_{c_{-j}}, \mathbf{a}_s, \omega, \mathbf{D}, \sigma^2 \right.) \propto \mathcal{N}\left(\hat{a}_{c_j}, \sigma^2 (\mathbf{X}_{a_{c_j}}{}^T \mathbf{X}_{a_{c_j}})^{-1}\right), \tag{2.12}$$

where

$$\hat{a}_{c_j} = \frac{\hat{\mathbf{D}}^{(1)} \mathbf{X}_{a_{c_j}}}{\mathbf{X}_{a_{c_j}}^T \mathbf{X}_{a_{c_j}}}, \quad \mathbf{X}_{a_{c_j}} = \begin{bmatrix} \cos(\omega_j t_1) \\ \vdots \\ \cos(\omega_j t_N) \end{bmatrix} \tag{2.13}$$

and

$$\hat{\mathbf{D}}^{(1)} = \left\{ d_1^{(1)}, d_2^{(1)}, ..., d_N^{(1)} \right\} \tag{2.14}$$

whose components are defined by $d_i - \sum_{l=1}^{k} a_{c_l} \cos(\omega_l t_i)\delta_{lj} + a_{s_l} \sin(\omega_l t_i)$, $(i = 1, 2, 3..., N)$. The $\delta_{lj} = \begin{cases} 1 & l \neq j \\ 0 & l = j \end{cases}$ helps to eliminate the contribution, which comes from the cosine term of the j th sinusoid. When σ^2 is unknown, Eq. (2.12) becomes a univariate Student's t distribution:

$$p(a_{c_j} \left| \mathbf{a}_{c_{-j}}, \mathbf{a}_s, \omega, \mathbf{D}, \sigma^2 \right.) \propto \mathfrak{t}\left(\hat{a}_{c_j}, s^2_{a_{c_j}} (\mathbf{X}_{a_{c_j}}{}^T \mathbf{X}_{a_{c_j}})^{-1}, N-1\right), \tag{2.15}$$

with

$$s^2_{a_{c_j}} = \frac{1}{N-1}(\hat{\mathbf{D}}^{(1)} - \hat{a}_{c_j}\mathbf{X}_{a_{c_j}})^T(\hat{\mathbf{D}}^{(1)} - \hat{a}_{c_j}\mathbf{X}_{a_{c_j}}) \tag{2.16}$$

When $\left\{\mathbf{a}_c, \mathbf{a}_{s_{-j}}, \boldsymbol{\omega}\right\}$ is given, in a similar way, the conditional PDF of a_{s_j} given that $\mathbf{a}_c, \mathbf{a}_{s_{-j}}, \boldsymbol{\omega}, \mathbf{D}$ and σ^2 have already been known is

$$p(a_{s_j}\left|\mathbf{a}_c, \mathbf{a}_{s_{-j}}, \omega, \mathbf{D}, \sigma^2\right.) \propto \mathcal{N}\left(\hat{a}_{s_j}, \sigma^2(\mathbf{X}_{a_{s_j}}{}^T\mathbf{X}_{a_{s_j}})^{-1}\right), \tag{2.17}$$

where

$$\hat{a}_{s_j} = \frac{\hat{\mathbf{D}}^{(2)}\mathbf{X}_{a_{s_j}}}{\mathbf{X}^T_{a_{s_j}}\mathbf{X}_{a_{s_j}}}, \quad \mathbf{X}_{a_{s_j}} = \begin{bmatrix} \sin(\omega_j t_1) \\ \vdots \\ \sin(\omega_j t_N) \end{bmatrix} \tag{2.18}$$

and

$$\hat{\mathbf{D}}^{(2)} = \left\{d_1^{(2)}, d_2^{(2)}, ..., d_N^{(2)}\right\}, \tag{2.19}$$

whose components are defined by $\hat{d}_i^{(2)} = d_i - \sum_{l=1}^{k} a_{c_l}\cos(\omega_l t_i) + a_{s_l}\sin(\omega_l t_i)\delta_{lj}$, $(i = 1, ..., N)$.

