Intro to Bayesian Hierarchical Modeling

June 11, 2025

UTSSRP

Prof. Gwendolyn Eadie

Basics of a Hierarchical Model

- One way to think about it: *adding layers to the Bayesian model*

- In the m&m's example, we had the posterior

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Where our **prior** was $p(\theta) \propto Beta(\alpha, \beta)$ and α and β were chosen as fixed values.

- But we could have also set a *hyperprior distribution* on α and β, and then we'd have

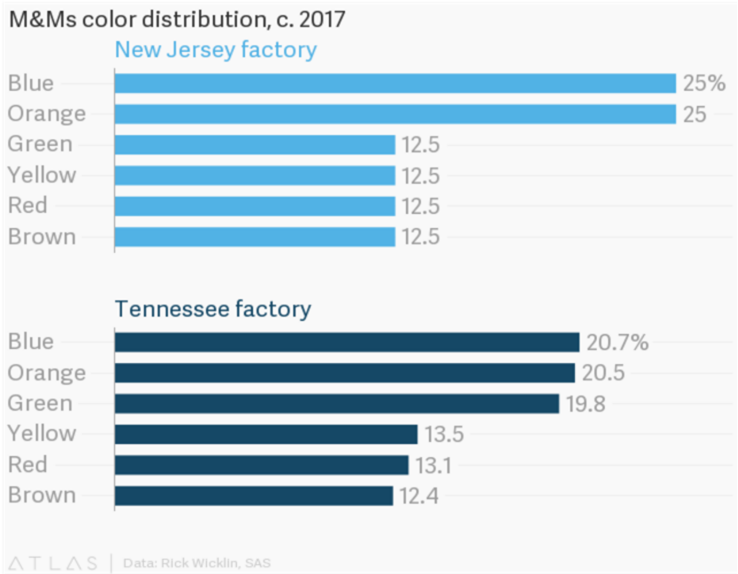
$$p(\theta, \alpha, \beta|y) \propto p(y|\theta)p(\theta|\alpha, \beta)\boldsymbol{p(\alpha, \beta)}$$

- This is now a hierarchical model.

Basics of a Hierarchical Model

- Another reason to do hierarchical modeling: *want to infer parameters at different levels in the hierarchy, and account for structure in the variation*

- In N. America, two factories make m&m's



Imagine we had m bags from New Jersey and q bags from Tennessee, but we only know that $\mathrm{m} + q = 45$.

What should we do if we want to know

1. what the colour distribution of the m&m's made in each factory is?

   AND

2. The value of m?

Again, a hierarchical model will work here.

Hierarchical Model Example

- The proportion of blues made at the factory may vary from one day to the next (random fluctuation)
  - The **true** θ (% of blue) changes subtly over time

- Imagine:
  - Each time I do the m&m's exercise with the class, I buy a box of m&m's.
  - I always buy boxes from the same country and factory (only one factory)

- Now I want to infer the variation in θ from class to class.

→ To do this, we need to *estimate* α and β

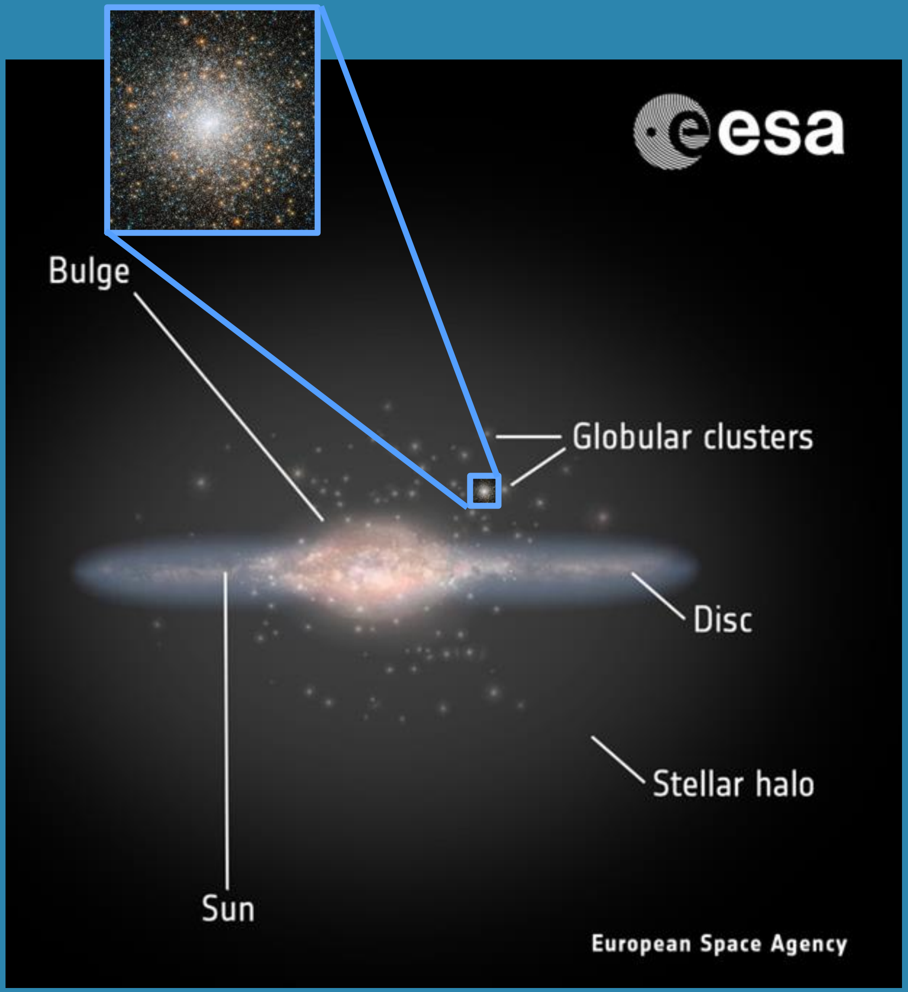
- Set a *hyperprior distribution* on α and β, and then we'd have

$$p(\theta,\alpha,\beta|y) \propto p(y|\theta)\boldsymbol{p(\theta|\alpha,\beta)}\boldsymbol{p(\alpha,\beta)}$$

- This is now a hierarchical model.

Detailed Example from Astronomy Research

Globular Cluster (GC)



Sketch of Milky Way

Estimating the mass of the Milky Way

Using hierarchical Bayes and "kinematic tracers"

Hierarchical Bayesian Model for MW Mass Estimate in Pictures

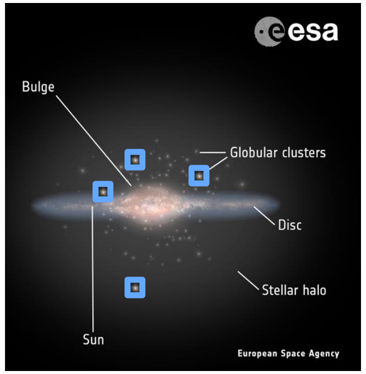
Likelihood



...


