Markov chain Monte Carlo methods for Bayesian gravitational radiation data analysis

Nelson Christensen
Department of Physics, The University of Auckland, Auckland, New Zealand

Renate Meyer
Department of Statistics, The University of Auckland, Auckland, New Zealand

(Received 12 February 1998; published 8 September 1998)

The LIGO and VIRGO kilometer length laser interferometric gravitational radiation detectors should observe numerous mergers of compact binary systems. The accurate determination of the binary's signal parameters is a critical task for the observers. Important cosmological information, such as an independent measurement of the Hubble constant, can be derived if an accurate determination of the distance to the event is achieved. A Bayesian approach to the parameter estimation problem has become a popular topic. Unfortunately the multidimensional integrals that are inherent in the calculation of the Bayes estimator can be computationally prohibitive. In this paper we show that computational difficulties can be overcome by using the Gibbs sampler to calculate posterior distributions. The Bayesian approach and its implementation via Markov chain Monte Carlo calculations is illustrated by way of an example involving four parameters.
[S0556-2821(98)07918-1]

PACS number(s): 04.80.Nn, 02.70.Lq, 06.20.Dk

I. INTRODUCTION

A number of collaborations around the world will be operating laser interferometric gravitation radiation antennas within the next few years. In the United States the Laser Interferometric Gravitational Wave Observatory (LIGO) is under construction, with 4 km arm length interferometers in Hanford, Washington, and Livingston, Louisiana [1]. A similar French-Italian detector will be built in Europe (VIRGO) [2]. Coalescing binaries containing neutron stars (NSs) or black holes (BHs) promise to be the cleanest and most promising source of detectable radiation [3]. Ultimately the LIGO-VIRGO network may observe binaries out to a distance of 2 Gpc [4].

The detection of coalescing binary events will provide physicists with extremely useful cosmological information. Initially Schutz [5] noted that a detected signal contains enough information to decipher the absolute distance to the system, and hence the determination of the Hubble constant would be achieved through the observed distribution of several binaries. Subsequent work [6] indicates that the uncertainty in the measured distance can be comparable to the distance itself, but important cosmological tests will still be possible through the observation of numerous mergers [7].

In addition to the cosmological importance, accurate parameter estimation in the observed coalescing binaries will provide a host of information of great physical significance. Observation of the time of tidal disruption of an NS-NS binary system may permit a determination of the NS radii and information on the NS equation of state [8]. The characteristics of radiation in the post-Newtonian regime will provide insight into highly non-linear general relativistic effects [6,9]. The formation of a BH at the end of a NS-NS coalescence, or the merger of two BHs, will produce gravitational radiation as the system decays to a Kerr BH; this is an extremely interesting radiation production regime [9].

Application of Bayes' theorem is well suited to astrophysical observations [10]. The *Bayesian* versus *frequentist* approaches to gravitational radiation data analysis are well presented by Finn [11]. Parameter estimation from the gravity wave signals of coalescing compact binaries provides an important application of Bayesian methods [4,6,12,13]. The inevitable difficulty in calculating the multidimensional integrals necessary for the determination of posterior probability distributions has hindered the further development of Bayesian parameter estimation methods. However, these impediments have been overcome by the progress made within the last decade in Bayesian computational technology via Markov chain Monte Carlo (MCMC) methods (see [14] for an introduction). Since the seminal paper by Geman and Geman [15] in the context of digital image analysis, MCMC methods have already revolutionized many areas of applied statistics and will have an impact on gravitational wave data analysis. MCMC approaches to dynamic modeling represent some of the currently critical research frontiers in Bayesian statistical modeling.

The organization of this paper is as follows. To encourage its use in gravitational radiation data analysis, the Bayesian approach to statistical inference and its implementation via Markov chain Monte Carlo methods will be discussed in Sec. II. In Sec. III we apply the MCMC methods to a four parameter Bayesian estimation; this example is identical to that of Cutler and Flanagan [6] in their application of the Marković approximation [16] towards the calculations of the distance error ΔD. Section IV presents our conclusions.

