LIGO-G000058-00-D

New Parameter Estimation, Data Analysis and Statistical Strategies for LIGO

Nelson Christensen
Physics and Astronomy
Carleton College ~ ~19 majors/yr physics
Northfield, MN 55057

nchriste@carleton.edu

A "random walk" through templates which also generates statistics and confidence intervals.

(i.e. coalescing binaries)

Serial walk through templates does not generate statistics

Estimation of Parameters for Coalescing Binary Signals

Likelihood * *A Priori* Distributions

Generate Probability Distribution Functions for each parameter in problem

Use Markov Chain Monte Carlo (MCMC) Methods for this computational exercise

Bayes Theorem

Given a measurement $\mathbf{y}(t) = \mathbf{n}(t) + \mathbf{s}(t)$ (signal and noise) and prior belief $p(\vec{\theta})$ about the parameters $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$, the researcher's opinion as to the state of nature is summarized by the posterior distribution of $\vec{\theta}$ given by the Bayes theorem

$$p(\vec{\theta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \vec{\theta})p(\vec{\theta})}{m(\mathbf{y})} \propto p(\vec{\theta})p(\mathbf{y} \mid \vec{\theta})$$

where

$$m(\mathbf{y}) = \int p(\vec{\theta})p(\mathbf{y} \mid \vec{\theta})d\vec{\theta}$$

is the marginal density of $\mathbf{y}$, which can be regarded as a normalizing constant as it is independent of $\vec{\theta}$.

WE CAN THEN DETERMINE THE PROBABILITY FOR HAVING CERTAIN PARAMETER VALUES GIVEN THE DATA.

The Likelihood:

$$L(\vec{\theta}) = p(y|\vec{\theta})$$

In the gravity wave interferometer world

$y(t) = n(t) + s(t,\vec{\theta})$, for $n(t)$ noise, and signal $s(t,\vec{\theta})$ defined by parameters $\vec{\theta}$,

$L(\vec{\theta}) = K \exp\left[2\langle y, s(\vec{\theta})\rangle - \langle s(\vec{\theta}), s(\vec{\theta})\rangle\right]$ with normalization constant K,

$$\langle y,s\rangle = \int_{-\infty}^{\infty} df \frac{\tilde{y}(f)\tilde{s}^*(f)}{S_n(|f|)}, \quad \tilde{y}(f) = \int_{-\infty}^{\infty} y(t)e^{2\pi i t}dt,$$

and $S_n(|f|)$ is the one-sided power spectral density of the detector's noise.

Templates estimate the signal $\Rightarrow s(t,\vec{\theta})$

With information on the a priori probabilities for the parameters $p(\theta_i)$
→ *posterior density of parameters* $p(\vec{\theta}\mid y) \propto L(y\mid\vec{\theta})\prod_i p(\theta_i)$.

Hence, when one desires the PDF for the parameter θ_i one must normalize the posterior density and integrate over the remaining parameters, namely

Very hard to compute

$$p(\theta_i \mid y) = N \int d\theta_1 \ldots \int d\theta_{i-1} \int d\theta_{i+1} \ldots \int d\theta_n L(y\mid\vec{\theta}) p(\theta_1)\ldots p(\theta_{i-1}) p(\theta_{i+1})\ldots p(\theta_n).$$

This normalization and multiple parameter integration procedure is **prohibitive** for a large number of parameters (say greater than ~5) except when one uses MCMC methods.

Coalescing binary parameters:
Template matching $M_1, M_2, \phi_c, t_c, \varepsilon, \vec{S}_1, \vec{S}_2, \vec{L}$ or even $\tau_0, \tau_1, \tau_{1.5}, \ldots$ as a test of GR

and also the "amplitude affecting parameters"
$D, \cos\iota, \theta, \varphi, \psi$

Gravity wave researchers have recognized the importance of the statistically correct posterior density $p(\vec{\theta} \mid y) \propto L(y \mid \vec{\theta}) \prod_i p(\theta_i)$

but neglected to use it due to complexity of integrations over numerous parameter dimensions.