Each GC has Individual parameters:
- True position
- True velocity

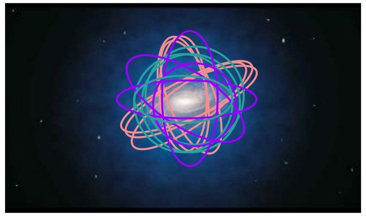
Prior



Shared population parameters for galaxy:

- Spatial density of GCs
- Gravitational potential
- Velocity anisotropy

Hyperprior



Hyperparameters:
- Bounds for model parameters
- Mean and variance for parameters

Velocity Anisotropy Parameter

$$\beta = 1 - \frac{\sigma_\theta^2 + \sigma_\phi^2}{2\sigma_r^2}$$

Hierarchical Bayesian Model for Milky Way

Posterior Distribution $\propto$ Likelihood $\cdot$ Prior $\cdot$ Hyperprior

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \Delta) \propto \prod_{i}^{N} \mathcal{L}(\boldsymbol{y_i}|\vartheta_i, \Delta_i)\, p(h(\vartheta_i)|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

measurement model

physical model

Priors on model parameters

Likelihood

$$\prod_{i}^{N} \mathcal{L}(\boldsymbol{y_i} | \vartheta_i, \Delta_i)$$

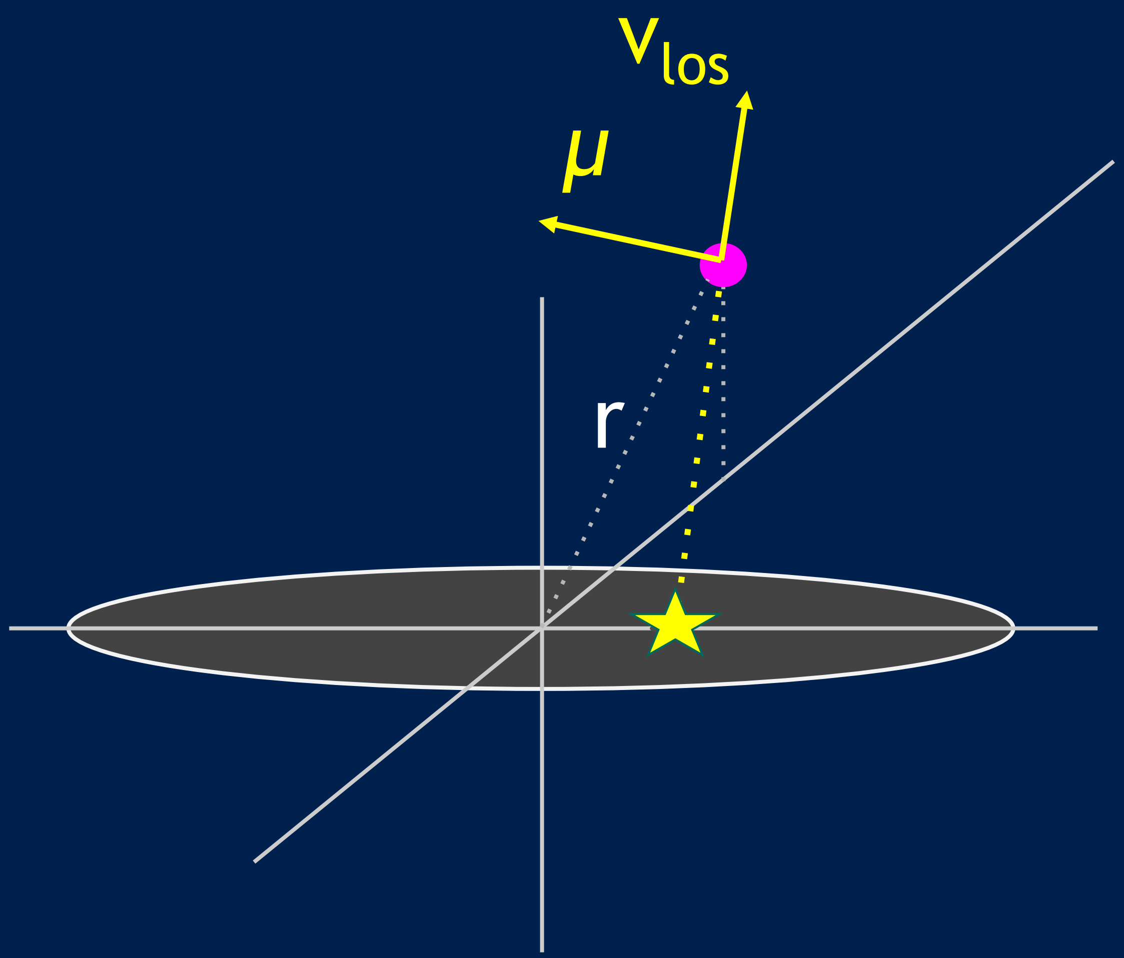
measurement model

Likelihood

$$\prod_{i}^{N} \mathcal{L}(\boldsymbol{y_i}|\vartheta_i, \Delta_i)$$

measurement model

Velocity components observed (data):



Likelihood

$$\prod_{i}^{N} \mathcal{L}(\boldsymbol{y_i}|\vartheta_i, \Delta_i)$$

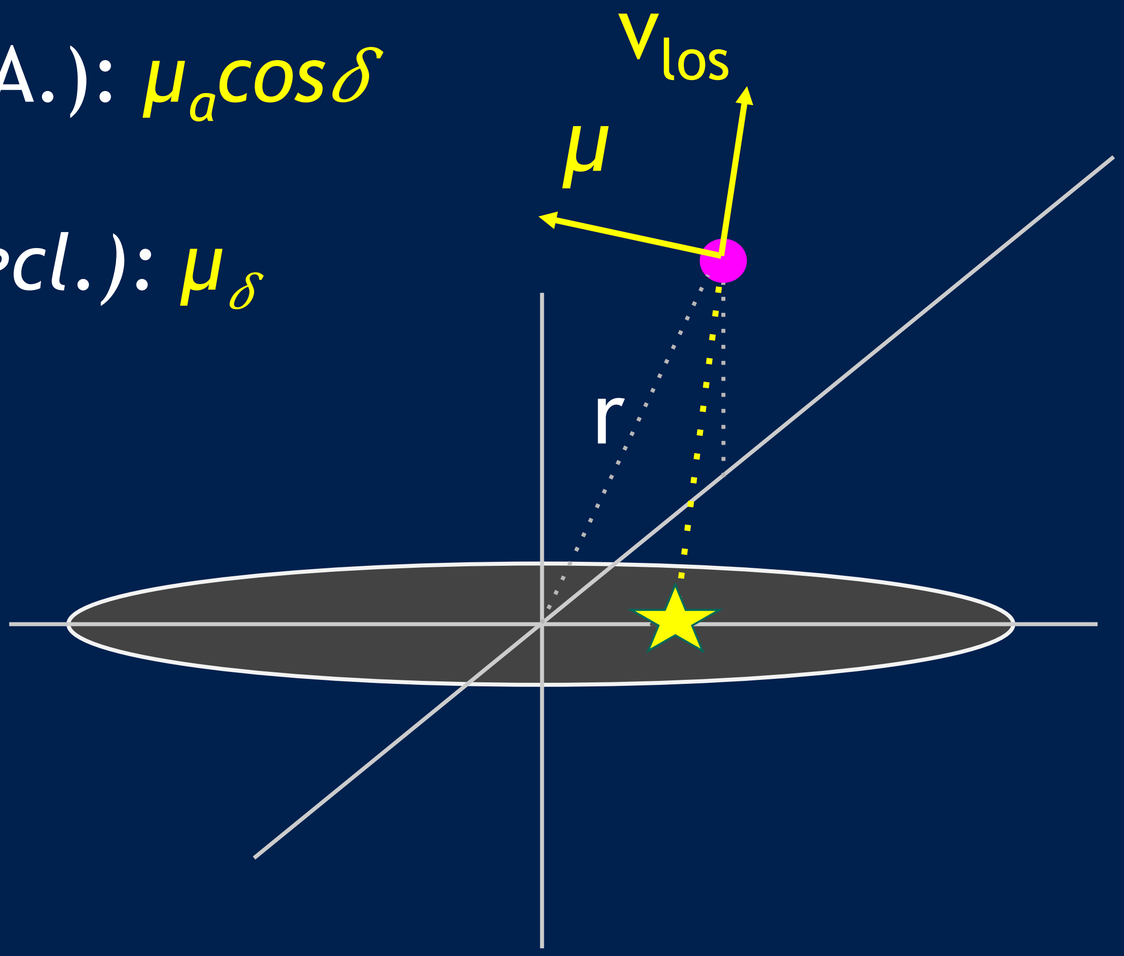
measurement model

Velocity components observed (data):

