Bayesian Analysis of RR Lyrae Distances and Kinematics

William H. Jefferys, Thomas R. Jefferys and Thomas G. Barnes

University of Texas at Austin, USA

Thanks to: Jim Berger, Peter Müller, Charles Friedman

Abstract

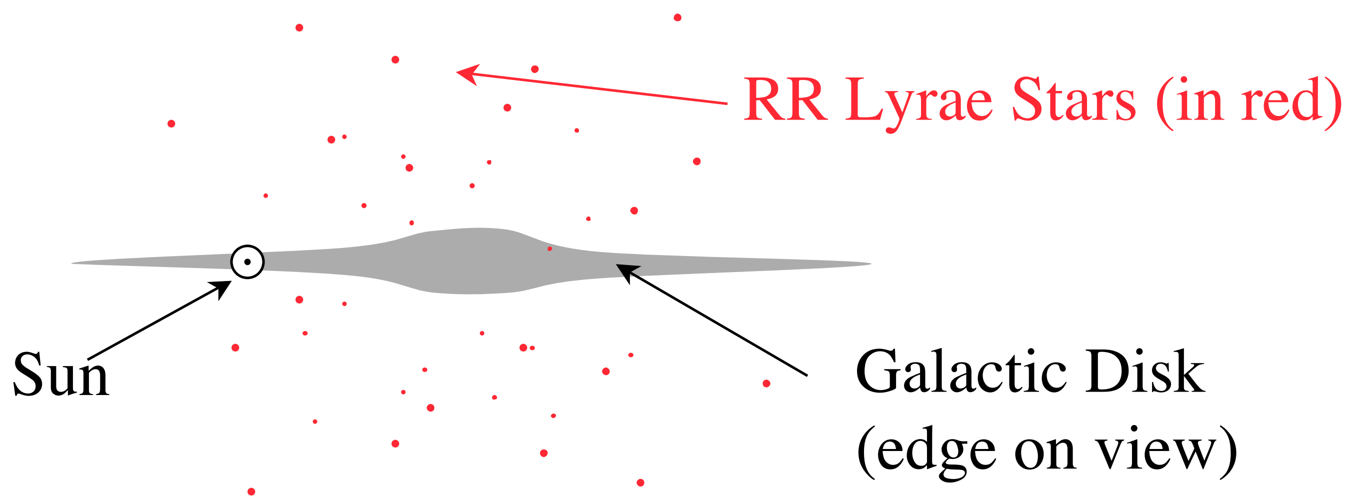
- We have developed a hierarchical Bayes model to analyze the distances, luminosities, and kinematics of RR Lyrae stars. Our model relates these parameters to the observed proper motions, radial velocities, apparent luminosities and metallicities of the stellar sample. We use a Metropolis-within-Gibbs sampler to draw an MCMC sample from the full posterior distribution of the parameters (including latent variables), and draw inferences on the quantities of interest in the usual way. We are testing our model with a small database from the literature and will eventually apply it to a new large database from the European HIPPARCOS satellite.

Outline

- A little astronomy—what's the goal?
- Basic Bayesian Inference
- Markov Chain Monte Carlo
- Mathematical model and likelihood function
- Priors
- Sampling strategy
- Results
- Future research

Description of the Problem

- RR Lyrae stars are a class of pulsating variable stars. They are readily recognizable from their periods 0.75±0.25 days) and characteristic light curves.
- They are old (evolved) stars with a roughly spherical spatial distribution in the galaxy and distinctive kinematics (statistical description of their motions in the galaxy)



Description of the Problem

- They are fairly bright (40 times as bright as the Sun) and so can be seen to fair distances in the galaxy.
- Their intrinsic mean visual-band luminosities are nearly constant.
  - This is known from studies of RR Lyrae stars in clusters, where all the stars are at the same distance.
- Useful as "standard candles" for estimating the distance of an object (like a star cluster).
  - Observe the apparent luminosities of RR Lyrae stars in the object, know their intrinsic luminosities, and use inverse square law of light to calculate the distance to the object.

Goals of Our Investigation

- Determine the absolute magnitude (log luminosity) of these stars
- Investigate the kinematics of the stars as a group (i.e., it is believed that the velocities of these stars have a roughly multivariate normal distribution; what are the parameters of that distribution?)
- Investigate the "cosmic scatter" of the magnitude (i.e., the variation about the mean unexplained by other variables)
- Investigate any variation of absolute magnitude with "metallicity" (i.e., content of elements heavier than helium)

How We Do It

- Our raw data are the proper motions $\boldsymbol{\mu}$ (vector of angular motion/time unit of the motions of the stars across the sky), the radial velocities ρ (kilometers/second of the motions towards or away from the Sun, obtained by Doppler shift), and the apparent magnitudes m of the stars, assumed measured without error.
- The proper motions are related to the cross-track velocities (in km/sec) by multiplying the former by the distance s to the star:

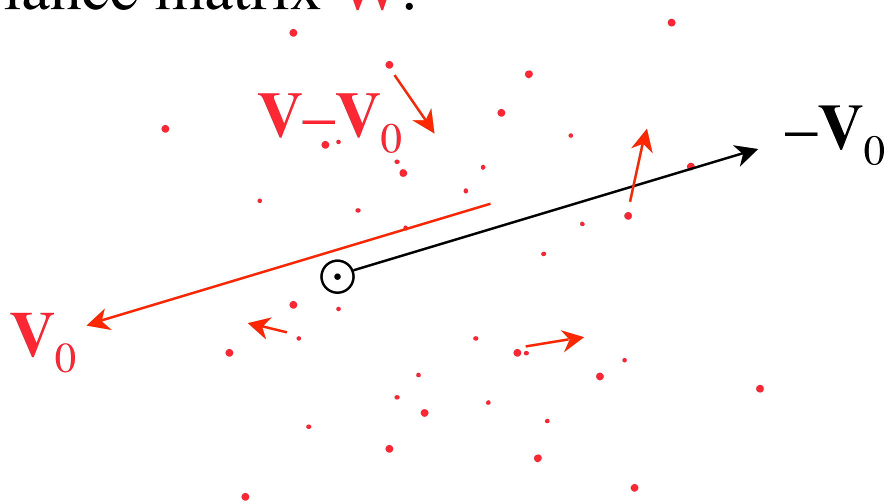
$$s\boldsymbol{\mu} \propto \mathbf{V}^{\perp}$$

If we assume that the proper motions and radial velocities are characterized by the same kinematical parameters, we can (statistically) infer s and then the magnitude M of the star through a defined relationship:

$$s = 10^{0.2(m-M+5)}$$

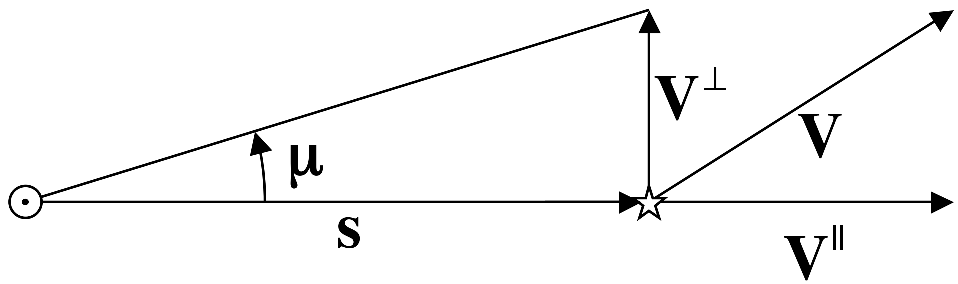
How We Do It

- As the Sun plunges through the swarm of RR Lyrae stars, we measure their proper motions and radial velocities, which are due to two sources.
  - The motion of the Sun at high velocity through the relatively stationary swarm, with velocity $-\mathbf{V}_0$
  - The peculiar velocities $\mathbf{V}-\mathbf{V}_0$ of the stars in the swarm relative to the swarm. The latter is assumed characterized by a multivariate normal distribution with mean zero and covariance matrix $\mathbf{W}$.