When σ^2 is unknown, Eq. (2.17) turns out to be

$$p(a_{s_j}\left|\mathbf{a}_{s_{-j}}, \mathbf{a}_c, \omega, \mathbf{D}, \sigma^2\right.) \propto t(\hat{a}_{s_j}, s^2_{a_{s_j}}(\mathbf{X}_{a_{s_j}}{}^T\mathbf{X}_{a_{s_j}})^{-1}, N-1) \tag{2.20}$$

with

$$s^2_{a_{s_j}} = \frac{1}{N-1}(\hat{\mathbf{D}}^{(2)} - \hat{a}_{s_j}\mathbf{X}_{a_{s_j}})^T(\hat{\mathbf{D}}^{(2)} - \hat{a}_{s_j}\mathbf{X}_{a_{s_j}}). \tag{2.21}$$

To be able to use the theory of GS for the nonlinear parameter $\boldsymbol{\omega}$, we need to introduce some reasonable approximations to linearize the nonlinear model function $f(t_i, \boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$ under the condition of the known amplitudes $\{\mathbf{a}_c, \mathbf{a}_s\}$. This can be done by expanding it around $\hat{\boldsymbol{\omega}}$ in a region where the posterior PDF is concentrated:

$$\begin{aligned} f(t_i, \hat{\boldsymbol{\omega}}) \cong & \sum_{l=1}^{k} a_{c_l}\cos(\hat{\omega}_l t_i) + a_{s_l}\sin(\hat{\omega}_l t_i) \\ & + \left(-a_{c_j} t_i \sin(\hat{\omega}_j t_i) + a_{s_j} t_i \cos(\hat{\omega}_j t_i)\right)(\omega_j - \hat{\omega}_j), \end{aligned} \tag{2.22}$$

where $\hat{\omega}_j = \arg\min_{\hat{\omega}\in\boldsymbol{\omega}} \sum_{i=1}^{N} (d_i - f(t_i, \boldsymbol{\omega}))^2$ and $\hat{\boldsymbol{\omega}} = \{\omega_1, ..., \omega_{j-1}, \hat{\omega}_j, \omega_{j+1}, ..., \omega_k\}$. Thus, the conditional PDF of ω_j given that $\boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}$ and σ^2 have already been known is a univariate Gaussian distribution:

$$p(\omega_j \,|\, \boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}, \sigma^2) \propto \mathcal{N}(\hat{\omega}_j, \sigma^2 (\mathbf{X}_{\omega_j}{}^T \mathbf{X}_{\omega_j})^{-1}), \tag{2.23}$$

where

$$\mathbf{X}_{\omega_j} = \begin{bmatrix} -a_{c_j} t_1 \sin(\hat{\omega}_j t_1) + a_{s_j} t_1 \sin(\hat{\omega}_j t_1) \\ \vdots \\ -a_{c_j} t_N \sin(\hat{\omega}_j t_N) + a_{s_j} t_N \sin(\hat{\omega}_j t_N) \end{bmatrix}. \tag{2.24}$$

If σ^2 is unknown, Eq. (2.23) becomes is a univariate Student's t distribution

$$p(\omega_j \,|\, \boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}, \sigma^2) \propto t\left(\hat{\omega}_j, s_{\omega_j}^2 (\mathbf{X}_{\omega_j}{}^T \mathbf{X}_{\omega_j})^{-1}, N-1\right). \tag{2.25}$$

with

$$s_{\omega_j}^2 = \frac{1}{N-1} (\mathbf{D} - \hat{\mathbf{D}})^T (\mathbf{D} - \hat{\mathbf{D}}), \tag{2.26}$$

where $\hat{\mathbf{D}} = \{\hat{d}(t_1), \hat{d}(t_2), ..., \hat{d}(t_N)\}$ whose components are defined by $\hat{d}(t_i) = \sum_{l=1}^{k} a_{c_l} \cos(\hat{\omega}_l t_i) + a_{s_l} \sin(\hat{\omega}_l t_i)$