Line-of-sight velocity: v_{los}

Proper motion (R.A.): $\mu_a cos\delta$

Proper motion (Decl.): μ_δ



Likelihood

$$\prod_i^N \mathcal{L}(y_i|\vartheta_i, \Delta_i)$$

measurement model

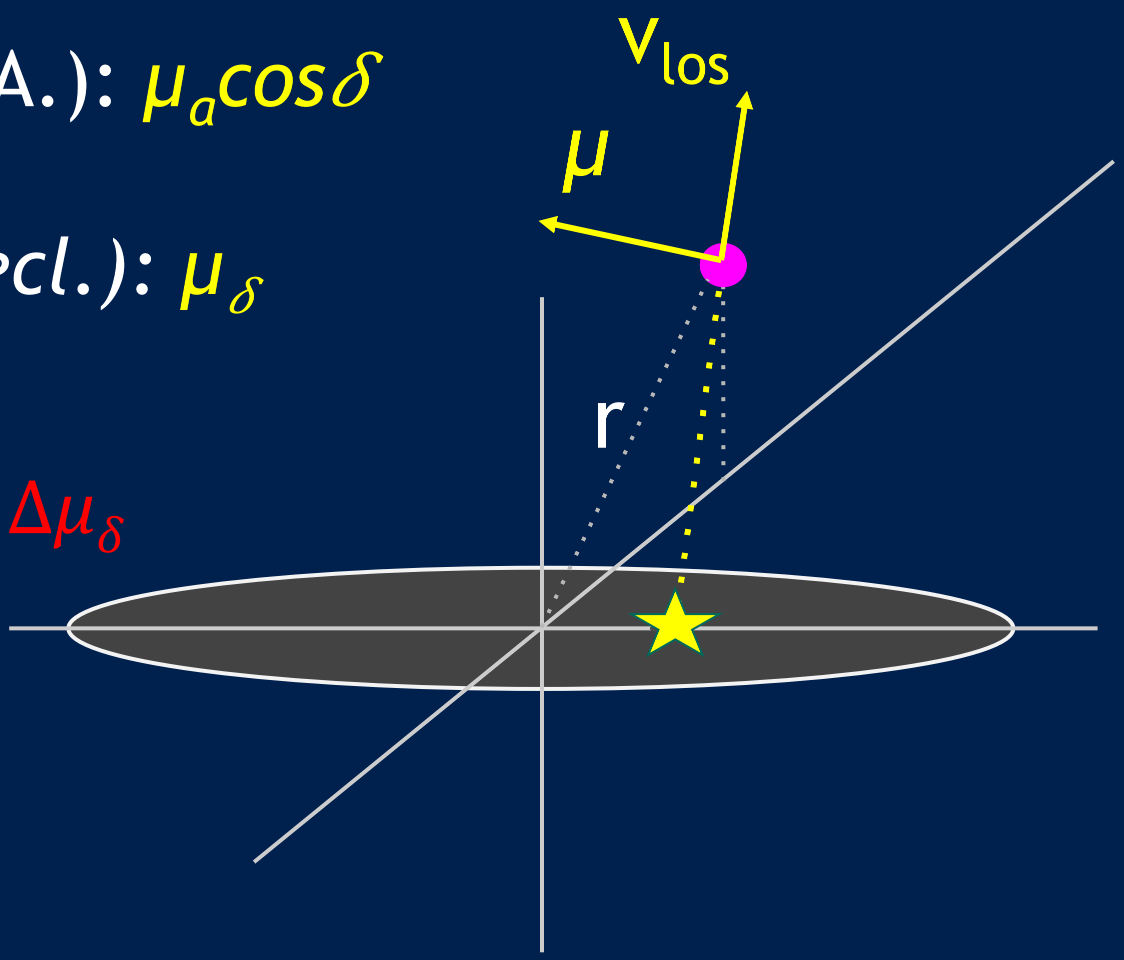
Velocity components observed (data):