II. BAYESIAN COMPUTATION VIA MARKOV CHAIN MONTE CARLO METHODS

A. Bayesian inference

In Bayesian data analysis, the model consists of a joint distribution over all unobserved (parameters) and observed (data) quantities, and one conditions on the data to obtain the posterior distribution of the parameters. This Bayesian approach to statistical inference can be integrated in the usual decision-theoretic framework (see e.g. [17]). The starting point is a statistical experiment described by a parametric family of probability distribution functions (PDFs) $\{p(\mathbf{y}|\boldsymbol{\theta});\ \boldsymbol{\theta}\in\Theta\}$ on the sample space $\mathcal{Y}$. The researcher's subjective belief and knowledge about the unknown parameters is comprised in the prior distribution $p(\boldsymbol{\theta})$, $\boldsymbol{\theta}\in\Theta\subseteq\mathbb{R}^n$ which together with the sampling distribution $p(\mathbf{y}|\boldsymbol{\theta})$, the likelihood, specifies the Bayesian model. The loss function $l(\boldsymbol{\theta},a)$ defined on the product of the parameter space Θ and the set of allowable actions A gives the loss incurred when $\boldsymbol{\theta}$ is the true state of nature and a decision for a is made. Certain choices of these correspond to point or interval estimation and hypothesis testing. Given a measurement $\mathbf{y}\in\mathcal{Y}$ and prior belief $p(\boldsymbol{\theta})$, the rationale underlying the Bayesian approach to the selection of a decision rule $d(\mathbf{y})$ is to choose the action $a=d(\mathbf{y})$ that minimizes the posterior risk

$$R(p,d(\mathbf{y}))=E_{\boldsymbol{\theta}|\mathbf{y}}[l(\boldsymbol{\theta},d(\mathbf{y}))]=\int l(\boldsymbol{\theta},d(\mathbf{y}))p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \tag{2.1}$$

since in the light of the data $\mathbf{y}$ the researcher's opinion as to the state of nature is summarized by the posterior distribution of $\boldsymbol{\theta}$ given by the Bayes theorem

$$p(\boldsymbol{\theta}|\mathbf{y})=\frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{m(\mathbf{y})}\propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}) \tag{2.2}$$

where

$$m(\mathbf{y})=\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{2.3}$$

is the marginal density of $\mathbf{y}$, which can be regarded as a normalizing constant as it is independent of $\boldsymbol{\theta}$. In the one-dimensional case under squared error loss, for instance, $l(\theta,a)=(\theta-a)^2$ and the minimum of the risk is attained at the posterior mean $E[\theta|y]$, the ''Bayes estimator'' (see also Appendix A of [6]).

The Bayesian approach in general consists of two conceptually and practically distinct steps:

(1) Constructing a full probability model which consists of a joint probability distribution for all observable and unobservable quantities.
(2) By conditioning on the observed data, calculating the posterior distribution, i.e., the conditional probability distribution of the unobservables of interest, given the observed data.

In the first step, the joint probability density $p(\mathbf{y},\boldsymbol{\theta})$ of the observations $\mathbf{y}$ and the unobservables $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_n)$ can be written as the product of two densities, referred to as the prior density $p(\boldsymbol{\theta})$ and the sampling density or likelihood function $p(\mathbf{y}|\boldsymbol{\theta})$, i.e.

$$p(\mathbf{y},\boldsymbol{\theta})=p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}). \tag{2.4}$$

In the light of the data, our opinion as to the state of nature is then updated to the posterior distribution, Eq. (2.2). One could then marginalize the joint posterior over certain components of $\boldsymbol{\theta}$ to obtain the marginal posterior densities and characteristics of interest, such as the posterior mean, median, or mode of a specific component θ_i.

Thus, the second step, though conceptionally easy, is indeed a formidable problem in general because it requires high-dimensional integration to obtain the normalization constant $m(\mathbf{y})$ and to calculate one-dimensional characteristics.

B. Markov chain Monte Carlo methods

Before the development of MCMC methods there had been essentially three different approaches to handle this multidimensional integration:

(i) Laplace expansion (i.e. normal approximation).
(ii) Numerical integration via Gaussian quadrature techniques.
(iii) Monte Carlo integration via importance sampling, hit-and-miss etc.

The first two methods require a high degree of mathematical sophistication of the data analyst. Laplace expansion relies on large sample asymptotics and the approximations can be very bad in small sample situations. Gaussian quadrature suffers from the curse of dimensionality in that the amount of computation rises exponentially with the number of parameters. Monte Carlo methods substitute the deterministic integration by a statistical estimation problem, that of estimating the mean of a certain multivariate distribution. This can be done by drawing a random sample and estimating the expectation by the sample mean. Although applicable to high-dimensional problems, conventional Monte Carlo methods can be very inefficient in certain situations.

A major breakthrough for the routine implementation of Bayesian inference was the realization that any high-dimensional integration can be performed by using MCMC methods of which the Gibbs sampler (described below) is an important special case.

Instead of generating a sequence of independent samples from the joint posterior, in MCMC methods a Markov chain is constructed, whose equilibrium distribution is just the joint posterior. Thus, after running the Markov chain for a certain ''burn-in'' period, one obtains (correlated) samples from the limiting distribution (provided that the Markov chain has reached convergence). Various methods to assess convergence have been developed (see e.g. [18]) and implemented in CODA [19], a software package for convergence diagnostics and statistical and graphical output analyses for Markov chains. Statistical theory ensures that averaging of a function of interest over realizations from a single run of the chain provides a consistent estimate of its expectation.