A revolution in computational statistics in last few years

Markov Chain Monte Carlo (MCMC) Methods

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

$$p(\theta_i \mid y, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$$

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

This define the walk through the parameter space

simulate $\theta_1^{(1)} \sim p(\theta_1 \mid y, \theta_2^{(0)}, \ldots, \theta_n^{(0)})$

simulate $\theta_2^{(1)} \sim p(\theta_2 \mid y, \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_n^{(0)})$

•

•

•

simulate $\theta_n^{(1)} \sim p(\theta_1 \mid y, \theta_1^{(1)}, \ldots, \theta_{n-1}^{(1)})$

Generates a PDF for each parameter

yields $\vec{\theta}^{(m)} = (\theta_1^{(m)}, \ldots, \theta_n^{(m)})$ after m cycles. Defines a Markov chain with transition kernel

$$k(\vec{\theta}^{(m+1)}, \vec{\theta}^{(m)}) = \prod_{i=1}^{n} p(\theta_i^{(m+1)} \mid y, \theta_1^{(m+1)}, \ldots, \theta_{i-1}^{(m+1)}, \theta_{i+1}^{(m)} \ldots, \theta_n^{(m)})$$

that converges to the joint posterior as its equilibrium distribution.

The problem of sampling from an n-variate PDF is reduced to sampling from n univariate PDFs.

→ Cross correlation matrix of components another result of MCMC calculation. spin & masses?

The MCMC computational time scales linearly with parameter number.

Not prohibitive as parameter number increases

A random, but wise, walk through parameter space with templates.

1) Create probability distribution functions for parameters.

2) Create confidence intervals

3) The MCMC "random walk" spends more time sampling where signal is likely, less (~nil) in unlikely regions

Maximum of the Likelihood is not necessarily the peak value of the parameter's marginal distribution (its PDF) or its mean.
(See Cutler and Flanagan, Phys Rev D 49, 2658 (1994), or Christensen and Meyer, Phys Rev D 58, 082001, (1998))

Maximum Likelihood is not good for parameter estimation; for a given parameter, the maximization depends on whether or not other parameters have been integrated out.

The posterior density $p(\vec{\theta} \mid \mathrm{y}) \propto L(\mathrm{y} \mid \vec{\theta}) \prod_{i} p(\theta_i)$ is the correct statistical tool, and MCMC will provide the PDF for each individual parameter.

If PDF is NOT sharply peaked, then the 97% *fitting-factor* is certainly overkill.

Cross-correlation between parameters will broaden distributions.

Instead of a serial walk through parameters in template generation, MCMC will conduct a *Brownian* or *random-walk* through parameter space.

It is yet to be determined if this MCMC approach will be more efficient in parameter determination, HOWEVER, the MCMC process will also provide the valuable PDF for each parameter. Can determine confidence intervals, statistics, cross correlation between parameters.

As parameter number of increases, computational time for a MCMC march through the templates scales linearly.

Previous work: MCMC Example based on problem posed in Cutler and Flanagan Phys Rev D 49, 2658 (1994)

Uncertainty in distance to coalescing binary. Assumed signal detected by templates, and sky location known.

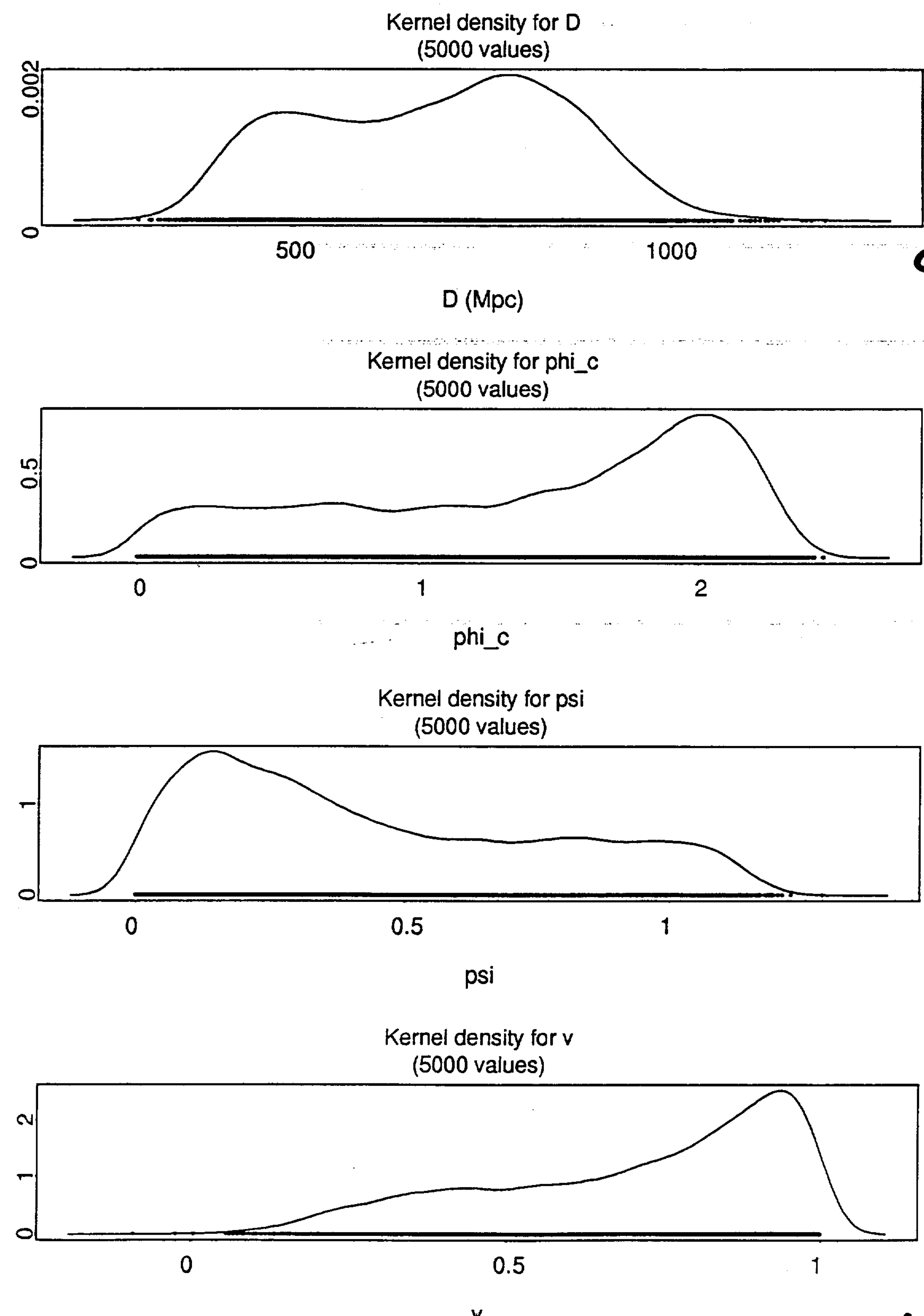
Four Parameters:

- Distance D
- Cosine of angle of inclination of orbital plane $\nu = \cos\iota$
- Polarization angle of gravity wave ψ
- Phase of waveform at "collision time" ϕ_c

Given "maximum likelihood" values $(D_0, \nu_0, \psi_0, \phi_{c0})$
Determine PDFs for each parameter

See : Markov Chain Monte Carlo Methods for Bayesian Gravitation Radiation Data Analysis, Christensen and Meyer, Phys Rev D 58, 082001, (1998))

FIG. 1. 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, $v_0 = 0.31$, $\psi_0 = 11.5°$, and $\phi_{c0} = 114.6°$.



0<D<7.2 Gpc

-1<v<1

TABLES

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

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

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

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

In order to *sample* parameter θ_2

during the mth loop of the Gibb's sampler

$$\theta_2^{(m)} \sim p\left(\theta_2 \mid \mathrm{y}, \theta_1^{(\mathrm{m})}, \theta_3^{(\mathrm{m}\text{-}1)} \ldots, \theta_n^{(m-1)}\right)$$

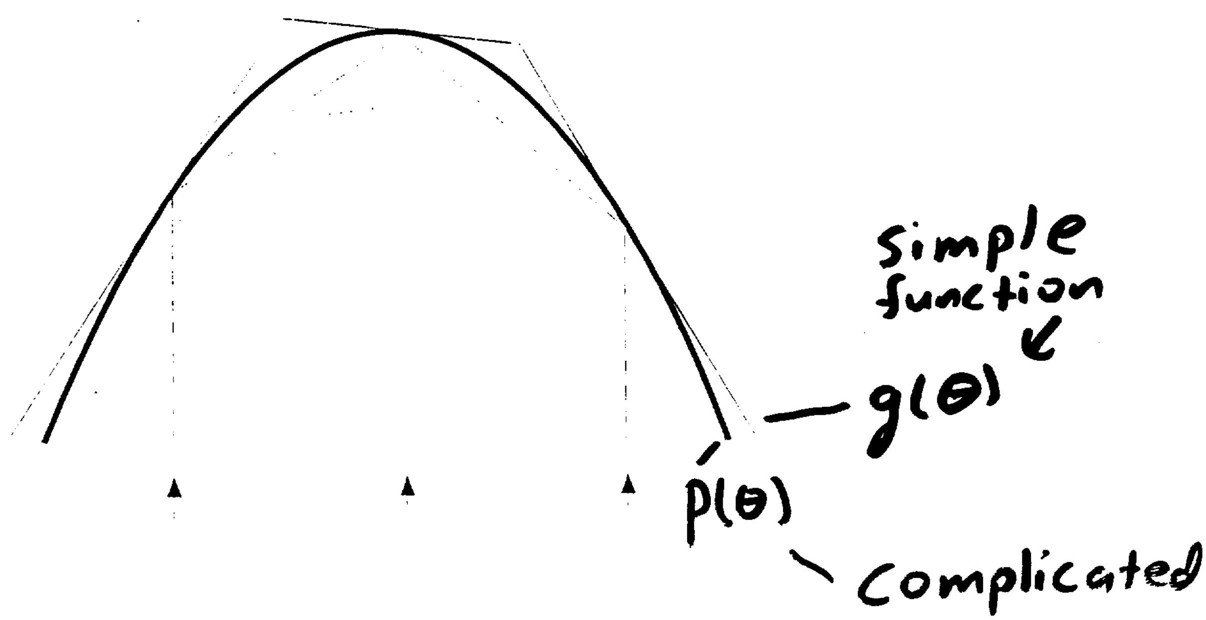
$$p\left(\theta_2 \mid \mathrm{y}, \theta_1^{(\mathrm{m})}, \theta_3^{(\mathrm{m}\text{-}1)} \ldots, \theta_n^{(m-1)}\right) \propto p(\theta_2) L\left(\mathrm{y} \middle| \vec{\theta}\right)$$