Line-of-sight velocity: v_{los}

Proper motion (R.A.): $\mu_a \cos\delta$

Proper motion (Decl.): μ_δ

"Known" measurement uncertainties: $\Delta\mu_\delta$



Likelihood

$$\prod_i^N \mathcal{L}(\boldsymbol{y_i}|\vartheta_i, \Delta_i)$$

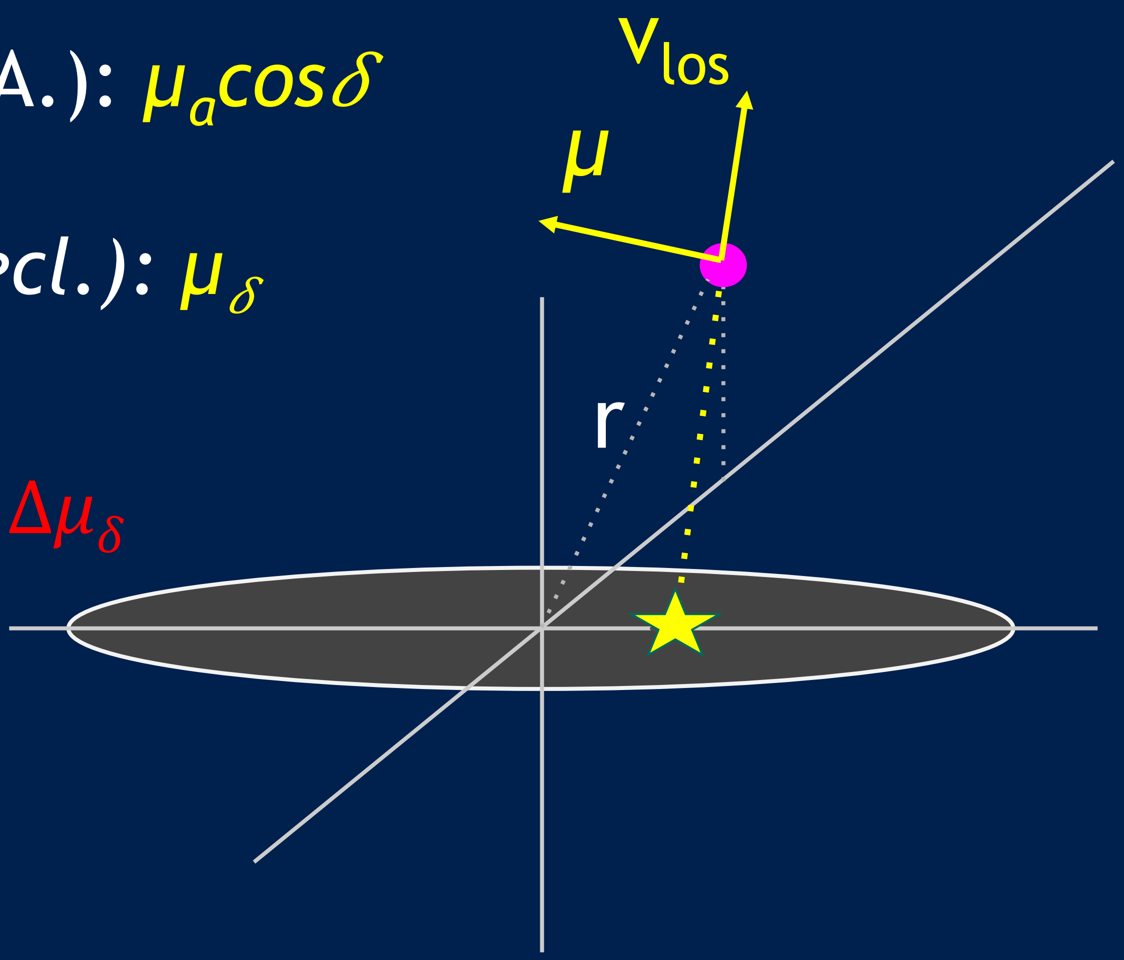
measurement model

Velocity components observed (data):

Line-of-sight velocity: v_{los}

Proper motion (R.A.): $\mu_a cos\delta$

Proper motion (Decl.): μ_δ



“Known” measurement uncertainties: $\Delta\mu_\delta$

$$\mu_\delta \sim N(\mu_\delta, \Delta\mu_\delta)$$

Hierarchical Bayesian Model

Posterior Distribution ∝ Likelihood · Prior · Hyperprior

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \Delta) \propto \prod_{i}^{N} \mathcal{L}(\boldsymbol{y_i}|\vartheta_i, \Delta_i)\, p(h(\vartheta_i)|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

measurement model

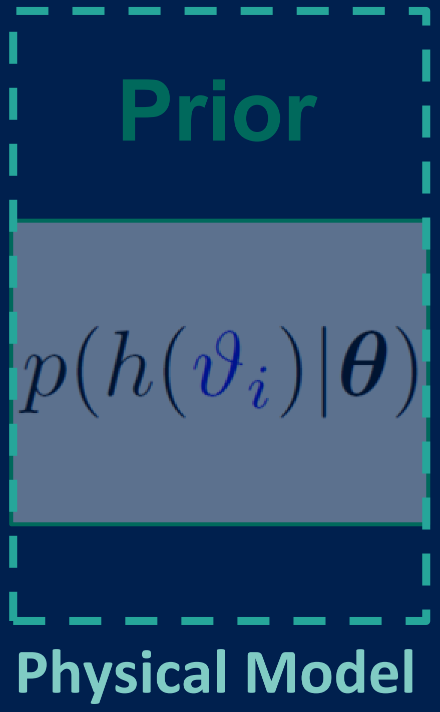
physical model

Priors on model parameters

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

physical model

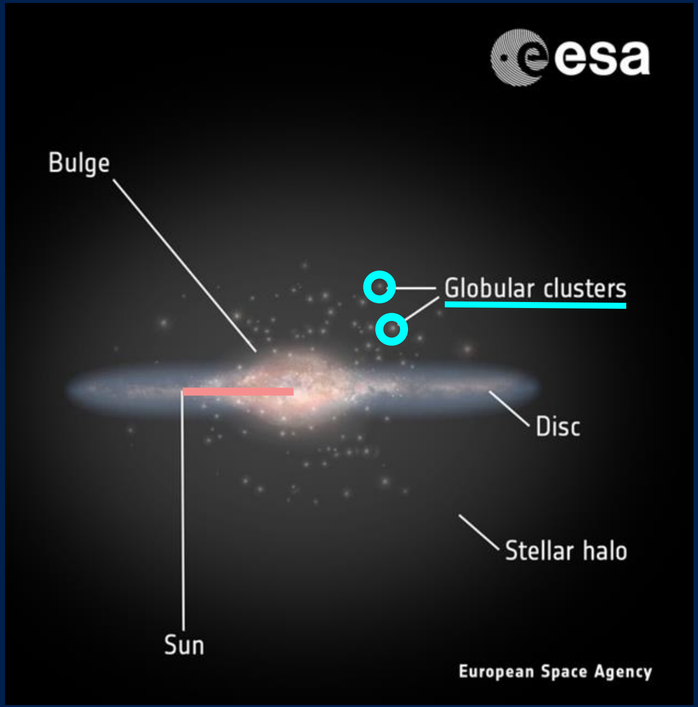


Prior
$p(h(\vartheta_i)|\boldsymbol{\theta})$
Physical Model

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model

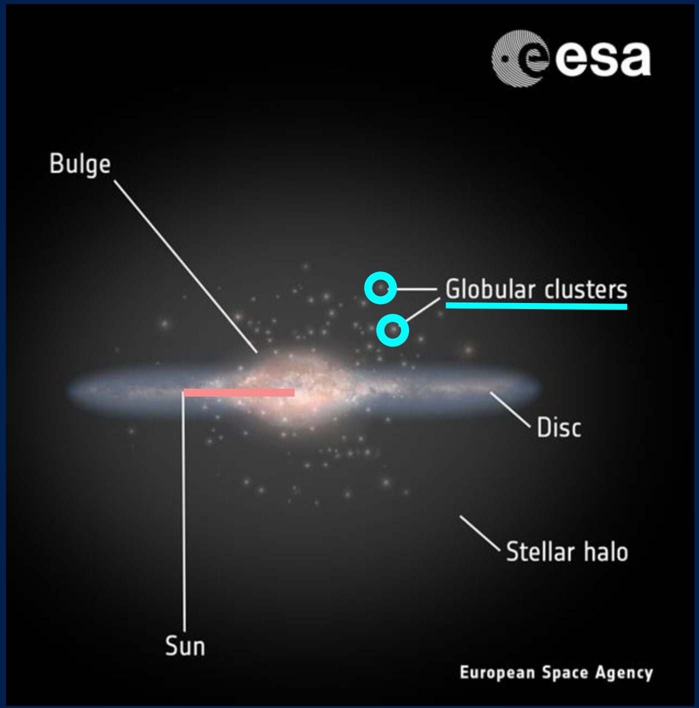


...