How We Do It

- Here's a picture, viewed from a coordinate system in which the Sun is stationary. The velocity $\mathbf{V}$ is the vector sum of the swarm's velocity relative to the Sun ($\mathbf{V}_0$) and the peculiar velocity of the star relative to the swarm ($\mathbf{V}-\mathbf{V}_0$).



$$s = |\mathbf{s}|$$

Outline

- A little astronomy — what's the goal?
- Basic Bayesian Inference
- Markov Chain Monte Carlo
- Mathematical model and likelihood function
- Priors
- Sampling strategy
- Results
- Future research

Basic Bayesian Inference

- The basic differences between Bayesian and classical (frequentist) inference can be summed up as follows:
  - The frequentist studies the behavior of functions $\delta(x)$ of the data vector x under repeated *iid* sampling on x (e.g., an estimator, a confidence interval).
  - She regards *data* as a random variable conditional on the *fixed but unknown* parameter vector θ that describes the states of nature.
  - Thus, for example, a frequentist confidence interval is constructed in such a way as to guarantee that if the prescribed procedure is computed for many *iid* trials, then the parameter θ will be captured by the interval in a prescribed proportion (e.g., 95%) of the trials

Basic Bayesian Inference

- The Bayesian, regards the *actual* data x that have been observed as *fixed and known* (but generated by a random process), and the parameter vector θ as a *random variable*. She conditions everything on x.
- She regards the randomness of θ as measuring her *state of knowledge or ignorance* about these parameters.
- Thus, when she computes a Bayesian credible interval, she says that there is a probability 0.95 (say) that the parameter lies in that interval, conditioned on x and her prior knowledge.

Basic Bayesian Inference

- The Bayesian starts with a *prior distribution* $\pi(\theta)$ on the parameters θ and, when given data x with sampling distribution $f(x|\theta)$ and likelihood function $l_x(\theta)\propto f(x|\theta)$, updates the distribution using *Bayes' rule* to obtain a *posterior distribution* $\pi(\theta|x)$:

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{p(x)} = \frac{p(x,\theta)}{p(x)}$$

- Thus, the "Bayesian mantra" (the key identity of Bayesian inference):

posterior ∝ prior × likelihood

- All results are derived from the posterior distribution, e.g., point estimators of parameters, variances, credible intervals, marginal distributions, and other summaries of the inference.

Basic Bayesian Inference

- Priors may be informative (e.g., they reflect actual prior information about a parameter) or "noninformative" (supposedly neutral in some sense about where the parameter lies, so as to "let the data speak for themselves")
- When a particular "noninformative" prior is used routinely, it is desirable that it also have good frequentist properties
  - For example, one would want the Bayes estimators associated with the solution to be *admissible* under an appropriate loss function
- We will see examples of both concerns later

Basic Bayesian Inference

- A great advantage of Bayesian methods is their uniformity and simplicity. The same basic components (prior, likelihood, posterior) are used in basically the same way in all problems. This makes it very easy for the statistician to model a wide variety of problems. To solve any problem, we
  - Identify and write down the likelihood
  - Decide on priors for all parameters or states of nature
  - Compute the posterior (up to a constant factor) using the "Bayesian mantra"

    $$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

  - Make inferences from the posterior (conceptually simple but may require computationally intense methods)

Hierarchical Bayes

- A very rich class of models is provided by *hierarchical Bayes* models, of which the present problem is an example
- One may find that some parameters in the likelihood depend on other parameters not mentioned in the likelihood. For example:

$$\begin{aligned}
\text{Likelihood}: &\ l_x(\theta) \propto f(x \mid \theta) \\
\text{Prior on } \theta: &\ \pi(\theta \mid \sigma) \\
\text{Prior on the new parameter } \sigma: &\ \pi(\sigma) \\
\text{Posterior}: &\ \pi(\theta, \sigma \mid x) \propto f(x \mid \theta)\pi(\theta \mid \sigma)\pi(\sigma)
\end{aligned}$$

Hierarchical Bayes

- We represent dependencies by a directed acyclic graph (DAG)

$$\boxed{\sigma} \rightarrow \boxed{\theta} \rightarrow \boxed{x}$$

$$\text{Likelihood}: \; l_x(\theta) \propto f(x \mid \theta)$$

$$\text{Prior on } \theta: \; \pi(\theta \mid \sigma)$$

$$\text{Prior on the new parameter } \sigma: \; \pi(\sigma)$$

$$\text{Posterior}: \; \pi(\theta, \sigma \mid x) \propto f(x \mid \theta)\pi(\theta \mid \sigma)\pi(\sigma)$$

Outline

- A little astronomy — what's the goal?
- Basic Bayesian Inference
- Markov Chain Monte Carlo
- Mathematical model and likelihood function
- Priors
- Sampling strategy
- Results
- Future research

Markov Chain Monte Carlo

- The most difficult part of Bayesian inference has been the requirement to integrate functions of the posterior distribution over the parameters θ. This is needed for many purposes, e.g.,
  - Posterior mean as a point estimator of θ_i
$$\bar{\theta}_i = \int \theta_i \pi(\theta \mid x) d\theta$$
  - Posterior mean as an estimator of some function of θ
$$\bar{g} = \int g(\theta) \pi(\theta \mid x) d\theta$$
  - …for example, the posterior variance of θ_i
$$\mathrm{Var}(\theta_i) = \int (\theta_i - \bar{\theta}_i)^2 \pi(\theta \mid x) d\theta$$

Markov Chain Monte Carlo

- To get any of these one needs to know the normalization constant in Bayes' rule, which is also gotten by integration. Rewriting Bayes' rule we find that

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{p(x)} = \frac{f(x \mid \theta)\pi(\theta)}{\int f(x \mid \theta)\pi(\theta)d\theta}$$

$p(x)$ is just the normalization factor required to deliver a probability; it is called the marginal probability of the data.

Markov Chain Monte Carlo

- Another example would be computing a marginal distribution. The standard Bayesian way to treat nuisance parameters is simply to integrate them out, deriving the marginal posterior distribution of the parameter(s) of interest

$$\pi(\theta_i \mid x) = \int \pi(\theta \mid x) d\theta_{-i}$$

where θ_{-i} denotes the components of θ excepting for θ_i

Markov Chain Monte Carlo

- Most integrals can't be computed in closed form
- θ may contain a very large number of components
  - Numerical integration generally unfeasible
- Integration has been a huge obstacle to doing any but the simplest Bayesian calculations
  - Despite its conceptual appeal, power, and simplicity, Bayesian methods were not widely practiced.
- Over the past 15 years the development of Markov chain Monte Carlo (MCMC) methods has completely changed the Bayesian landscape, so that Bayesian methods are being used increasingly for complex problems by statisticians of all philosophical persuasions.