The Gibbs sampler is a specific MCMC method where in a cycle we sample from each of the full *conditional* posterior distributions:

$$p(\theta_i|\mathbf{y},\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_n). \tag{2.5}$$

Given an arbitrary set of starting values $\theta_1^{(0)},\ldots,\theta_n^{(0)}$ the algorithm proceeds as follows:

$$\begin{aligned}&\text{simulate}\quad \theta_1^{(1)}\sim p(\theta_1|\mathbf{y},\theta_2^{(0)},\ldots,\theta_n^{(0)})\\ &\text{simulate}\quad \theta_2^{(1)}\sim p(\theta_2|\mathbf{y},\theta_1^{(1)},\theta_3^{(0)},\ldots,\theta_n^{(0)})\\ &\qquad\vdots\\ &\text{simulate}\quad \theta_n^{(1)}\sim p(\theta_n|\mathbf{y},\theta_1^{(1)}\ldots,\theta_{n-1}^{(1)})\end{aligned} \tag{2.6}$$

and yields $\boldsymbol{\theta}^{(m)}=(\theta_1^{(m)},\ldots,\theta_n^{(m)})$ after m such cycles. This defines a Markov chain with transition kernel

$$k(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\theta}^{(m)}) = \prod_{i=1}^{n} p(\theta_i^{(m+1)}|\mathbf{y},\theta_1^{(m+1)},\ldots,\theta_{i-1}^{(m+1)},\theta_{i+1}^{(m)},\ldots,\theta_n^{(m)}), \tag{2.7}$$

which converges to the joint posterior as its equilibrium distribution (see [14]). Consequently, if all the full conditional posterior distributions are available, all that is required is sampling iteratively from these. Thereby, the problem of sampling from an n-variate probability distribution function (PDF) is reduced to sampling from n univariate PDFs.

C. Sampling from non-log-concave densities

In many applications where the prior distribution is conjugate to the likelihood, the full conditionals in fact reduce analytically to closed-form distributions and we can use highly efficient special purpose Monte Carlo methods for generating from these (see e.g. [20]). In non-conjugate Bayesian models the full conditional density of a certain parameter can in general be constructed from those few terms of the joint distribution which depend upon it. This demonstrates the need for fast and efficient black-box methods to sample from an arbitrarily complex full conditional posterior distribution which is known only up to a constant of proportionality for use in each cyclic step of the Gibbs sampler.

At first sight, this seems to be an arduous task without any further assumptions on certain properties of the full conditional distributions that are under consideration. A rich class of distributions, however, is given by the class of those with *log-concave* densities. Examples of log-concave densities are listed in the table of Gilks and Wild [21] or Devroye [20], p. 287.

Various fast and efficient simulation methods for sampling from log-concave distributions have been proposed in the literature: see e.g. [20]. However, these require the location of the mode of the density, thereby necessitating a time-intensive and computer-expensive maximization step. Using the fact that any log-concave density can be bounded from above and below by its tangents and chords, the *adaptive rejection sampling* (ARS) developed by Gilks and Wild [21] was able to dispense with the awkward and time-consuming optimization. It is based on the usual Monte Carlo rejection sampling using squeezing functions (cf. [22]). A further advantage is its ''adaptivity'' in that with each rejection step the bounding function gets closer to the target density and thereby reduces the rejection probability in the subsequent step. To calculate the tangents, the first derivatives of the density are required in the original algorithm. A derivative-free version was developed by [23]. For an efficient rejection sampling algorithm it is also essential that the envelope density be easy to sample from. This is the case in ARS, the envelope being a piecewise exponential density.