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model



Specific energy:

$$\mathcal{E} = -\frac{1}{2}(v_r^2 + v_t^2) + \Phi(r)$$

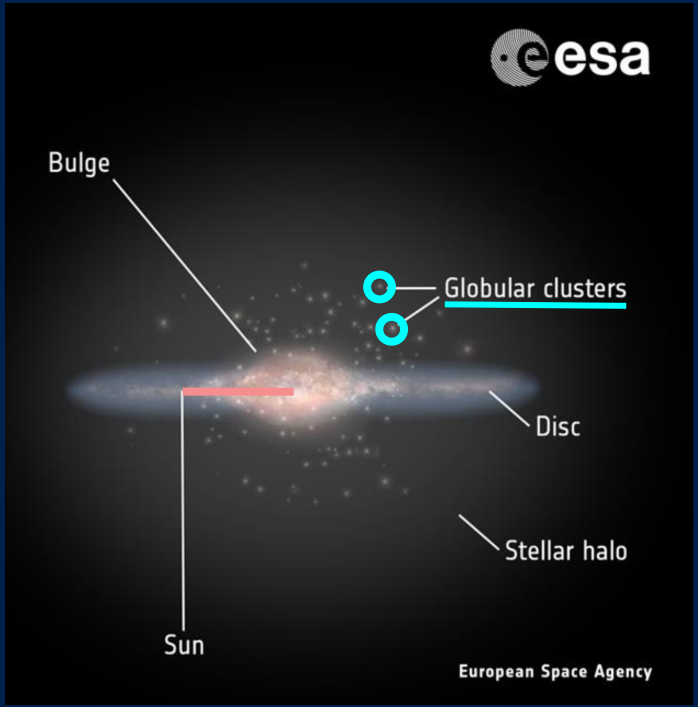
Angular momentum:

$$L = r v_t$$

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model



Specific energy:

$$\mathcal{E} = -\frac{1}{2}(v_r^2 + v_t^2) + \Phi(r)$$

Angular momentum:

$$L = rv_t$$

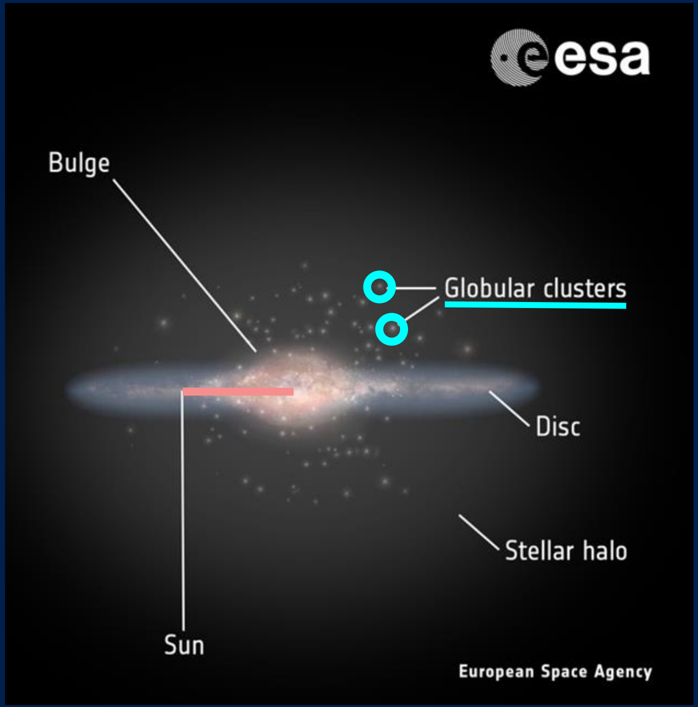
Probability density for specific energy and angular momentum

$$f(\mathcal{E}, L)$$

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model



Specific energy:

$$\mathcal{E} = -\frac{1}{2}(v_r^2 + v_t^2) + \Phi(r)$$

Angular momentum:

$$L = r v_t$$

Probability density for specific energy and angular momentum

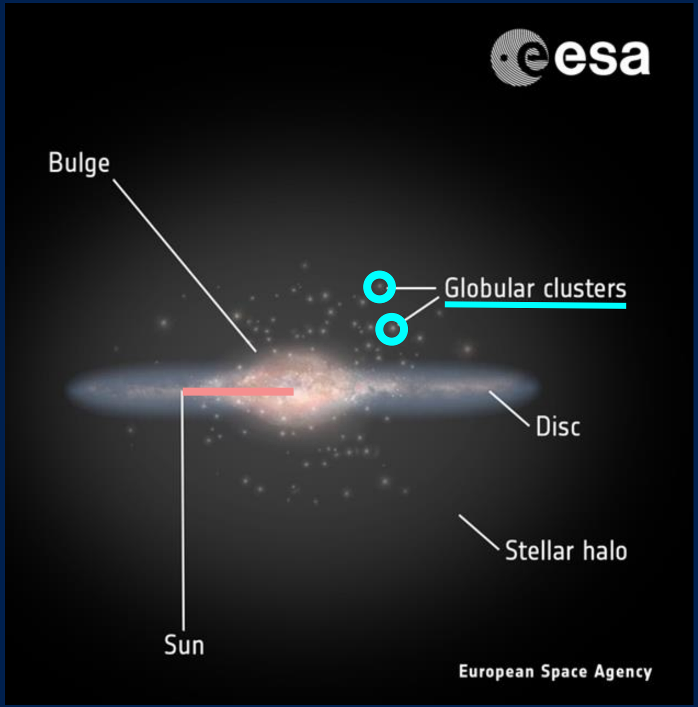
Cuddeford (1991)

$$f(\mathcal{E}, L) \propto L^{-2\beta} f(\mathcal{E})$$

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model



Specific energy:

$$\mathcal{E} = -\frac{1}{2}(v_r^2 + v_t^2) + \Phi(r)$$

Angular momentum:

$$L = r v_t$$

...

Probability density for specific energy and angular momentum

Cuddeford (1991)

$$f(\mathcal{E}, L) \propto L^{-2\beta} f(\mathcal{E})$$

Probability density for the specific energy:

Binney & Tremaine (2008), *Galactic Dynamics*, Princeton.

$$f(\mathcal{E}) = \frac{1}{\sqrt{8}\,\pi^2} \int_0^{\mathcal{E}} \frac{1}{\sqrt{\mathcal{E} - \Phi}} \left(\frac{d^2\rho_t}{d\Phi^2}\right) d\Phi + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{d\rho_t}{d\Phi}\right)_{\Phi=0}$$

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model

Probability density for specific energy and angular momentum

Cuddeford (1991)

$$f(\mathcal{E}, L) \propto L^{-2\beta} f(\mathcal{E})$$

Probability density for the specific energy:

Binney & Tremaine (2008), *Galactic Dynamics*, Princeton.

$$f(\mathcal{E}) = \frac{1}{\sqrt{8}\,\pi^2} \int_0^{\mathcal{E}} \frac{1}{\sqrt{\mathcal{E} - \Phi}} \left(\frac{d^2\rho_t}{d\Phi^2} \right) d\Phi + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{d\rho_t}{d\Phi} \right)_{\Phi=0}$$

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model

Gravitational Potential: $\Phi(r)$

Number density profile of globular clusters: $\rho(r)$

Assuming: spherical symmetry and equilibrium state

Probability density for specific energy and angular momentum

Cuddeford (1991)

$$f(\mathcal{E},L) \propto L^{-2\beta} f(\mathcal{E})$$

Probability density for the specific energy:

Binney & Tremaine (2008), *Galactic Dynamics*, Princeton.

$$f(\mathcal{E}) = \frac{1}{\sqrt{8}\,\pi^2} \int_0^{\mathcal{E}} \frac{1}{\sqrt{\mathcal{E}-\Phi}} \left(\frac{d^2\rho_t}{d\Phi^2}\right) d\Phi + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{d\rho_t}{d\Phi}\right)_{\Phi=0}$$

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model

Gravitational Potential:

$$\Phi(r) = \frac{\Phi_o}{r^\gamma}$$

Number density profile of globular clusters:

$$\rho(r) \propto \frac{1}{4\pi r^2} \times \frac{1}{r^{\alpha-2}}$$

Assuming: spherical symmetry and equilibrium state

Probability density for specific energy and angular momentum

Cuddeford (1991)

$$f(\mathcal{E}, L) \propto L^{-2\beta} f(\mathcal{E})$$

Probability density for the specific energy:

Binney & Tremaine (2008), *Galactic Dynamics*, Princeton.

$$f(\mathcal{E}) = \frac{1}{\sqrt{8}\,\pi^2} \int_0^{\mathcal{E}} \frac{1}{\sqrt{\mathcal{E}-\Phi}} \left(\frac{d^2\rho_t}{d\Phi^2}\right) d\Phi + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{d\rho_t}{d\Phi}\right)_{\Phi=0}$$

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model

Gravitational Potential:

$$\Phi(r) = \frac{\Phi_o}{r^\gamma}$$

Number density profile of globular clusters:

$$\rho(r) \propto \frac{1}{4\pi r^2} \times \frac{1}{r^{\alpha-2}}$$

Assuming: spherical symmetry and equilibrium state

Probability density for specific energy and angular momentum

Cuddeford (1991)

$$f(\mathcal{E}, L) \propto L^{-2\beta} f(\mathcal{E})$$

Probability density for the specific energy:

Binney & Tremaine (2008), *Galactic Dynamics*, Princeton.

$$f(\mathcal{E}) = \frac{1}{\sqrt{8}\,\pi^2} \int_0^{\mathcal{E}} \frac{1}{\sqrt{\mathcal{E} - \Phi}} \left(\frac{d^2\rho_t}{d\Phi^2}\right) d\Phi + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{d\rho_t}{d\Phi}\right)_{\Phi=0}$$

Prior

$p(h(\vartheta_i)|\boldsymbol{\theta})$

Physical Model

Gravitational Potential:

$$\Phi(r) = \frac{\Phi_o}{r^\gamma}$$

Number density profile of globular clusters:

$$\rho(r) \propto \frac{1}{4\pi r^2} \times \frac{1}{r^{\alpha-2}}$$

Assuming: spherical symmetry and equilibrium state

Probability density function:

$$f(\mathcal{E}, L) = \frac{L^{-2\beta}\, \mathcal{E}^{\frac{\beta(\gamma-2)}{\gamma}+\frac{\alpha}{\gamma}-\frac{3}{2}}\, \Gamma(\frac{\alpha}{\gamma} - \frac{2\beta}{\gamma} + 1)}{\sqrt{8\pi^3 2^{-2\beta}}\, \Phi_0^{\frac{-2\beta}{\gamma}+\frac{\alpha}{\gamma}}\, \Gamma(\frac{\beta(\gamma-2)}{\gamma} + \frac{\alpha}{\gamma} - \frac{1}{2})\, \Gamma(1-\beta)}$$

Evans et al 1997, Deason et al 2011, 2012 (*notations differ!*)

Prior

$p(h(\vartheta_i)|\theta)$

Physical Model

Gravitational Potential:

$$\Phi(r) = \frac{\Phi_o}{r^\gamma}$$

Number density profile of globular clusters:

$$\rho(r) \propto \frac{1}{4\pi r^2} \times \frac{1}{r^{\alpha-2}}$$

Assuming: spherical symmetry and equilibrium state

Probability density function:

$$f(\mathcal{E}, L) = \frac{L^{-2\beta}\mathcal{E}^{\frac{\beta(\gamma-2)}{\gamma}+\frac{\alpha}{\gamma}-\frac{3}{2}}\Gamma(\frac{\alpha}{\gamma}-\frac{2\beta}{\gamma}+1)}{\sqrt{8\pi^3 2^{-2\beta}}\Phi_0^{\frac{-2\beta}{\gamma}+\frac{\alpha}{\gamma}}\Gamma(\frac{\beta(\gamma-2)}{\gamma}+\frac{\alpha}{\gamma}-\frac{1}{2})\Gamma(1-\beta)}$$

Evans et al 1997, Deason et al 2011, 2012 (*notations differ!*)

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model

Gravitational Potential:

$$\Phi(r) = \frac{\Phi_o}{r^\gamma}$$

Number density profile of globular clusters:

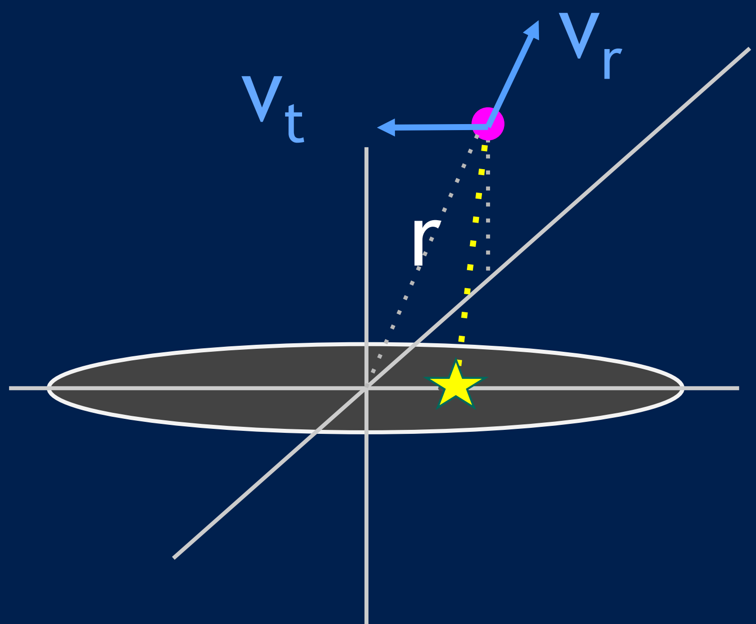
$$\rho(r) \propto \frac{1}{4\pi r^2} \times \frac{1}{r^{\alpha-2}}$$

Assuming: spherical symmetry and equilibrium state

Probability density function:

$$f(\mathcal{E}, L)$$

Galactocentric



Evans et al 1997, Deason et al 2011, 2012 (*notations differ!*)

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model

Gravitational Potential:

$$\Phi(r) = \frac{\Phi_o}{r^\gamma}$$

Number density profile of globular clusters:

$$\rho(r) \propto \frac{1}{4\pi r^2} \times \frac{1}{r^{\alpha-2}}$$

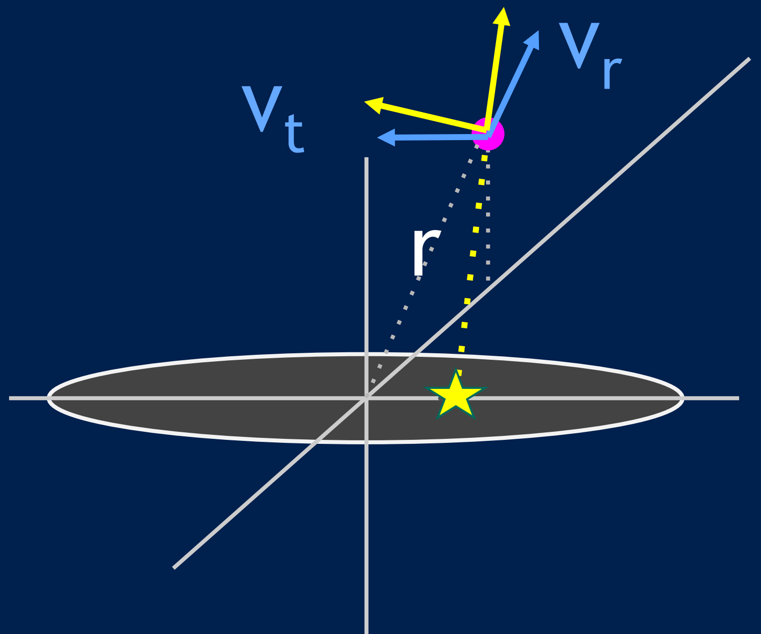
Assuming: spherical symmetry and equilibrium state