Markov Chain Monte Carlo

- The idea behind MCMC is very simple: Given the easily obtained joint distribution of x and θ

$$p(x,\theta)=f(x|\theta)\pi(\theta)$$

generate a large sample from the posterior distribution.

- Fortunately, this can be done without knowing the normalization constant! Thus we avoid having to compute this constant.

Markov Chain Monte Carlo

- Given a sample from the posterior distribution, we can use ordinary frequentist methods to generate desired results from the sample:
  - Posterior mean of some function $g(\theta)$ approximated by the sample mean of θ from the sample

$$\bar{g} \approx \frac{1}{N}\sum_{i=1}^{N} g(\theta_i)$$

  - To obtain a marginal distribution, simply ignore the coordinates of the nuisance variables
  - Quantiles approximated by sample quantiles
  - Precision of the results is a function of how many points are in the sample; one is limited only by the speed of ones computer, ones patience and ones cleverness

Gibbs Sampling

- Suppose one has available the full conditional distributions

$$\pi(\theta_i \mid \theta_{-i}, x)$$

- Then a sample from the posterior distribution can be obtained (under some mild side conditions) by starting at any point $\theta^{(1)}$ in the sample space Θ and generating new points sequentially by drawing from the conditional distributions in turn, always conditioning on the most recent values of θ_j

$$\theta^{(0)} \in \Theta$$

$$\theta_j^{(n)} \sim \pi(\theta_j \mid \theta_{\hat{j}}^{(n-1)}, x),\ j = 1, \ldots, k$$

where

$$\theta_{\hat{j}}^{(n-1)} = (\theta_1^{(n)}, \theta_2^{(n)}, \ldots, \theta_{j-1}^{(n)}, \theta_{j+1}^{(n-1)}, \ldots, \theta_k^{(n-1)})$$

for $n = 1, \ldots, N$ to obtain a large sample of large size N

Metropolis-Hastings Sampling

- Gibbs sampling is possible only when you can generate samples conveniently from each of the full conditional distributions. Sometimes this is not possible, and under these conditions the Metropolis-Hastings method can be substituted for some (or all) of the Gibbs steps. It works as follows:
  - Generate a *candidate* θ^* from a *proposal distribution* $q(\theta^* | \theta)$. The distribution q is subject to some mild side conditions but can be quite arbitrary
  - Compute the quantity

$$\alpha = \min\left(\frac{\pi(\theta^* \mid x) q(\theta \mid \theta^*)}{\pi(\theta \mid x) q(\theta^* \mid \theta)}, 1 \right)$$

  - With probability α, accept θ^* as the next value of θ; otherwise assign the current value θ as the next value

Rejection Sampling

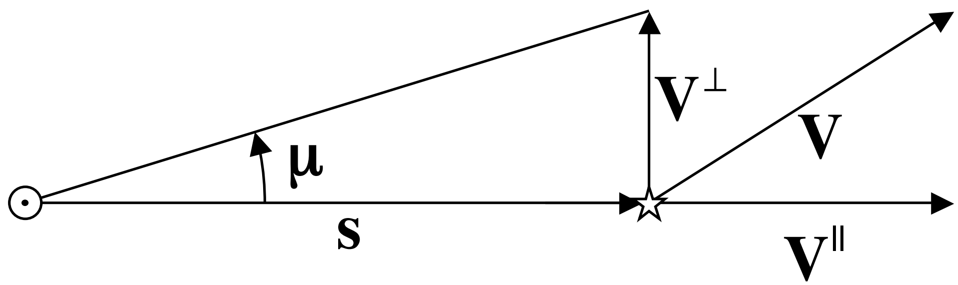
- One may not be able to sample directly from $\pi(\theta|x)$, but one may have a distribution $q(\theta)$ from which one can sample easily, subject to
  - support(π) $\subseteq$ support(q)
  - $\pi(\theta|x) \leq Mq(\theta)$ for all θ and some known constant M.
- Then a sample from $\pi(\theta|x)$ can be obtained by repeating the following steps until acceptance is obtained:
  - Generate a sample θ^* from $q(\theta^*)$
  - Accept θ^* as the desired sample with probability $\pi(\theta^*|x)/Mq(\theta^*)$
  - If θ^* is rejected, return to the first step. Repeat until acceptance is achieved.

Outline

- A little astronomy — what's the goal?
- Basic Bayesian Inference
- Markov Chain Monte Carlo
- Mathematical model and likelihood function
- Priors
- Sampling strategy
- Results
- Future research

The Likelihood

- Linear algebra and the geometry of the figure



yield for each star

$\hat{\mathbf{s}} = \mathbf{s}/s$ is the unit vector towards the star,

$$V^{\parallel} = \hat{\mathbf{s}}'\mathbf{V}$$

$$\hat{\mathbf{s}}'\mathbf{V}^{\perp} = \mathbf{0}$$

$$\mathbf{V} = (V_{\alpha}^{\perp}, V_{\delta}^{\perp}, V^{\parallel}) = \mathbf{V}^{\perp} + \mathbf{V}^{\parallel} = \mathbf{V}^{\perp} + \hat{\mathbf{s}}V^{\parallel}$$

$\mathbf{V}^{\parallel} = \hat{\mathbf{s}}\hat{\mathbf{s}}'\mathbf{V}$ (projects $\mathbf{V}$ onto $\hat{\mathbf{s}}$)

$\mathbf{V}^{\perp} = \mathbf{V} - \mathbf{V}^{\parallel} = (\mathbf{I} - \hat{\mathbf{s}}\hat{\mathbf{s}}')\mathbf{V}$ (projects $\mathbf{V}$ onto plane of sky)

The Likelihood

- If $\boldsymbol{\mu}^{\text{o}}=(\mu_\alpha{}^{\text{o}}, \mu_\delta{}^{\text{o}})$are the two components of the *observed* proper motion vector (in the plane of the sky), and ρ^{o} is the *observed* radial velocity, we assume

$$\mu_\alpha^{\text{o}} \sim N(V_\alpha^{\perp}/s, \sigma_{\mu_\delta}^2)$$

$$\mu_\delta^{\text{o}} \sim N(V_\delta^{\perp}/s, \sigma_{\mu_\delta}^2)$$

$$\rho^{\text{o}} \sim N(V^{\parallel}, \sigma_\rho^2)$$

- The variances are measured at the telescope and assumed *known perfectly*. The variables without superscripts are the “true” values.
- The likelihood is just the product of all these normal distributions over all stars in the sample

Outline

- A little astronomy—what's the goal?
- Basic Bayesian Inference
- Markov Chain Monte Carlo
- Mathematical model and likelihood function
- Priors
- Sampling strategy
- Results
- Future research

Quantities Being Estimated

- The posterior distribution is a function of the following quantities:
  - The true proper motions $\boldsymbol{\mu}_i = \mathbf{V}_i^{\perp} / s_i$ and radial velocities $\rho_i = V_i^{\parallel}$ of the stars where $\mathbf{V}_i = s_i \boldsymbol{\mu}_i + \hat{\mathbf{s}}_i \rho_i$
  - The mean velocity $\mathbf{V_0}$ of the stars relative to the Sun
  - The three-dimensional covariance matrix **W** that describes the distribution of the stellar velocities
  - The distance s_i to each star, (in turn a function of each star's associated absolute magnitude M_i)
  - The mean absolute magnitude M of the RR Lyrae stars as a group