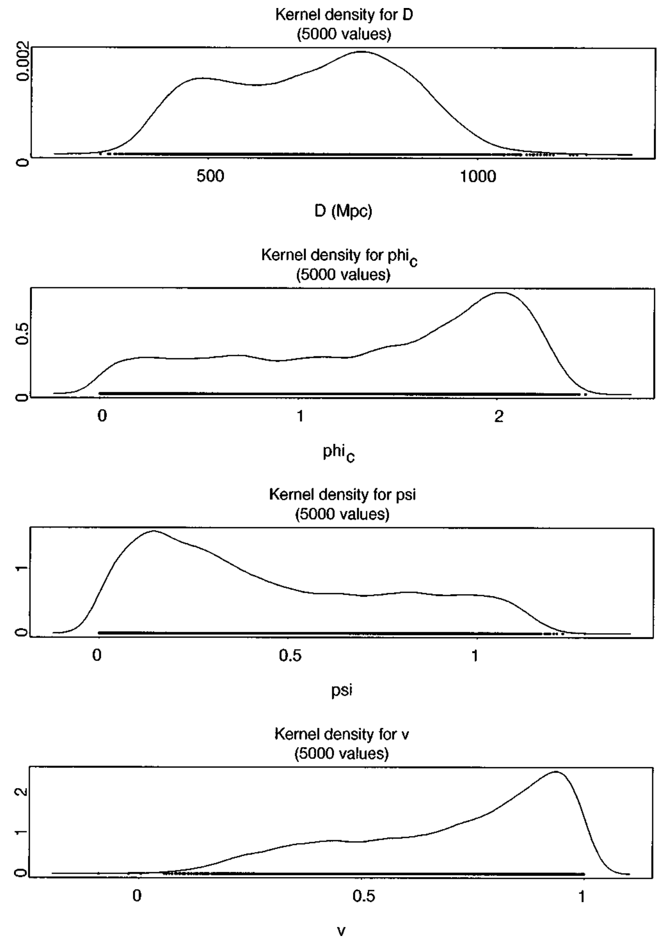

FIG. 1. Kernel density estimates of the marginal posteriors of the variables D, υ, ψ, and ϕ_c assuming the initial ''best-fit'' parameters of $D_0=432$ Mpc, $\upsilon_0=0.31$, $\psi_0=11.5°$, and $\phi_{c0}=114.6°$.

In an attempt to dispose of the crucial requirement of log-concavity, Gilks *et al.* [24] proposed a ''Metropolized'' version of adaptive rejection sampling (ARMS), basically an application of Tierney's *rejection sampling chains* [25]. For a comprehensive non-technical explanation of rejection sampling chains the reader is also referred to [26], for a detailed review of these variants of adaptive rejection sampling see [24], and for an empirical computational comparison see [27].

We used the universal algorithm ARMS [24,28] to sample from all the full conditional posterior densities in the following example.

TABLE I. The posterior mean, standard deviation (SD), time series standard error (SE), and lower quartile, median, and upper quartile of the parameters D, v, ψ, and ϕ_c (cf. Fig. 1).

Parameter	Mean	SD	SE	25%	Median	75%
D (Mpc)	689	168	5.34	546	704	820
v	0.709	0.234	0.0077	0.537	0.774	0.908
ψ	0.456	0.329	0.0125	0.171	0.366	0.725
ϕ_c	1.39	0.675	0.0259	0.817	1.58	1.97

III. EXAMPLE OF BAYESIAN ESTIMATION FOR A COALESCING BINARY SYSTEM

We now give an example whereby we apply MCMC methods in order to calculate the Bayes estimators for parameters in a coalescing binary system. We choose to reexamine an example used by Cutler and Flanagan [6]; the definition of parameters will come from their paper, and hence our notation will subsequently follow theirs. Using the Marković approximation [16] (i.e. there is no uncertainty of the location in the sky of the binary), the dependence on the detected wave amplitude parameter $\mathcal{A}_A$ [Eq. (4.16) of [6]] is reduced to the distance to the binary D, the cosine of the inclination angle ι to the line of sight $v=\cos\,\iota$, the polarization angle of the gravity wave ψ, and the phase of the waveform at the ''collision time'' ϕ_c.

In this example it is assumed that the LIGO-VIRGO network has detected a NS-NS binary merger with a signal-to-noise ratio of $\rho=12.8$ in the direction given by $(\theta,\varphi)=(50°,276°)$. For this direction the network parameters [Eq. (4.24) of [6]] are $\sigma_D=1.03$ and $\epsilon_D=0.74$. It is further assumed that from the detected signal the ''best-fit'' parameters [31] are the distance of $D_0=432$ Mpc, masses $M_1=M_2=1.4\,M_\odot$, the cosine of the inclination angle $v_0=0.31$, and the polarization angle $\psi_0=11.5°$. Note that in the caption of Fig. 10 of [6] an incorrect value of ψ_0 was stated, but the displayed PDF was calculated with $\psi_0=11.5°$ [29]. In Cutler and Flanagan's derivation of their PDF for D they found that their approximation technique eliminated the dependence of ϕ_{c0}; we will use a value of $\phi_{c0}=114.6°$, but we have found that the Bayes estimators for the other three parameters are independent of the choice of this angular value, in agreement with the previous analysis [6].