Probability density function:

$$f(\mathcal{E}, L)$$

Galactocentric

Heliocentric



Evans et al 1997, Deason et al 2011, 2012 (*notations differ!*)

Prior

$$p(h(\vartheta_i)|\boldsymbol{\theta})$$

Physical Model

Gravitational Potential:

$$\Phi(r) = \frac{\Phi_o}{r^\gamma}$$

Number density profile of globular clusters:

$$\rho(r) \propto \frac{1}{4\pi r^2} \times \frac{1}{r^{\alpha-2}}$$

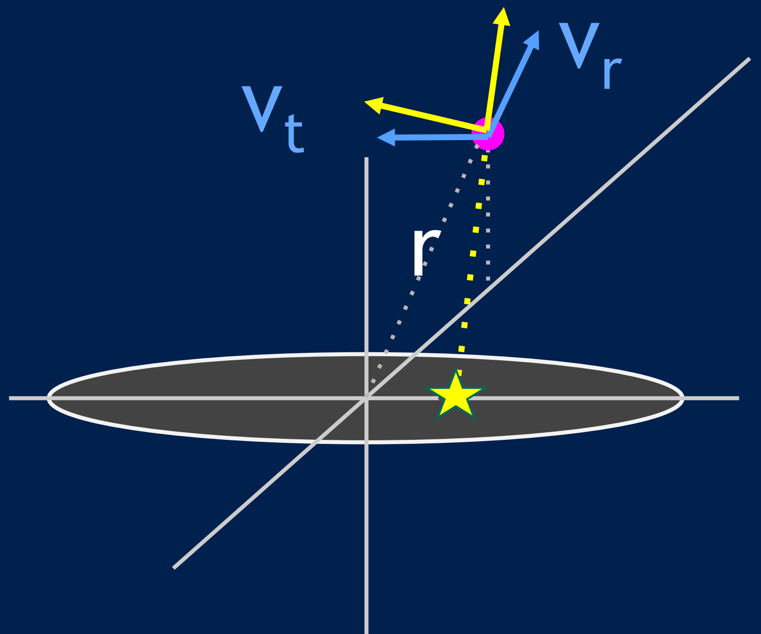
Assuming: spherical symmetry and equilibrium state

Probability density function:

$$f(\mathcal{E}, L)$$

Galactocentric

Heliocentric



Evans et al 1997, Deason et al 2011, 2012 (*notations differ!*)

Hierarchical Bayesian Model

Posterior Distribution $\propto$ Likelihood • Prior • Hyperprior

$$p(\boldsymbol{\theta}|\boldsymbol{y},\Delta) \propto \prod_i^N \mathcal{L}(\boldsymbol{y_i}|\vartheta_i,\Delta_i)\, p(h(\vartheta_i)|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

measurement model

physical model

Priors on model parameters

- Hyperprior

$p(\boldsymbol{\theta})$

Priors on model parameters

Hyperprior

$p(\boldsymbol{\theta})$

Priors on model parameters

Φ_o the scale factor for the gravitational potential

γ the power-law slope of the gravitational potential

α the power-law slope of the satellite population

β the velocity anisotropy parameter

Hyperprior

$p(\boldsymbol{\theta})$

Priors on model parameters

Hyperprior Probability Distributions

Parameter	Distribution	Hyperparameters
Φ_o	Uniform	$\Phi_{o,\mathrm{min}} = 1$, $\Phi_{o,\mathrm{max}} = 200$
γ	Uniform	$\gamma_{\mathrm{min}} = 0.3$, $\gamma_{\mathrm{max}} = 0.7$
α	Gamma	$b = 0.4$ kpc, $c = 0.001$, $p = 0.001$
β	Uniform	$\beta_{\mathrm{min}} = -0.5$, $\beta_{\mathrm{max}} = 1$