Priors Defining the Hierarchical Model

- The priors on the $\mathbf{V}_i$ are assigned by assuming that the velocities of the individual stars are drawn from a (three-dimensional) multivariate normal distribution:

$$\mathbf{V}_i \mid \mathbf{V_0}, \mathbf{W} \sim N(\mathbf{V_0}, \mathbf{W})$$

- We choose an improper flat prior on $\mathbf{V_0}$, the mean velocity of the RR Lyrae stars relative to the velocity of the Sun

Priors Defining the Hierarchical Model

- The prior on **W** is requires some thought. One wants a proper posterior distribution (flat priors are improper, for example, but the posterior distribution must be proper)
- Another desirable feature is that estimators derived from the analysis should have good frequentist properties. In particular, we would like them to be *admissible* under an appropriate loss (e.g., here, quadratic loss)
- To satisfy these desiderata, while maintaining a reasonably noncommittal (vague) prior, we chose a "hierarchical independence Jeffreys prior" on **W**, which for a three-dimensional distribution implies

$$\pi(\mathbf{W}) \propto |\mathbf{I} + \mathbf{W}|^{-2}$$

Priors Defining the Hierarchical Model

- The distances s_i are related to the M_i and m_i by a defined (exact) relationship:

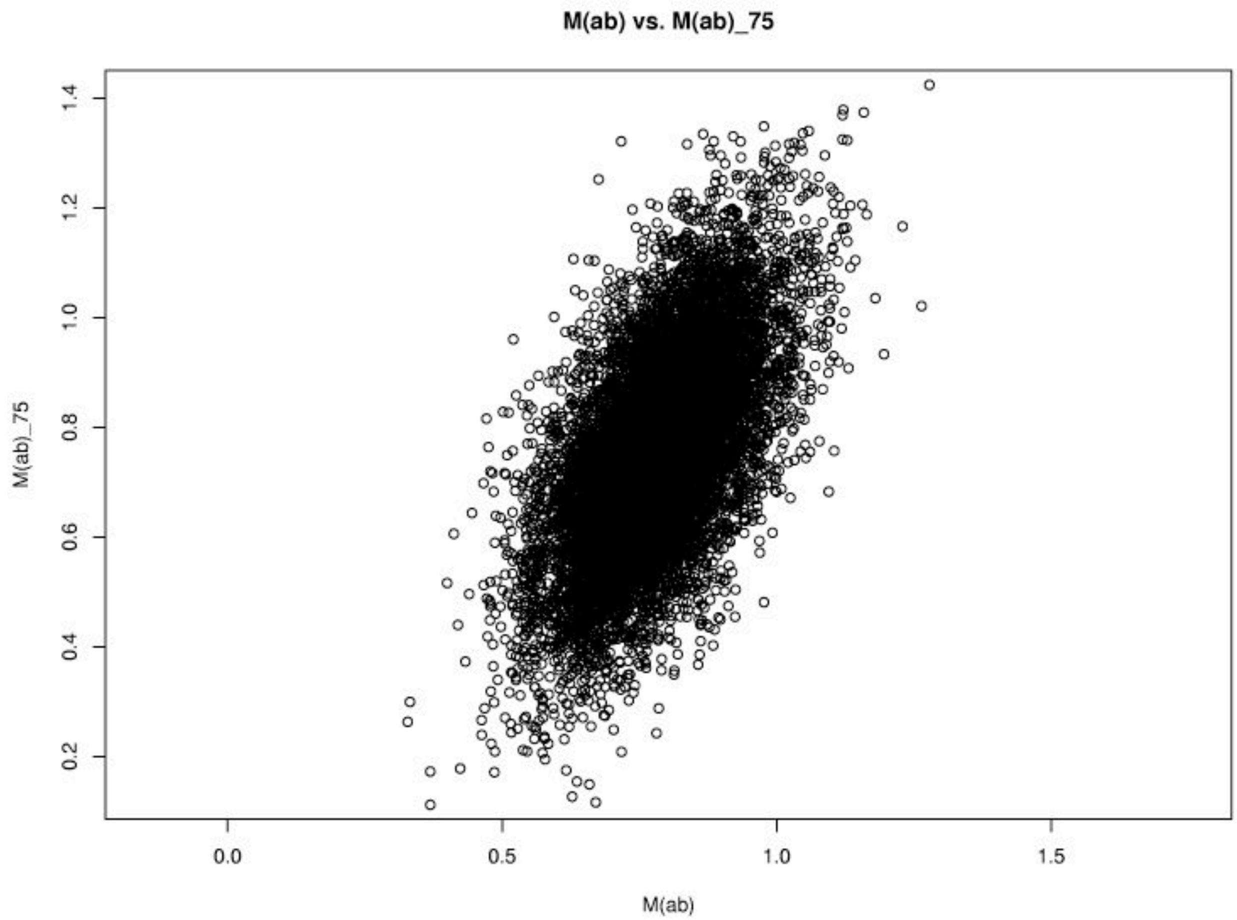
$$s_i = 10^{0.2(m_i - M_i + 5)}$$

- We introduce new variables $U_i = M_i - M$, because preliminary calculations indicated that sampling would be more efficient on the variables U_i (the M_i are highly correlated with each other and with M, but the U_i are not). Therefore we write