A systematic form of GS algorithm [11, 12, 10] contains choosing initially arbitrary starting values $\{\mathbf{a}_c^{(0)}, \mathbf{a}_s^{(0)}, \boldsymbol{\omega}^{(0)}\}$ and drawing successively random samples from the full conditional distributions:

$$\begin{aligned} a_{c_j}^{(1)} &\sim p(a_{c_j} \,|\, \{a_{c_1}^{(1)}, ..., a_{c_{j-1}}^{(1)}, a_{c_{j+1}}^{(0)}, ..., a_{c_k}^{(0)}\}, \mathbf{a}_{s_j}^{(0)}, \omega^{(0)}, \mathbf{D}) \\ a_{s_j}^{(1)} &\sim p(a_{s_j} \,|\, \mathbf{a}_{c_j}^{(1)}, \{a_{s_1}^{(1)}, ..., a_{s_{j-1}}^{(1)}, a_{s_{j+1}}^{(0)}, ..., a_{s_k}^{(0)}\}, \omega^{(0)}, \mathbf{D}) \\ \omega_j^{(1)} &\sim p(\omega_j \,|\, \mathbf{a}_{c_j}^{(1)}, \mathbf{a}_{s_j}^{(1)}, \{\omega_1^{(1)}, ..., \omega_{j-1}^{(1)}, \omega_{j+1}^{(0)}, ..., \omega_k^{(0)}\}, \mathbf{D}), (j = 1, ..., k). \end{aligned} \tag{2.27}$$

At each iteration of the Gibbs sampler, we cycle through the set of conditional distributions and draw one sample from each. When a sample is drawn from one conditional distribution, the succeeding distributions are updated with the new value of that sample. At the K'th iteration we obtain the following drawings:

$$
\begin{aligned}
a_{c_j}^{(K+1)} &\sim p(a_{c_j} \left| \{a_{c_1}^{(K+1)}, ..., a_{c_{j-1}}^{(K+1)}, a_{c_{j+1}}^{(K)}, ..., a_{c_k}^{(K)}\}, \mathbf{a}_{s_j}^{(K)}, \omega_j^{(K)}, \mathbf{D}\right.) \\
a_{s_j}^{(K+1)} &\sim p(a_{s_j} \left| \mathbf{a}_{c_j}^{(K+1)}, \{a_{s_1,}^{(K+1)}, ..., a_{s_j-1}^{(K+1)}, a_{s_{j+1}}^{(K)}, ..., a_{s_k}^{(K)}\}, \omega_j^{(K)}, \mathbf{D}\right.) \\
\omega_j^{(K+1)} &\sim p(\omega_j \left| \mathbf{a}_{c_j}^{(K+1)}, \mathbf{a}_{s_j}^{(K+1)}, \{\omega_1^{(K+1)}, ..., \omega_{j-1}^{(K+1)}, \omega_{j+1}^{(K)}, ..., \omega_k^{(K)}\}, \mathbf{D}\right.).
\end{aligned}
\tag{2.28}
$$

For a large enough K, $a_c^{(K+1)}$, $a_s^{(K+1)}$ and $\omega^{(K+1)}$ can be considered as random variables drawn from their posterior PDF distributions. Therefore we are able to generate samples of these posterior PDFs for each parameter. Using these samples, all of the estimates about the their corresponding can then be found, such as the most probable values for them, the mean value, the marginal variances with respect to the most probable value etc. When σ^2 is unknown, we do the same thing as above except that the random numbers are drawn from the Student's t distribution.