Φ_o the scale factor for the gravitational potential

γ the power-law slope of the gravitational potential

α the power-law slope of the satellite population

β the velocity anisotropy parameter

Hierarchical Bayesian Model

Posterior Distribution $\propto$ Likelihood • Prior • Hyperprior

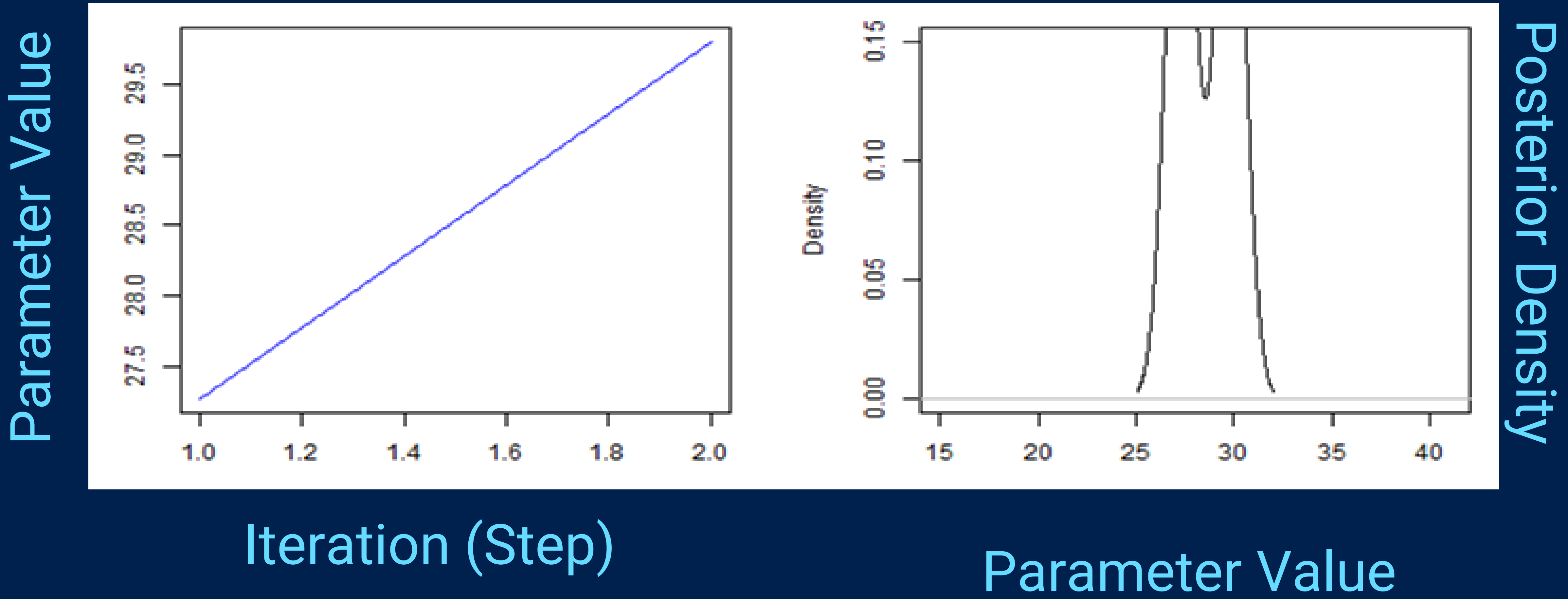
measurement model

physical model

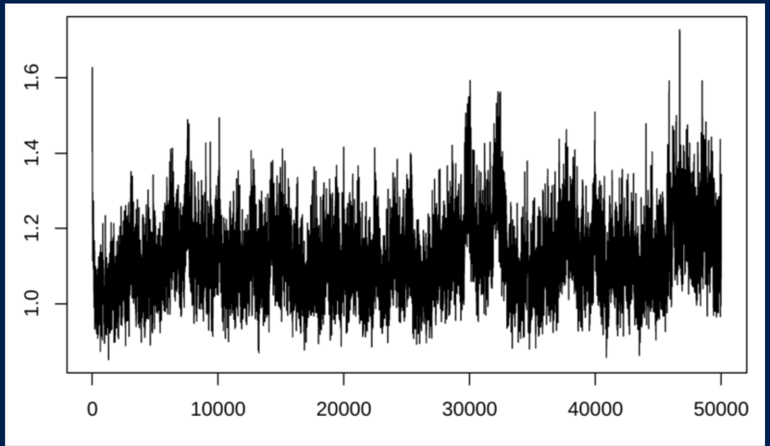
parameter assumptions

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \Delta) \propto \prod_{i}^{N} \mathcal{L}(\boldsymbol{y_i}|\vartheta_i, \Delta_i)p(h(\vartheta_i)|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

Sampling the posterior distribution → MCMC



Parameter Value



Iteration (Step)

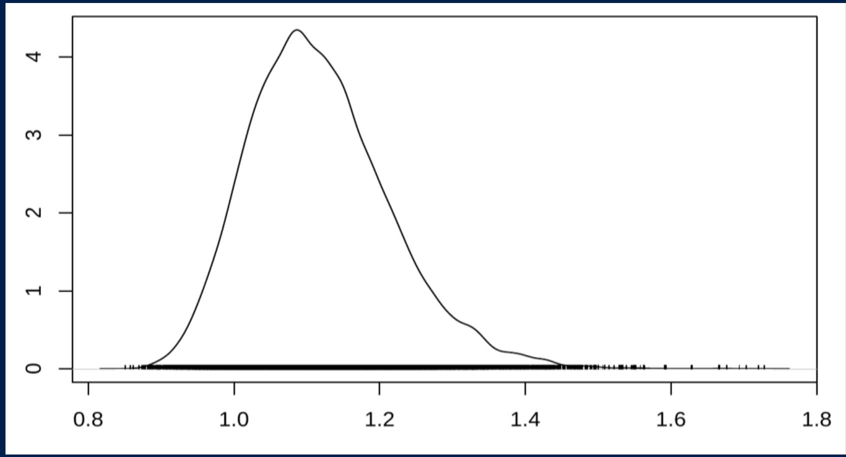
Trace Plot of Markov Chain

- Value of parameter as Markov chain walks through parameter space

Density Plot

- Smoothed histogram of parameter values in Markov chain

Density



Parameter Value

Sampling the posterior distribution

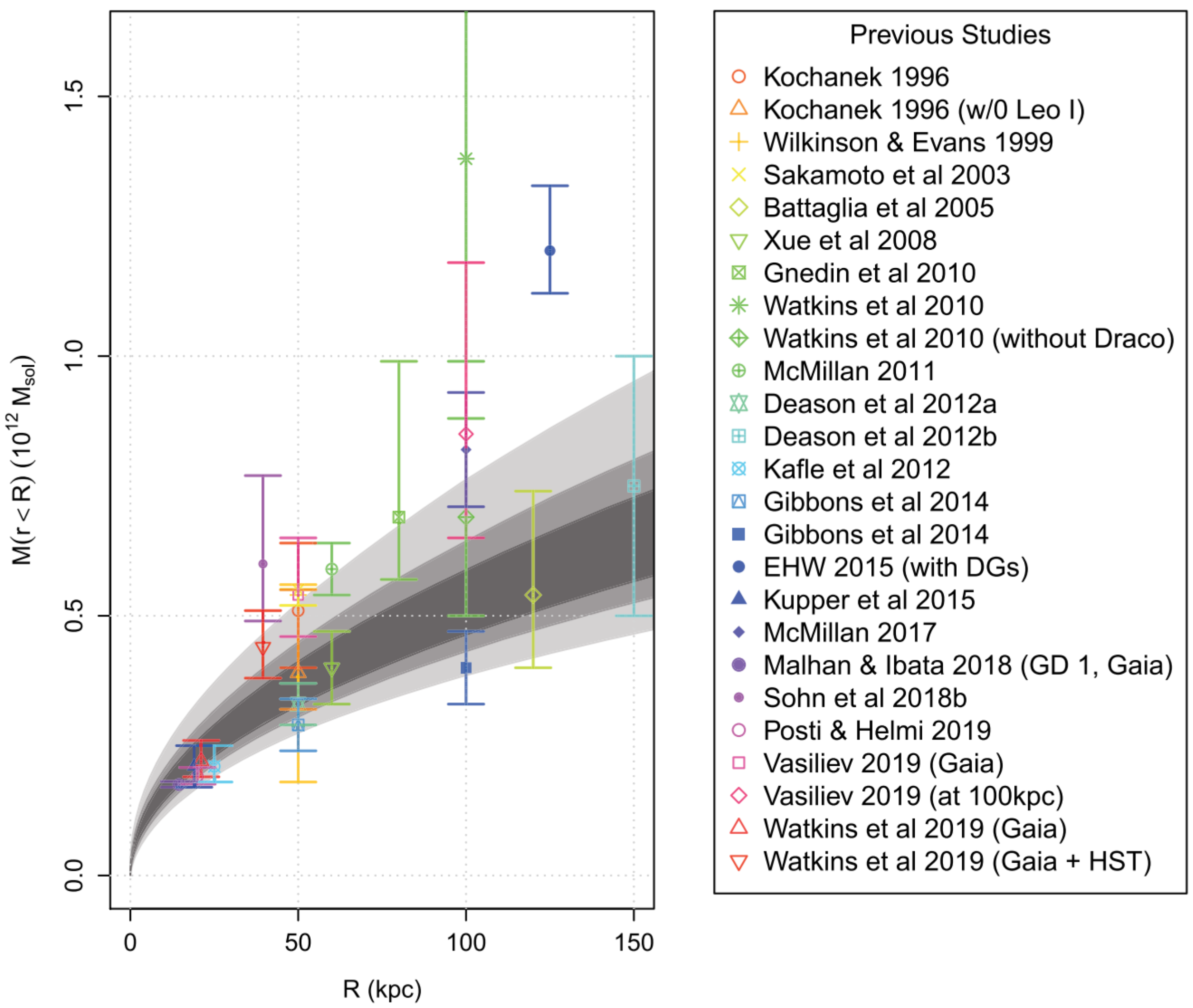
- Adaptive-tuning algorithm for efficient mixing
  - Roberts, G. O., & Rosenthal, J. S. 2009, J. Comput. Graphical Stat., 18, 349
- R-hat statistic to monitor convergence
  - Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003) *Bayesian Data Analysis* Chapman & Hall/CRC Press)
- Hybrid-Gibbs sampler that incorporates incomplete data
  - Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003) *Bayesian Data Analysis* Chapman & Hall/CRC Press)

Posterior Distribution is then used to calculate a

cumulative mass profile

with credible regions

THE ASTROPHYSICAL JOURNAL, 875:159 (9pp), 2019 April 20



Eadie & Juric (2019), ApJ 875:159