$$s_i = 10^{0.2(m_i - M - U_i + 5)}$$

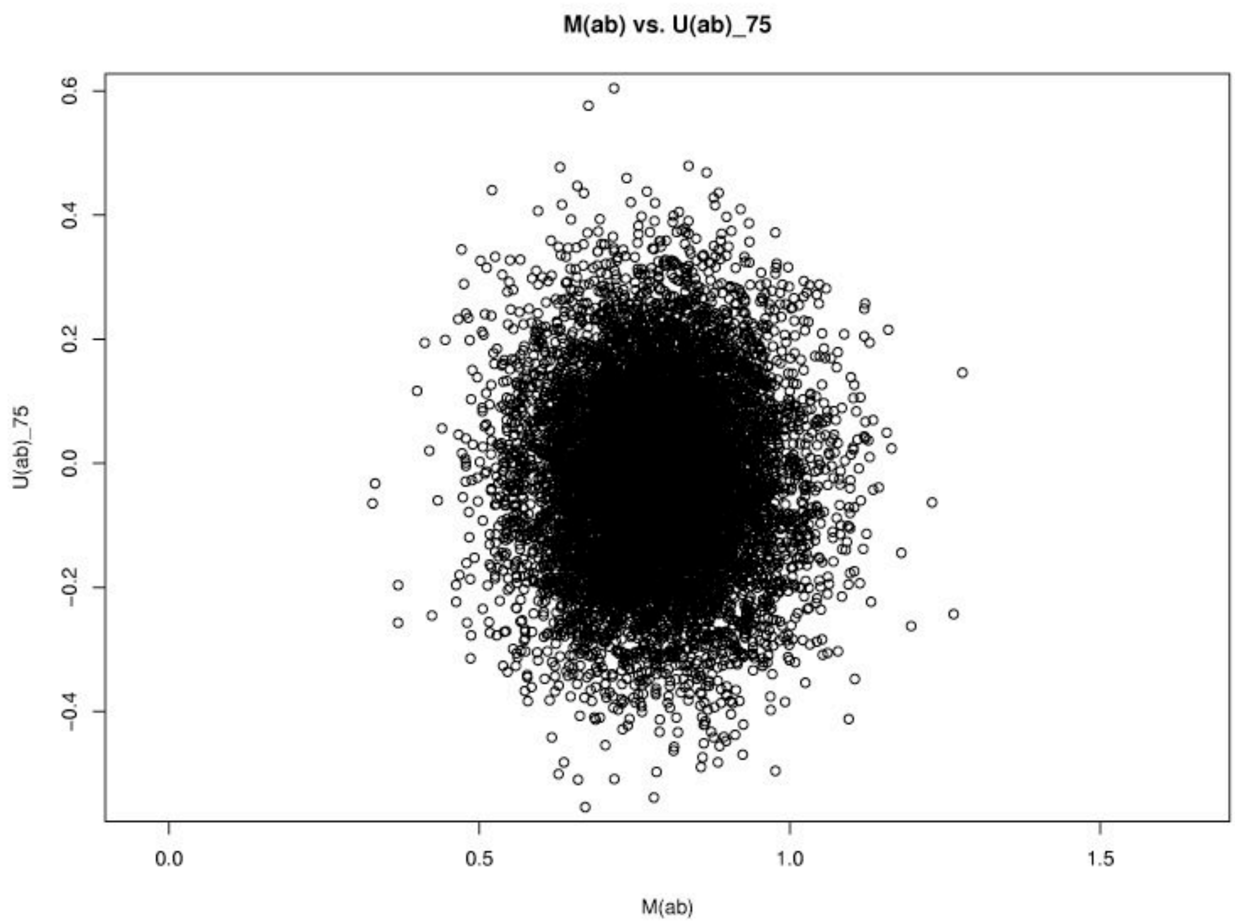
Priors Defining the Hierarchical Model

- Illustrating the correlation between M and one of the M_i. Note that the problem is much more acute with many variables