Our calculation will then commence with the posterior probability density function

$$p(\mathcal{A}_A)=\mathcal{N}p^{(0)}(\mathcal{A}_A)\times\exp\left[-\frac{r_0^2}{2}(\mathcal{A}_A-\hat{\mathcal{A}}_A)^*(\mathcal{A}_B-\hat{\mathcal{A}}_B)\Theta^{AB}\right] \tag{3.1}$$

where $\mathcal{N}$ is the normalization constant, $p^{(0)}$ the prior density, and

$$\mathbf{\Theta}=\sigma_D\begin{pmatrix}1+\epsilon_D & 0\\ 0 & 1-\epsilon_D\end{pmatrix} \tag{3.2}$$

[cf. Eq. (4.51) of [6]]. With the help of MATHEMATICA and Eq. (4.16) of [6] we explicitly write out $\mathcal{A}_A$ in terms of the parameters (D,v,ψ,ϕ_c) and $\hat{\mathcal{A}}_A$ in terms of $(D_0,v_0,\psi_0,\phi_{c0})$. We assume independent priors for the four parameters. For the marginal prior PDFs [cf. Eq. (4.56) of [6]] we choose ψ to be uniformly distributed between 0 and $\pi/2$, ϕ_c uniformly distributed between 0 and π, v uniformly distributed between -1 and 1, and the sources will be uniformly distributed within a sphere of radius $D_{max}=r_0=7.25$ Gpc (derived from the assumed noise dependence of advanced interferometer designs, Eq. (4.28) of [6]).

TABLE II. The cross-correlation matrix of the parameters D, v, ψ, and ϕ_c (cf. Fig. 1).

Variable	D	v	ψ	ϕ_c
D	1.0			
v	0.91	1.0		
ψ	0.416	0.48	1.0	
ϕ_c	−0.458	−0.528	−0.989	1.0

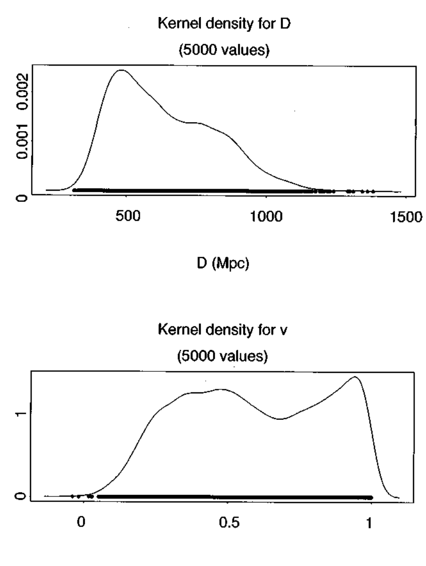

FIG. 2. Kernel density estimates of the marginal posteriors of the variables D,v starting with the probability distribution function obtained via Eq. (4.57) of [6] and assuming the initial ''best-fit'' parameters of $D_0=432$ Mpc, $v_0=0.31$, and $\psi_0=11.5°$.

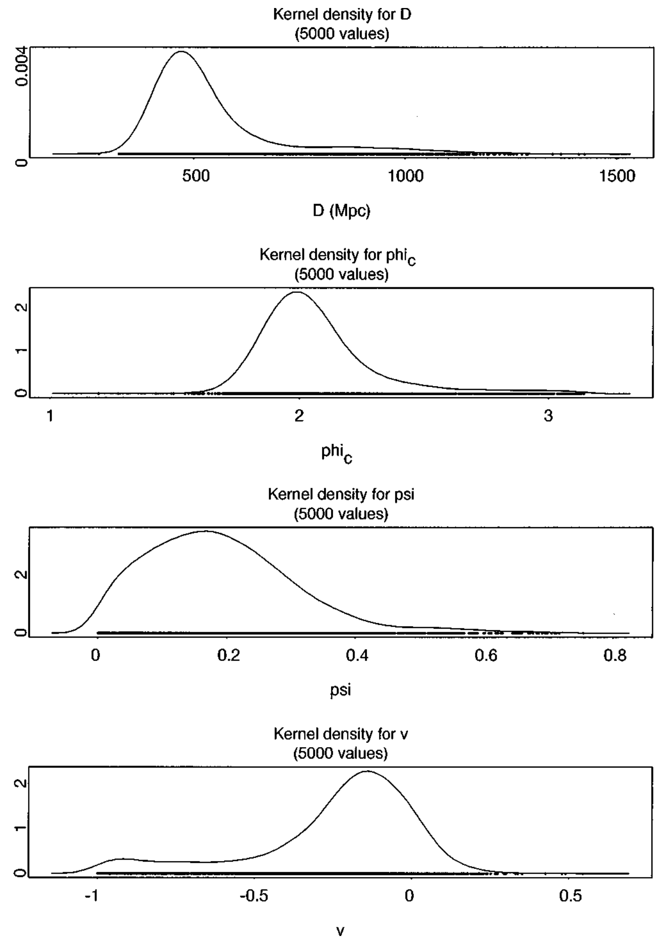

FIG. 3. Kernel density estimates of the marginal posteriors of the variables D, υ, ψ, and ϕ_c assuming the initial ''best-fit'' parameters of $D_0 = 432$ Mpc, $\upsilon_0 = -0.1$, $\psi_0 = 11.5°$, and $\phi_{c0} = 114.6°$.

