Quick Intro to Multinomial Distribution

May 7, 2025

Prof. Gwendolyn Eadie

Distribution of m&m's colour at a factory

- In exercise on Monday, we inferred the percentage of blue m&m's
  - Binomial distribution
    - *Blue* or *not blue*
- Now we want to infer percentage of each colour of m&m's
  - **Multinomial distribution with 6 categories**:
    - Red
    - Orange
    - Yellow
    - Green
    - Blue
    - Brown

Multinomial distribution

- When two or more unordered events, with probability of each type of event constant across trials, then for K types of events:

$$\Pr(y_1, \dots, y_K | n, \theta_1, \dots, \theta_K) = \frac{n!}{\prod_i y_i!} \prod_{i=1}^{K} \theta_i^{y_i}$$

- K is the number of colours ← k goes from 1 to $i = 6$
- n is the number of trials ← # of m&m's drawn from bag
- θ_i is the probability of success for category i ← percentage of colour i made
- y_i is the number of successes for category i ← number of colour i

Multinomial distribution

For k colours, the likelihood would be

$$p(y|\theta_1, \dots, \theta_k) \propto \left[\frac{n!}{\prod_i y_i!} \prod_{i=1}^{K} \theta_i^{y_i}\right] p(\theta_1, \dots, \theta_k)$$

And we could use the conjugate prior to the multinomial distribution, which is a Dirichlet distribution

$\boldsymbol{\theta} \sim Dirichlet(\alpha_1, \dots, \alpha_k)$ → note $\boldsymbol{\theta}$ is a vector of length $\boldsymbol{k}$

Intro to Hierarchical Modeling

May 7, 2025

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

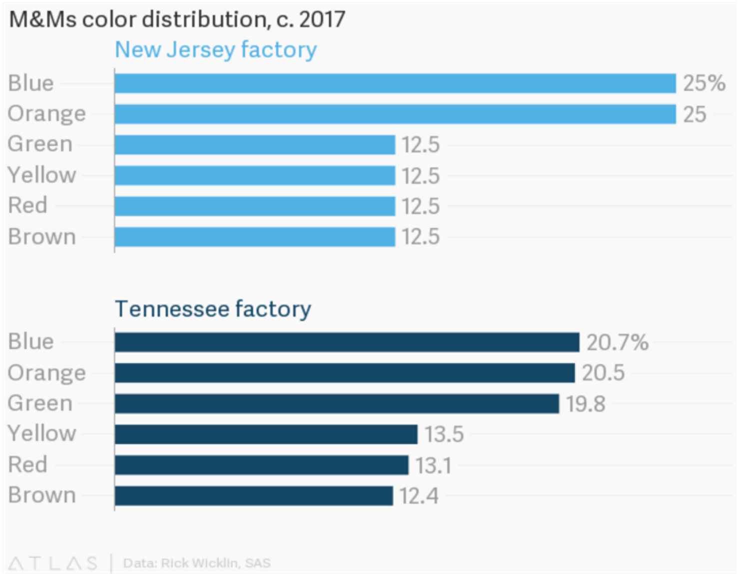
$$p(\theta, \alpha, \beta|y) \propto p(y|\theta)p(\theta|\alpha, \beta)\boldsymbol{p}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