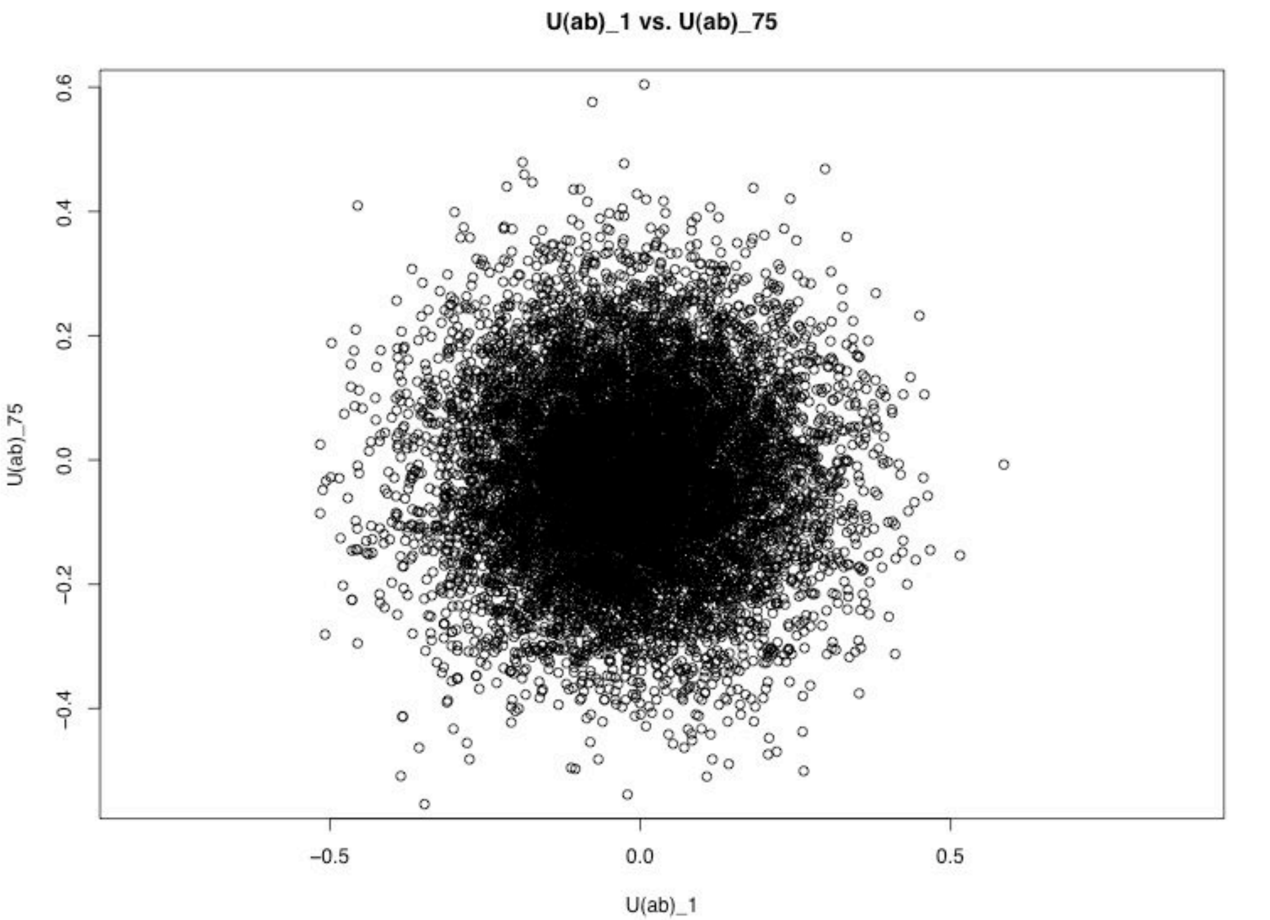
Priors Defining the Hierarchical Model

- Illustrating low correlation between M and one of the U_i



Priors Defining the Hierarchical Model

- Illustrating low correlation between two of the U_i



Priors Defining the Hierarchical Model

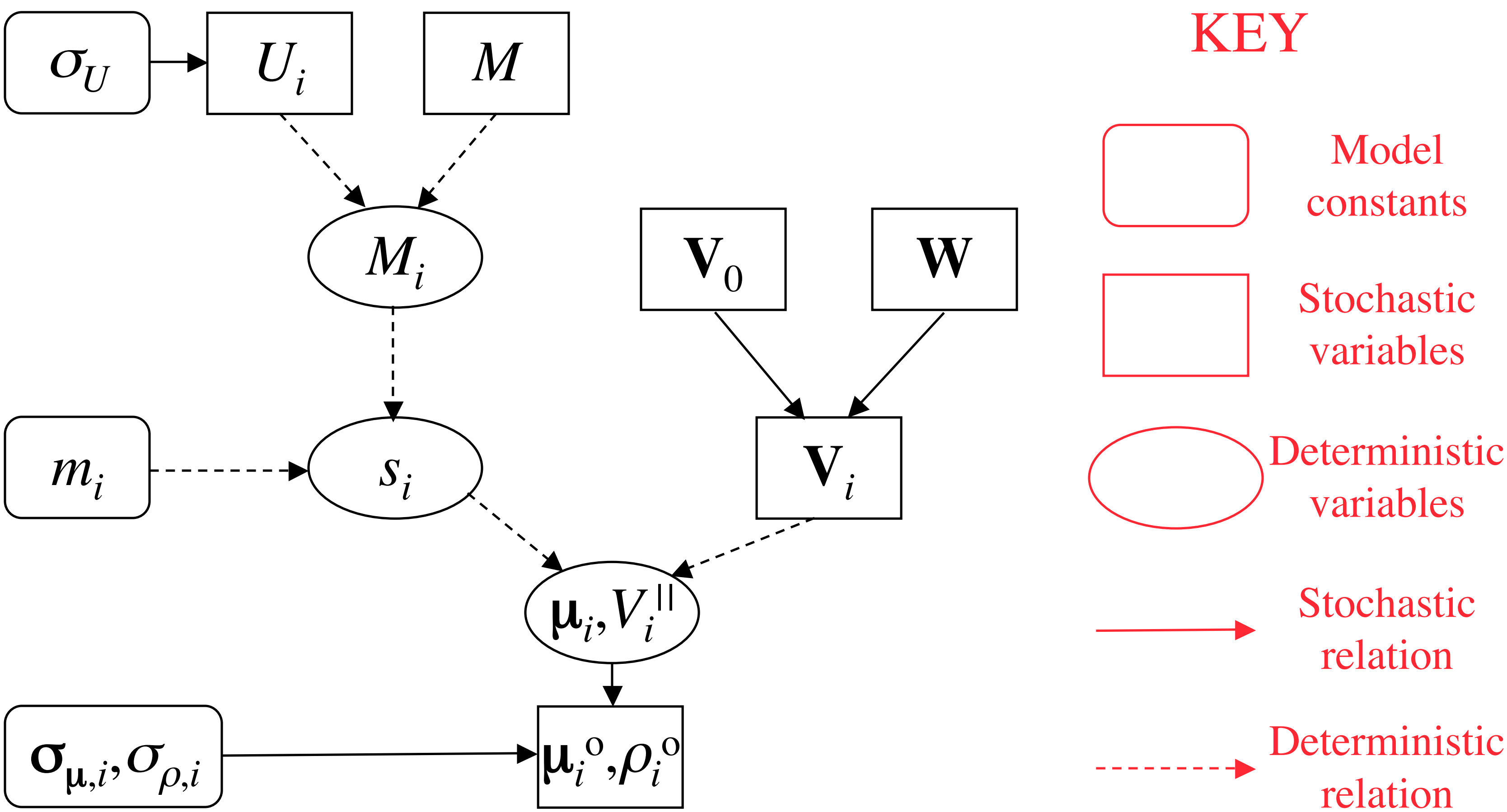
- We choose a flat prior on M (we have also tried a somewhat informative prior based on known data with not much change).
- Evidence from other sources (e.g., studies of RR Lyrae stars in clusters) indicates a "cosmic scatter" of about 0.15 magnitudes in M_i. Thus a prior on U_i of the form

$$U_i \sim N(0,(0.15)^2)$$

seems appropriate.

Priors Defining the Hierarchical Model

- The overall structure of the hierarchical model can be summarized by this DAG



Outline

- A little astronomy—what's the goal?
- Basic Bayesian Inference
- Markov Chain Monte Carlo
- Mathematical model and likelihood function
- Priors
- Sampling strategy
- Results
- Future research

Sampling Strategy

- We can use Gibbs sampling to sample on $\mathbf{V_0}$

$$\mathbf{V_0} \mid \{\mathbf{V}_i\}, \mathbf{W} \sim N(\overline{\mathbf{V}}, \mathbf{W}/N)$$

where N is the number of stars in our sample and

$$\overline{\mathbf{V}} = \frac{1}{N}\sum \mathbf{V_i}$$

Sampling Strategy

- We can also use Gibbs sampling to sample on **W**; however, we cannot generate **W** directly, but instead use a rejection sampling scheme. Generate a candidate $\mathbf{W}^*$ using

$$\mathbf{W}^* \mid \{\mathbf{V}_i\}, \mathbf{V_0} \sim \text{InverseWishart}(\mathbf{T}, df = N)$$

where

$$\mathbf{T} = \sum (\mathbf{V}_i - \mathbf{V_0})(\mathbf{V}_i - \mathbf{V_0})'$$

- Then accept or reject the candidate with probability

$$|\mathbf{W}^*|^2 / |\mathbf{I} + \mathbf{W}^*|^2$$

- Repeat until successful. This scheme is very efficient when the degrees of freedom is large

Sampling Strategy

- $\mathbf{V}_i = s_i\boldsymbol{\mu}_i + \hat{\mathbf{s}}_i\rho_i$ can be sampled in a Gibbs step; it involves data on the individual stars as well as the general velocity distribution, and the covariance matrices $\mathbf{S}_i$ of the velocities of the individual stars depend on s_i because of the relation $s\boldsymbol{\mu} = \mathbf{V}^{\perp}$
- Omitting the algebraic details, the resulting sampling scheme is

$$\mathbf{V}_i \sim N(\mathbf{u}_i, \mathbf{Z}_i)$$

where

$$\mathbf{Z}_i^{-1} = \mathbf{S}_i^{-1} + \mathbf{W}_i^{-1}$$

$$\mathbf{u}_i = \mathbf{Z}_i(\mathbf{S}_i^{-1}\mathbf{V}_i^{\mathbf{o}} + \mathbf{W}^{-1}\mathbf{V}_{\mathbf{0}})$$

$$\mathbf{V}_i^{\mathbf{o}} = s_i\boldsymbol{\mu}_i^{\mathbf{o}} + \hat{\mathbf{s}}_i\rho_i^{\mathbf{o}}$$

Sampling Strategy

- $\mathbf{V}_i = s_i\boldsymbol{\mu}_i + \hat{\mathbf{s}}_i\rho_i$ can be sampled in a Gibbs step; it involves data on the individual stars as well as the general velocity distribution, and the covariance matrices $\mathbf{S}_i$ of the velocities of the individual stars depend on s_i because of the relation $s\boldsymbol{\mu} = \mathbf{V}^{\perp}$
- Omitting the algebraic details, the resulting sampling scheme is

$$\mathbf{V}_i \sim N(\mathbf{u}_i, \mathbf{Z}_i)$$

where

$$\mathbf{Z}_i^{-1} = \mathbf{S}_i^{-1} + \mathbf{W}_i^{-1}$$ Precision=Sum of Precisions

$$\mathbf{u}_i = \mathbf{Z}_i(\mathbf{S}_i^{-1}\mathbf{V}_i^{\mathbf{o}} + \mathbf{W}^{-1}\mathbf{V}_{\mathbf{0}})$$ Weighted Average

$$\mathbf{V}_i^{\mathbf{o}} = s_i\boldsymbol{\mu}_i^{\mathbf{o}} + \hat{\mathbf{s}}_i\rho_i^{\mathbf{o}}$$

Sampling Strategy

- We sample the U_i using a Metropolis-Hastings step. Our proposal is $U_i^* \sim N(U_i, w)$ with an appropriate w, adjusted for good mixing. Recalling that s_i is a function of U_i, the conditional distribution is proportional to

$$s_i^2 N(s_i \boldsymbol{\mu}_i + \hat{\mathbf{s}}_i V_i^{\parallel} - \mathbf{V}_0, \mathbf{W})$$

with the informative prior on U_i we described earlier,

$$U_i \sim N(0, (0.15)^2)$$

Sampling Strategy

- We sample on M using a Metropolis-Hastings step. Our proposal for M^* is a t distribution centered on M with an appropriate choice of degrees of freedom and variance, adjusted for good mixing. Again recalling that s_i is a function of M, the conditional distribution is proportional to

$$\prod s_i^2 N(s_i \boldsymbol{\mu}_i + \hat{\mathbf{s}}_i V_i^{\parallel} - \mathbf{V_0}, \mathbf{W})$$

  with a prior on M that may or may not be informative (as discussed earlier).
- We found that df=10 and variance=0.01 on the t proposal distribution mixed well.