2.5 Cramer–Rao Lower Bound

Given an estimation problem, one may ask: What is the variance of the best possible unbiased estimator? The answer is given by the Cramer–Rao lower bound (CRLB) [24, 17], which we will study in this section and it provides a theoretical lower limit for variance of estimator. If we consider the parameter vector θ and the signal to noise ratio (SNR), then the CRLB to the variance of unbiased estimator of the parameters θ for the signal model is determined in the form:

$$
\mathrm{Var}_{\boldsymbol{\theta}}(\hat{\theta}) \geq \mathrm{CRLB}(\theta) = \mathbf{J}^{-1}(\theta), \tag{2.29}
$$

where Fisher information matrix $\mathbf{J}(\theta)$ [24] is defined as an expectation of the second derivatives of the log likelihood function with respect to θ:

$$
\mathbf{J}(\theta) = E\left[-\frac{\partial^2 \ln P(\mathbf{D}|\theta, I)}{\partial \theta^2} \right]. \tag{2.30}
$$

for large N, $\mathbf{J}(\theta)$ is a diagonal matrix and its inversion is straightforward. The diagonal elements of its inversion yield the lower bound on the variance of the estimates asymptotically. When the noise is white Gaussian, we can use the an alternative form of CRLB which is easier than the general case in Eq. (2.30). In this case the Fisher information matrix becomes

$$
\mathbf{J}(\theta) = \frac{1}{\sigma^2} \sum_{j=1}^{N} \frac{\partial f_j(\theta)}{\partial \theta} \left(\frac{\partial f_j(\theta)}{\partial \theta} \right)^T. \tag{2.31}
$$

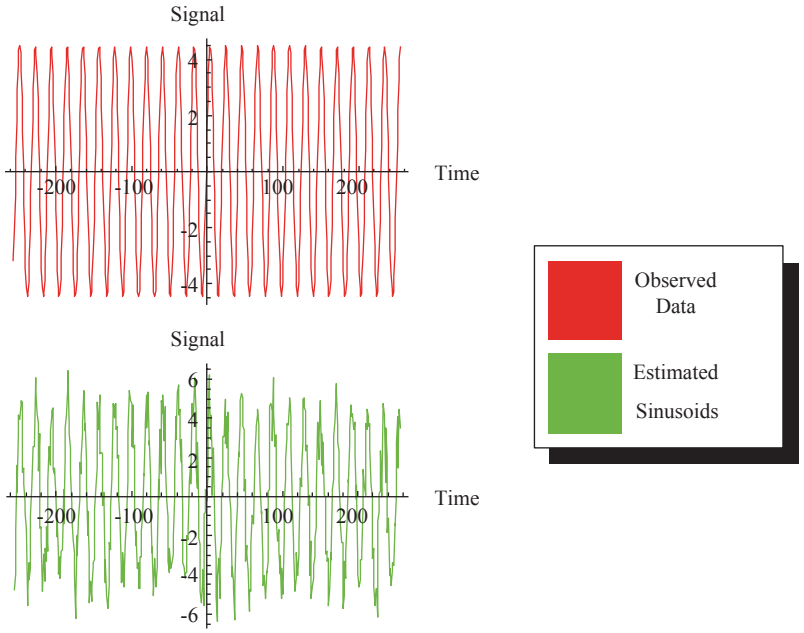

Fig. 2.1 Recovering signal from noisy data produced from a single harmonic frequency signal model

2.6 Computer Simulations

To demonstrate the proposed approach with examples which are used by previous researches [4, 11, 12, 38, 36, 37], we firstly created data samples according to a signal model with a single frequency:

$$d_i = 2\cos(0.3t_i) + 4\sin(0.3t_i) + e_i \quad (i = 1,...,512) \tag{2.32}$$

Here i runs in a symmetric time interval $-T$ and T (2T+1=N) and $e_i \sim N(0,1)$. We obtained noisy data samples ($N = 512$), shown in Fig. 2.1 and carried out Bayesian analysis. The proposed method requires initial values for the parameters to start the iteration. Instead of choosing them randomly from a uniform distribution [16], we first performed a Fast Fourier Transformation (FFT) of the data and then chose the locations of the peaks in the power spectrum density [4, 16], which is a squared magnitude of FFT, as an initial estimate for the frequencies. Once, initial frequencies were obtained, we carried on calculating the coefficients $\mathbf{a}_c$ and $\mathbf{a}_s$ as initial values for the amplitudes, respectively. The algorithm of GS, introduced in the paper was coded in *Mathematica* programming language and run on a workstation in two cases where the standard deviation of noise is known or not. In the case where the deviation of noise is unknown, the output of the computer simulation is illustrated in Table 2.1. The estimated parameter values are quoted

Table 2.1 Computer simulations for a single harmonic frequency model

Parameters	True values	Estimated values
ω_1	0.3	0.2999 ± 0.00009
a_{c_1}	2	2.041 ± 0.0623
a_{s_1}	4	3.992 ± 0.0628

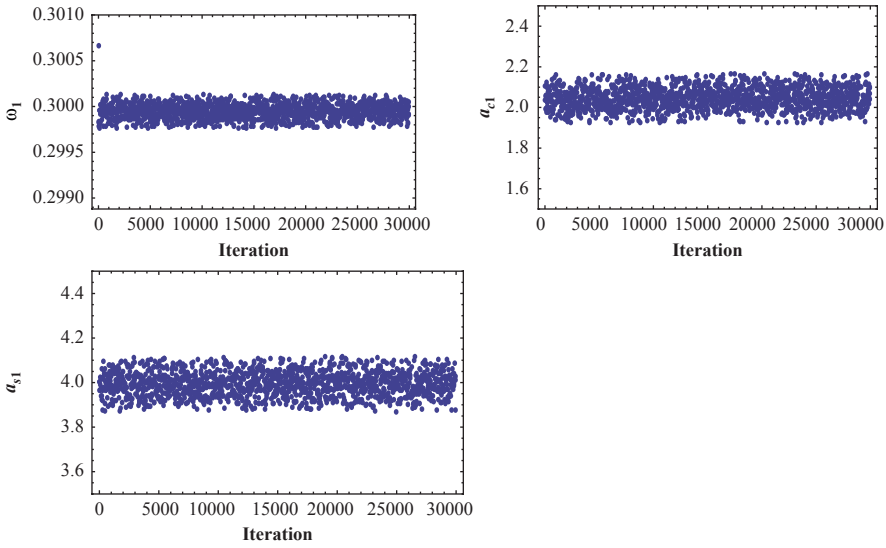

Fig. 2.2 MCMC parameter iterations

as (*value*) ± (*standard deviation*) and used to regenerate the given signal model, shown in Fig. 2.1.

It can be seen that a single frequency and its corresponding amplitudes are recovered very well.

In order to determine its convergence, there are several diagnostic tests [5, 34] we can do, both visual and statistical, to see if the chain appears to be converged. One intuitive and easily implemented diagnostic tool is a trace plot (or history plot) [34] which is a plot of the iteration number against the value of the draw of the parameter. If it has converged, the trace plot will move up and down around the mode of the distribution and the distribution of the parameters settles down to the target posterior PDF from which statistical inferences about the parameters can be made. A clear sign of non-convergence occurs when we observe some trending in the trace plot. In this case we can see whether our chain gets stuck in certain areas of the parameter space, which indicates bad mixing. Figure 2.2 shows the scatter plots of the model parameters, ω_1, a_{c_1} and a_{s_1}, respectively and indicates that the GS samples are densely placed around the estimated values of these parameters.

In our second example, we consider a signal model with two close harmonic frequencies:

http://www.springer.com/978-3-319-03966-4

Computational Problems in Engineering
Mastorakis, N.; Mladenov, V. (Eds.)
2014, XII, 311 p. 177 illus., 135 illus. in color.,
Hardcover
ISBN: 978-3-319-03966-4