We employ the Gibbs sampler (2.6) in conjunction with the universal algorithm ARMS to sample from the posterior distribution. This is implemented as a C-program on a SUN ULTRA workstation (143 MHz). We perform 120 000 cycles of the Gibbs sampler and thin the chain by taking every 20th observation to avoid highly correlated values. For the remaining 6000 samples we use a burn-in of 1000 which yields a final chain of length 5000. A typical run for the four parameter case takes about 20 min.

Extensive convergence diagnostics were calculated for the four parameters D, υ, ψ, and ϕ_c using the CODA software [19]. All chains passed the Heidelberger-Welch stationarity test. The Raftery-Lewis convergence diagnostics confirmed that the thinning and burn-in period were sufficient. Lags and autocorrelations within each chain were reasonably low. Geweke's Z-scores were low for all parameters. These convergence diagnostics are summarized in [19] (see also references therein).

In Fig. 1 we display the kernel density estimates [30] of the four marginal posterior densities using a bandwidth of one-eighth of the sample range. In contrast to elaborate asymptotic techniques precision estimates of the Bayes estimators, e.g. the distance measurement accuracy ($\Delta D/D = 0.24$ in this example), can easily be calculated using MCMC methods. Summary statistics including posterior mean, standard deviation, the time series standard error (the square root of the spectral density estimate divided by the sample size), and the 25%, 50%, and 75% quantiles are listed in Table I. In Table II we present the cross-correlation matrix.

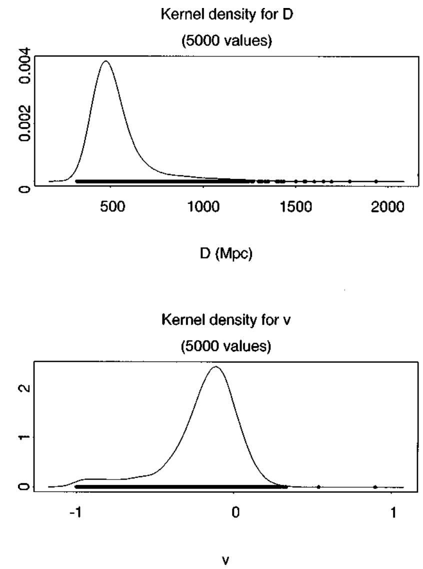

FIG. 4. Kernel density estimates of the marginal posteriors of the variables D, υ starting with the probability distribution function obtained via Eq. (4.57) of [6] and assuming the initial ''best-fit'' parameters of $D_0 = 432$ Mpc, $\upsilon_0 = -0.1$, and $\psi_0 = 11.5°$.

A number of noticeable features of the parameters are discernable from Fig. 1 and Tables I and II. One can observe that the angular distributions are relatively flat. For ϕ_c there is a peak at $\phi_{c0} = 114.6°$. The distribution of ψ has its peak at $\psi_0 = 11.5°$. There is a strong correlation of 0.91 between the parameters and υ and D. A small υ implies a smaller signal, which could also be caused by a greater distance D. This is reflected in the posterior PDF for D. The Bayes estimators, i.e. the posterior means, found via our analysis, $\overline{D} = 689$ Mpc and $\overline{\upsilon} = .709$, are not necessarily the same as the ''maximum-likelihood estimates'' D_0 and υ_0, i.e. the posterior modes; this effect was also noted before [6]. Asymptotic theory tells us that the posterior distribution can be approximated by a Gaussian distribution having mean equal to the posterior mode. This example illustrates the shortcomings of this Gaussian approximation [31].

A joint posterior PDF for D and υ was derived by Cutler and Flanagan [6] by expanding the argument of the exponential in Eq. (3.1) to second order in $\psi - \psi_0$ and $\phi - \phi_{c0}$ and integrating over ψ and ϕ_c. The result is given by Eq. (4.57) of [6], and by numerical integration over υ they obtain their Fig. 10 [6]. Starting with Eq. (4.57) of [6] we can also apply our MCMC method to obtain the marginal posterior PDF of D and υ. Figure 2 shows that our procedure gives a similar distribution for D as that displayed in Fig. 10 of [6]. However, we also note that when the four parameters for the problem are maintained throughout the calculation we obtain differing distributions for D and υ. Our MCMC method is