Outline

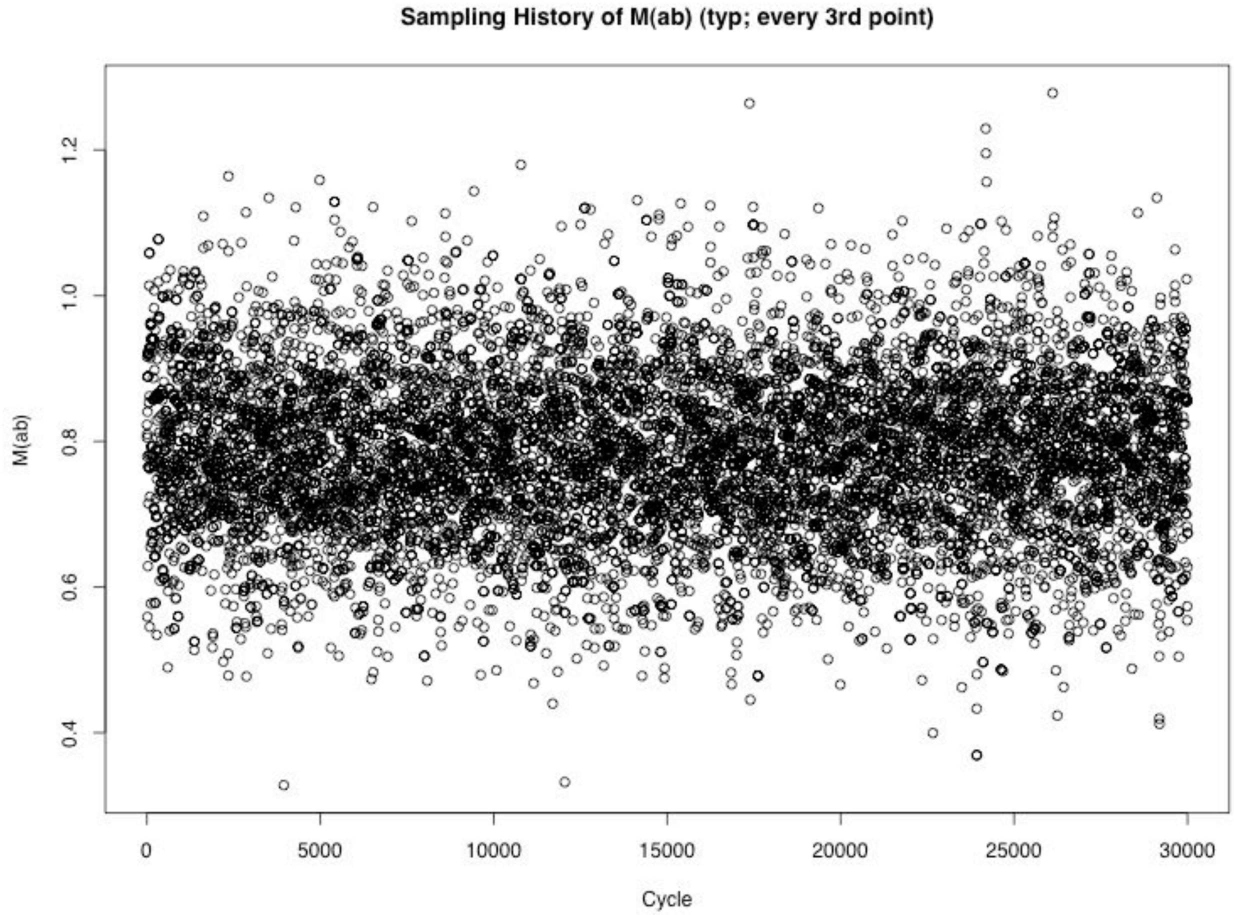
- A little astronomy — what's the goal?
- Basic Bayesian Inference
- Markov Chain Monte Carlo
- Mathematical model and likelihood function
- Priors
- Sampling strategy
- Results
- Future research

Results on M

- Our key astrophysically interesting result is M.
- We look at
  - Plot of samples on M
  - Histogram of M
  - Mean, variance of M

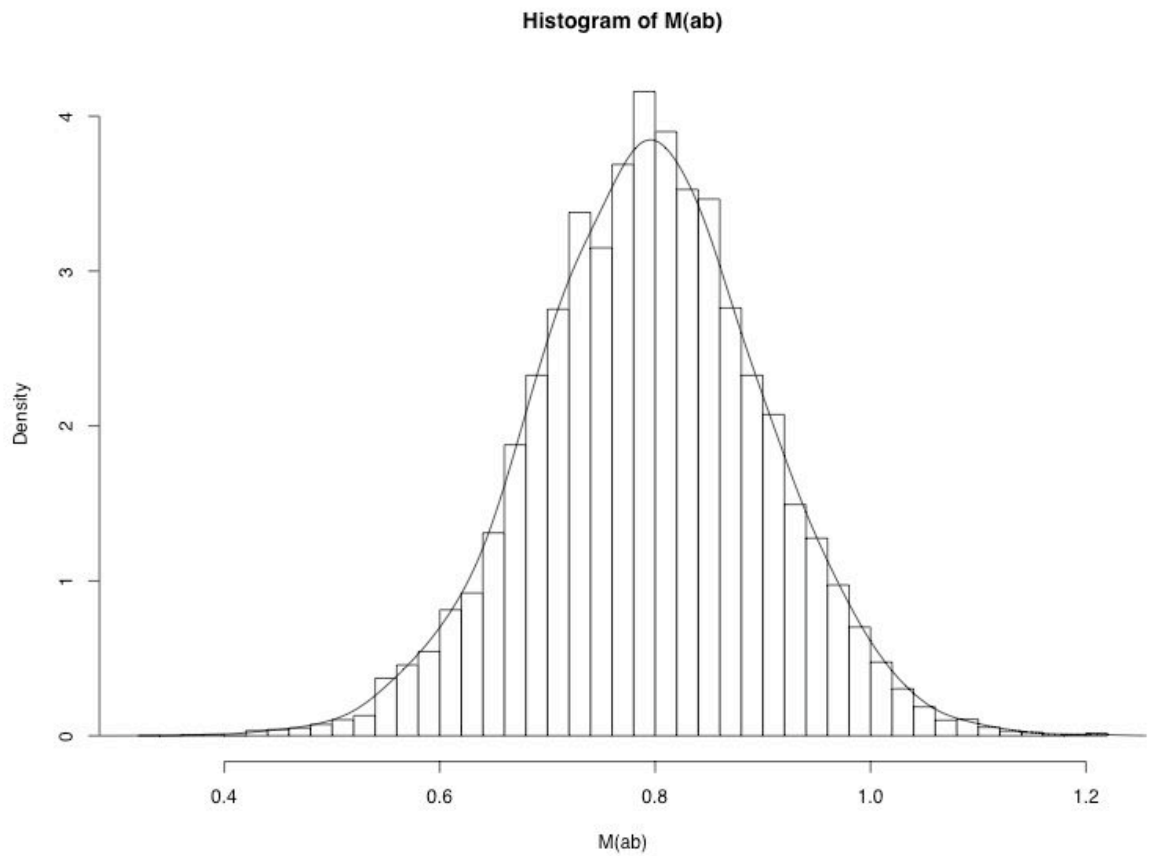
Sampling History of M(ab)