where $L\left(\mathrm{y} \middle| \vec{\theta}\right)$ is formed with parameters

$\theta_1 = \theta_1^{(m)}, \quad \theta_3 = \theta_3^{(m-1)}, \ldots, \theta_n = \theta_n^{(m-1)}$ **fixed and θ_2 a variable.**

Sample from $p\left(\theta_2 \mid \mathrm{y}, \theta_1^{(\mathrm{m})}, \theta_3^{(\mathrm{m}\text{-}1)} \ldots, \theta_n^{(m-1)}\right) \propto p(\theta_2) L\left(\mathrm{y} \middle| \vec{\theta}\right)$

via a *Metropolized* version of adaptive rejection sampling (ARMS)



Sample θ from g

Accept new $\hat{\theta}$ with probability $\dfrac{p(\hat{\theta})}{g(\hat{\theta})}$

Projects in the Immediate Future

(1) Data Analysis, Statistics and Parameter Estimation for Binary Inspiral Searches

Derive and develop optimized code for applying Markov Chain Monte Carlo (MCMC) techniques to the parameter estimation problem for coalescing binary observations.

i. Simulate (noisy) LIGO data containing (post-Newtonian) coalescing binary events.

ii. Generate post-Newtonian templates, again using with existing software.

iii. Construct a likelihood function with the simulated data and templates.

iv. Generate *a priori* distributions for the signal parameters via basic astrophysical assumptions.

v. Incorporate the data, template generation, likelihood function and *a priori* distributions into a Gibb's sampler MCMC, and thereby generate the probability distribution functions for each parameter of the model.

vi. Determine the computational cost of implementing these MCMC procedures for use in the LIGO environment.

Work in collaboration with Dr. R. Meyer, Department of Statistics, University of Auckland, New Zealand, and Prof. L.S. Finn, Department of Physics, Penn State.

(2) Maximum likelihood analysis versus integrated likelihood analysis

Investigate the relative efficiency of a maximum likelihood analysis as compared to an integrated likelihood (e.g., a Bayesian approach with a uniform prior) analysis of the same significance.

This would be explored in the context of the binary inspiral problem.

Need to more efficiently process the outputs from the parallel Wiener-filter (i.e. template) analysis in a way that uses present inherent information rather than simply identifying the "loudest" (i.e. maximum likelihood) event.

Quantify whether the integrated likelihood search is more powerful than a maximum likelihood analysis, and identify the computational costs of implementation.

a. Simulate LIGO data with and without (post-Newtonian) coalescing binary events.

b. Generate post-Newtonian templates.

c. Construct a likelihood function with the simulated data and templates.

d. Use uniform *a priori* distributions for the signal parameters and incorporate the data, template generation, likelihood function and *a priori* distributions into a Gibb's sampler MCMC.

e. Search sets of data (with and without embedded signal) over some time window and region of parameter space using both the maximum likelihood and integrated likelihood techniques. Compare the maximum likelihood and integrated likelihood techniques by examining the generated probability distribution functions.

f. For the maximum likelihood and integrated likelihood techniques determine the fraction of the time that an "actual" event is detected, and compare it to false alarm rates. Pick an acceptable false alarm rate for given signal rates and strengths find the associated false dismissal rate.

g. Study the computational requirements for these maximum likelihood and integrated likelihood techniques.

Work in collaboration with Dr. R. Meyer, Department of Statistics, University of Auckland, New Zealand, and Prof. L.S. Finn, Department of Physics, Penn State.

Note 1, Linda Turner, 05/09/00 09:17:51 AM
LIGO-G000058-00-D