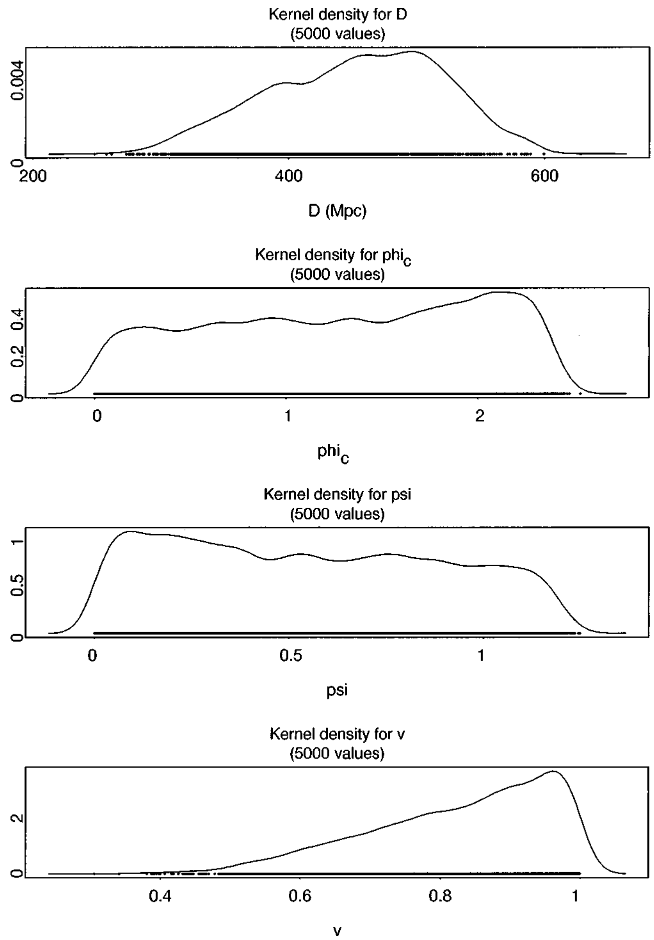

FIG. 5. Kernel density estimates of the marginal posteriors of the variables D, v, ψ, and ϕ_c assuming the initial ''best-fit'' parameters of $D_0=432$ Mpc, $v_0=0.8$, $\psi_0=11.5°$, and $\phi_{c0}=114.6°$.

not time consuming to run, and it correctly calculates the Bayes estimators without bringing in potential errors by utilizing approximation techniques.

By way of comparison we display in Fig. 3 the Bayes estimators for all four parameters with $v_0=-0.1$ (all other parameters the same), and compare it to the results from a two dimensional MCMC method (Fig. 4) based on the approximate two dimensional PDF of Eq. (4.57) of [6]. We likewise do the same for $v_0=0.8$ (Figs. 5, 6). A numerical integration of Eq. (4.57) of [6] over v produced similar distributions for D as seen in Figs. 4, 6. For these values there is better agreement between the marginal posterior PDFs from the four dimensional problem and the two-dimensional approximation. One can observe that the distributions for ψ and ϕ_c become narrowly peaked as v approaches 0; this displays the fact that the observed signal depends more critically on the precise angular values when the orbital plane of the binary system is edge-on to our line of sight.

IV. CONCLUSION

The application of MCMC methods should prove useful in gravitational radiation data analysis. Bayesian statistical

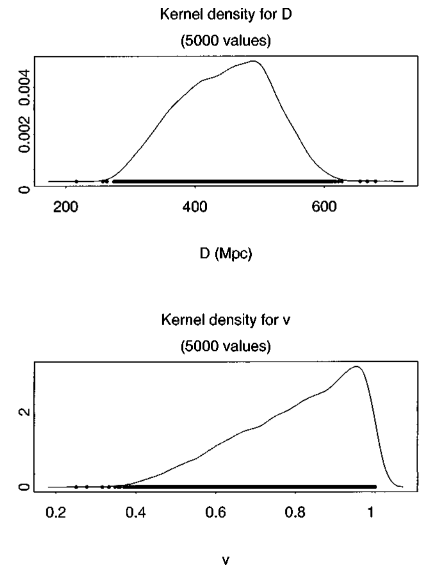

FIG. 6. Kernel density estimates of the marginal posteriors of the variables D, v starting with the probability distribution function obtained via Eq. (4.57) of [6] and assuming the initial ''best-fit'' parameters of $D_0=432$ Mpc, $v_0=0.8$, and $\psi_0=11.5°$.

techniques can be utilized for different types of waveforms, but coalescing compact binary systems offer obvious and exciting prospects. The expected ''clean'' signal from the system will allow excellent prospects for parameter estimation. This in turn will provide important insights into cosmology, general relativistic effects, and potentially even the equation of state of the neutron star. All of these important physics goals depend on a successful parameter estimation procedure. Bayesian techniques are well suited for this program, but unfortunately progress has been hindered by the complexity of the multidimensional integrals necessary for the calculation of the Bayes estimators.

The MCMC methods discussed in this paper can be directly applied to Bayesian gravitational radiation data analysis. Over the last ten years MCMC techniques have greatly simplified statistical calculations in numerous scientific disciplines. Even when one includes post-Newtonian effects in the dynamics of the binary merger, plus spin interaction effects [6,9], the computational ability of MCMC methods should permit the calculation of all posterior distributions.