- The samples on M show good mixing



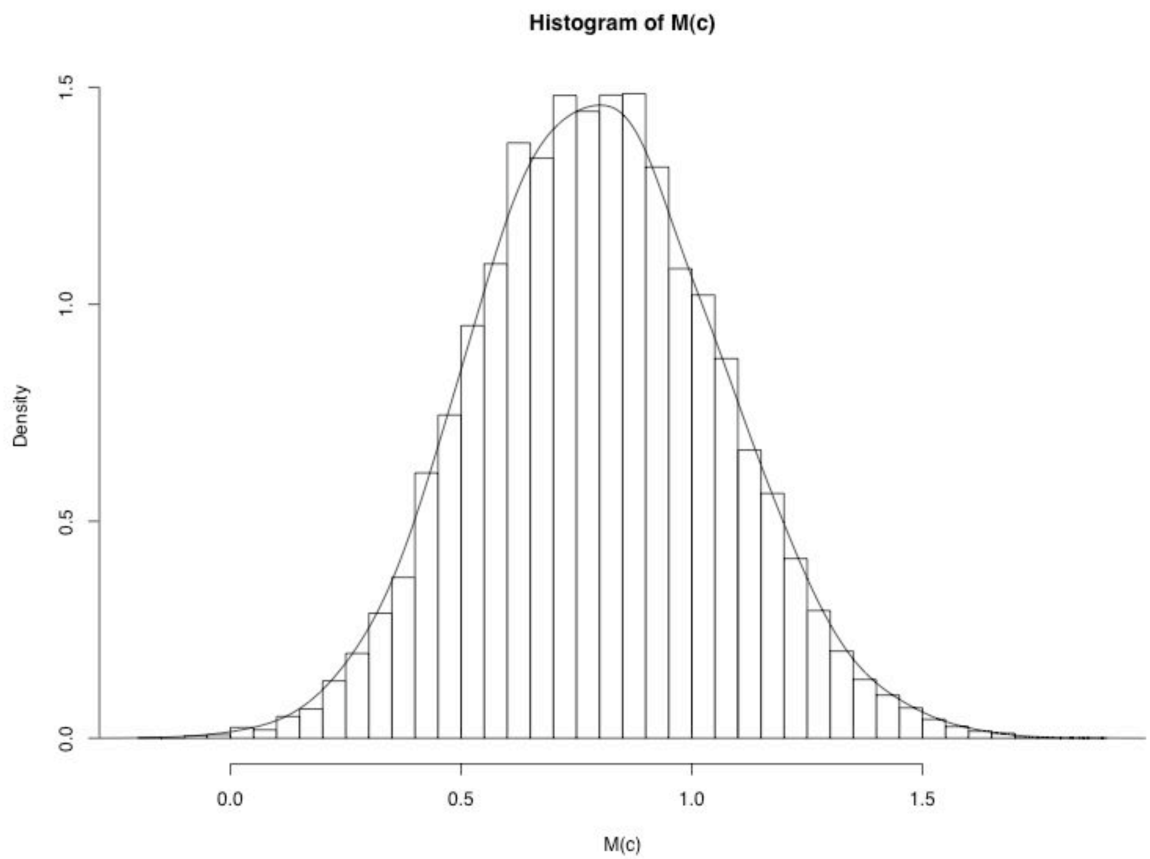
Marginal Density of M(ab)

- Plotting a smoothed histogram of *M* displays the posterior marginal distribution of *M*



Marginal Density of M(c)

- The same, for the overtone pulsators



Results on M

- Hawley et. al. used a maximum likelihood technique to study this problem. For comparison, we used their data.
- They ran cases with a number of subsets of stars, broken down by various criteria. We look only at the "ab" stars (normal pulsators, N=141) and the "c" stars (overtone pulsators, N=17).

	M (ab stars)
This study	0.79 ± 0.11
Hawley *et. al.*	0.76 ± 0.14
	M (c stars)
This study	0.75 ± 0.27
Hawley *et. al.*	1.09 ± 0.38

Other Methods for *M*

- The data of Hawley *et*. *al*. had some incorrect apparent magnitudes. Reanalysis with corrected data (not available to us for this study) shows that the value of *M* should be decreased by approximately 0.08 magnitudes.
- The best direct measurement of the distance of an RR Lyrae star, by the Hubble Space Telescope, gives M=0.61±0.10 magnitudes.
- The results of Skillen *et*. *al*., using the Surface Brightness method (similar to the method used by Barnes *et*. *al*. and reported at ISBA 2000) give M=0.65±0.10 magnitudes.

Results

- Other results of interest are
  - The reflex solar motion $\mathbf{V}_0$. This informs us of the mean velocity of the ensemble of RR Lyrae stars relative to the Sun (Units are km/sec).

ab stars	V_ϖ	V_θ	V_z
This study	-12 ± 11	-136 ± 9	-8 ± 7
Hawley *et. al.*	-10 ± 13	-155 ± 12	-9 ± 8
c stars			
This study	-26 ± 25	-113 ± 26	-2 ± 12
Hawley *et. al.*	-26 ± 25	-124 ± 25	-6 ± 13

Results

- Other results of interest are
  - Velocity Ellipsoid: This tells us how the galactocentric orbits of the RR Lyrae stars are distributed in a statistical sense, as described by the covariance matrix of the velocities. In galactic cylindrical coordinates (ϖ,θ,z) the matrix is believed roughly diagonal, with the on-diagonal dispersions decreasing from ϖ to θ to z. (Units are km/sec, see next chart).
  - There is a significant difference in the estimated standard deviations between this study and Hawley *et. al*. Hawley has stated (private communication) that she is not confident of her estimates.

Results

- Other results of interest are
  - Velocity Ellipsoid

ab stars	σ_{ϖ}	σ_{θ}	σ_{z}
This Study	133 ± 9	109 ± 7	75 ± 5
Hawley *et. al.*	150 ± 59	120 ± 47	87 ± 33
c stars			
This Study	101 ± 22	106 ± 23	46 ± 11
Hawley *et. al.*	101 ± 57	71	51

Results

- Compare with
  - Hawley, *et. al.*
    - The sample of "ab" stars agrees within the errors with the analysis of Hawley *et. al*. However, the "c" sample gives a discrepant value of M. Their value is high, compared to ours.
    - The reason is unknown, but in their analysis of these data they did not solve for the full covariance matrix of the velocities. Instead, they set the off-diagonal terms to zero, and fixed the ratios for the on-diagonal terms to that given by the reduction of the larger "ab" data.

Sampling on U_i

- Dominated by prior, seems that they are unidentified in this model
- J. Berger raised concerns about leaving it out, so we put it in; results essentially unchanged

Outline

- A little astronomy — what's the goal?
- Basic Bayesian Inference
- Markov Chain Monte Carlo
- Mathematical model and likelihood function
- Priors
- Sampling strategy
- Results
- Future research

Future Research

- We want to investigate the effect of metallicity on the results. It is known that M depends on this, but we want to see if our data can detect this effect
- There is a much larger and better sample of proper motions from the HIPPARCOS satellite, which together with new radial velocities should provide better results.