In this paper we have examined a relatively simple system whereby the amplitude of the gravity wave depends upon only four parameters. This system was used by Cutler and Flanagan [6] to display the increase in information that a Bayesian analysis can provide. Even with only a four dimensional integral the exact extraction of all the Bayes estimators proved to be numerically prohibitive. The MCMC method permits a quick and accurate solution. The computational time will increase only linearly with the number of parameters; this is a marked improvement over the exponentially increasing time scale for multidimensional numerical integrations. We have shown that our more exact solution can eliminate potential errors created by the instigation of approximation techniques.

The implementation of the MCMC method to the full parameter estimation problem of coalescing compact binaries will not be trivial. In this present paper we wish to bring this technique to the attention of the numerous physicists who are actively involved in data analysis studies for the LIGO-VIRGO systems. We are currently working on the extension of full Bayesian techniques, to be applied to the coalescing binary problem, and the application of MCMC methods for parameter estimation. Our initial results will be presented in a forthcoming publication.

ACKNOWLEDGMENTS

This work was supported by the Royal Society of New Zealand Marsden Fund and the University of Auckland Research Committee.

[1] A. Abramovici, W. E. Althouse, R. W. P. Drever, Y. Gürsel, S. Kawamura, F. J. Raab, D. Shoemaker, L. Sievers, R. E. Spero, K. S. Thorne, R. E. Vogt, R. Weiss, S. E. Whitcomb, and M. E. Zucker, Science **256**, 325 (1992).

[2] C. Bradaschia *et al.*, Nucl. Instrum. Methods Phys. Res. A **289**, 518 (1990).

[3] K. S. Thorne, in *300 Years of Gravitation*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1987), pp. 330–458.

[4] L. S. Finn and D. F. Chernoff, Phys. Rev. D **47**, 2198 (1993).

[5] B. F. Schutz, Nature (London) **323**, 310 (1986).

[6] C. Cutler and E. F. Flanagan, Phys. Rev. D **49**, 2658 (1994).

[7] L. S. Finn, Phys. Rev. D **53**, 2878 (1996).

[8] C. Cutler, T. A. Apostolatos, L. Bildsten, L. S. Finn, E. E. Flanagan, D. Kennefick, D. M. Marković, A. Ori, E. Poisson, G. J. Sussman, and K. S. Thorne, Phys. Rev. Lett. **70**, 2984 (1993).

[9] E. E. Flanagan and S. A. Hughes, Phys. Rev. D **57**, 4535 (1998).

[10] P. C. Gregory and T. J. Loredo, Astrophys. J. **398**, 146 (1992).

[11] L. S. Finn, gr-qc/9709077.

[12] L. S. Finn, Phys. Rev. D **46**, 5236 (1992).

[13] D. Nicholson and A. Vecchio, Phys. Rev. D **57**, 4588 (1998).

[14] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).

[15] S. Geman and D. Geman, IEEE Trans. Pattern. Anal. Mach. Intell. **6**, 721 (1984).

[16] D. Marković, Phys. Rev. D **48**, 4738 (1993).

[17] C. P. Robert, *The Bayesian Choice* (Springer, New York, 1996).

[18] M. K. Cowles, B. P. Carlin, J. Am. Stat. Assoc. **91**, 883 (1996).

[19] N. G. Best, M. K. Cowles, and S. K. Vines, *CODA manual version 0.30* (MRC Biostatistics Unit, Cambridge, England, 1995).

[20] L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York, 1986).

[21] W. R. Gilks and P. Wild, Appl. Stat. **41**, 337 (1992).

[22] B. D. Ripley, *Stochastic Simulation* (Wiley, New York, 1987).

[23] W. R. Gilks, in *Bayesian Statistics 4*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Oxford University Press, Oxford, 1992), pp. 89–114.

[24] W. R. Gilks, N. G. Best, and K. K. C. Chan, Appl. Stat. **44**, 455–472 (1995).

[25] L. Tierney, Ann. Stat. **22**, 1701 (1994).

[26] S. Chib and E. Greenberg, Am. Stat. **49**, 327 (1995).

[27] A. E. Gelfand and T. M. Lee, J. R. Stat. Soc. B **55**, 72 (1993).

[28] W. R. Gilks and R. M. Neal, Appl. Stat. **46**, 541 (1997).

[29] E. E. Flanagan (private communication).

[30] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986).

[31] The method for determining the ''best-fit'' or ''maximum-likelihood'' parameters is not addressed here. Our purpose in this present paper is only to display the MCMC methods for Bayesian computation in systems that may seem prohibitive due to the complexity of the multidimensional integrals. We feel that a proper Bayesian approach to parameter estimation in the coalescing compact binary systems has not been adequately described; this is a topic of current research by us and will be discussed in a forthcoming publication.