- 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

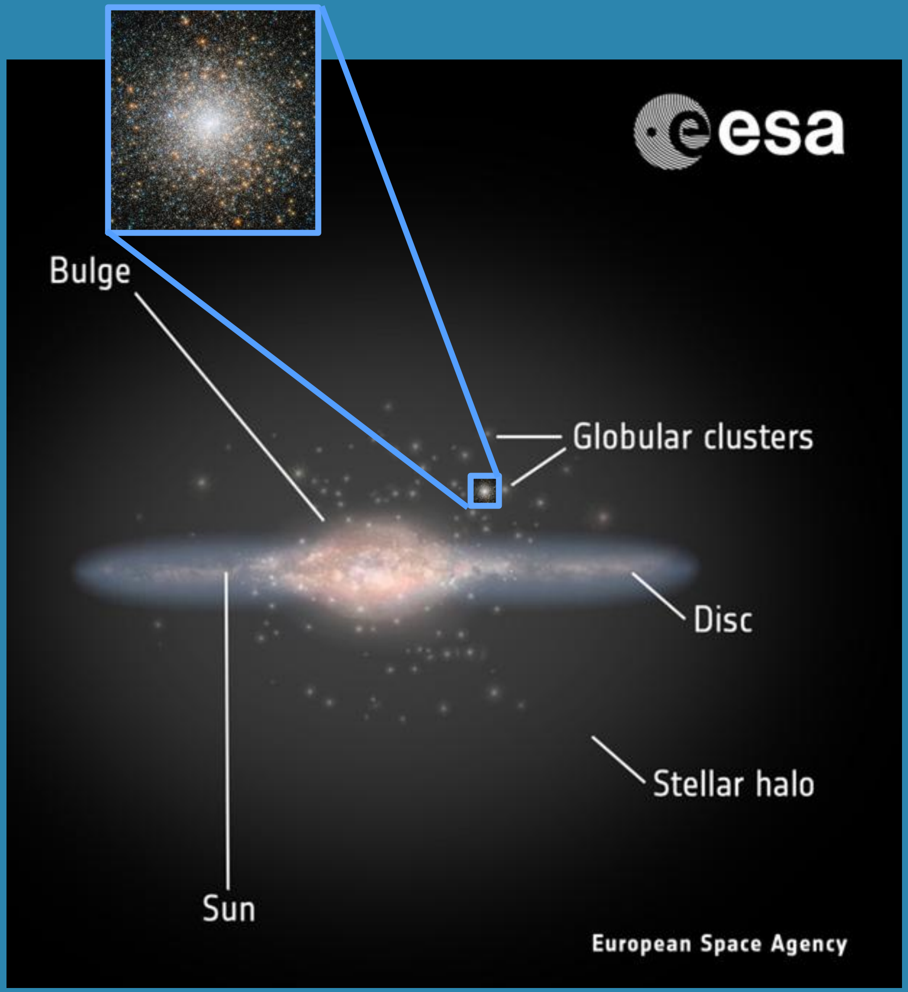
- Set a *hyperprior distribution* on α and β, and then we'd have

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

- This is now a hierarchical model.

Examples from Astronomy Research

Globular Cluster (GC)



Sketch of Milky Way

Estimating the mass of the Milky Way

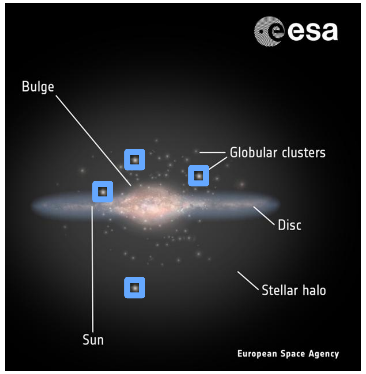
Using hierarchical Bayes and "kinematic tracers"

Hierarchical Bayesian Model for MW Mass Estimate in Pictures

Likelihood

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

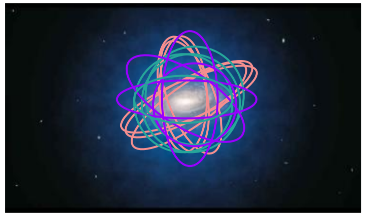
Prior



Shared population parameters for galaxy:

- Spatial density of GCs
- Gravitational potential
- Velocity anisotropy

Hyperprior



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

Hierarchical Bayesian Model for MW Mass Estimate in Math

Posterior Distribution ∝ Likelihood • Prior • Hyperprior

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

measurement model

physical model

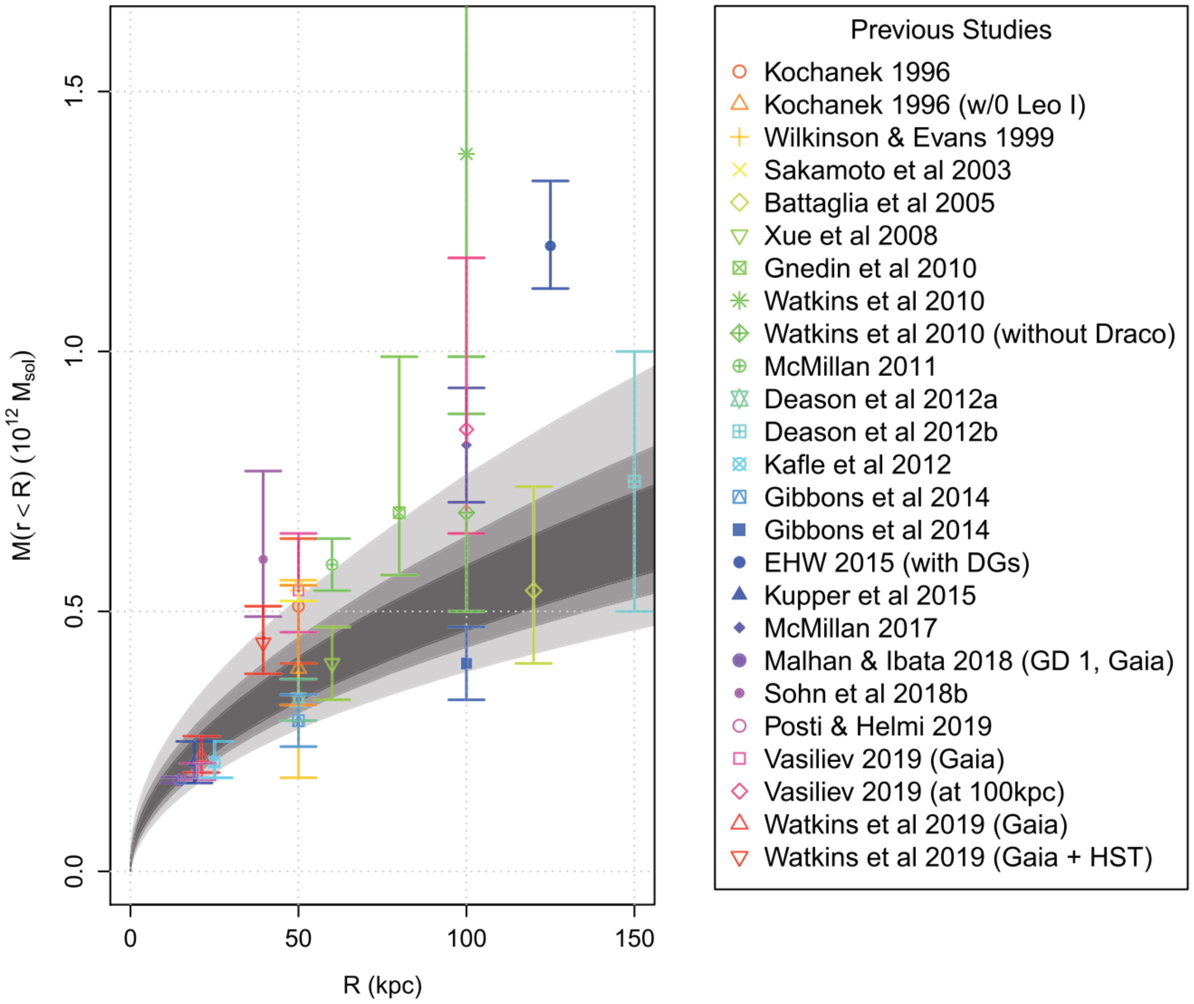
Priors on model parameters

Posterior Distribution is then used to calculate a

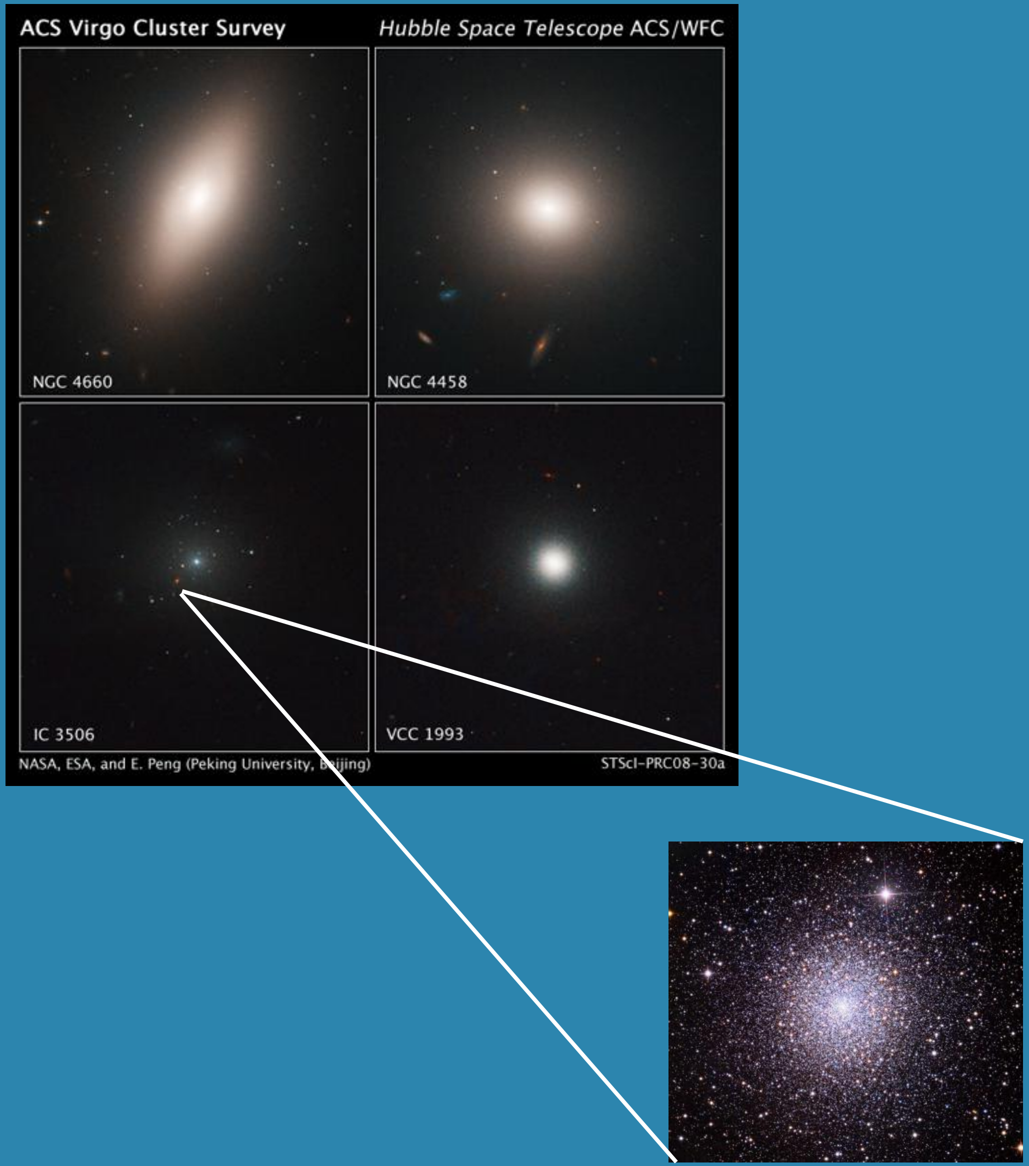
cumulative mass profile

with credible regions

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



Eadie & Juric (2019), ApJ 875:159



Inferring the relationship between GCs and their host galaxy mass

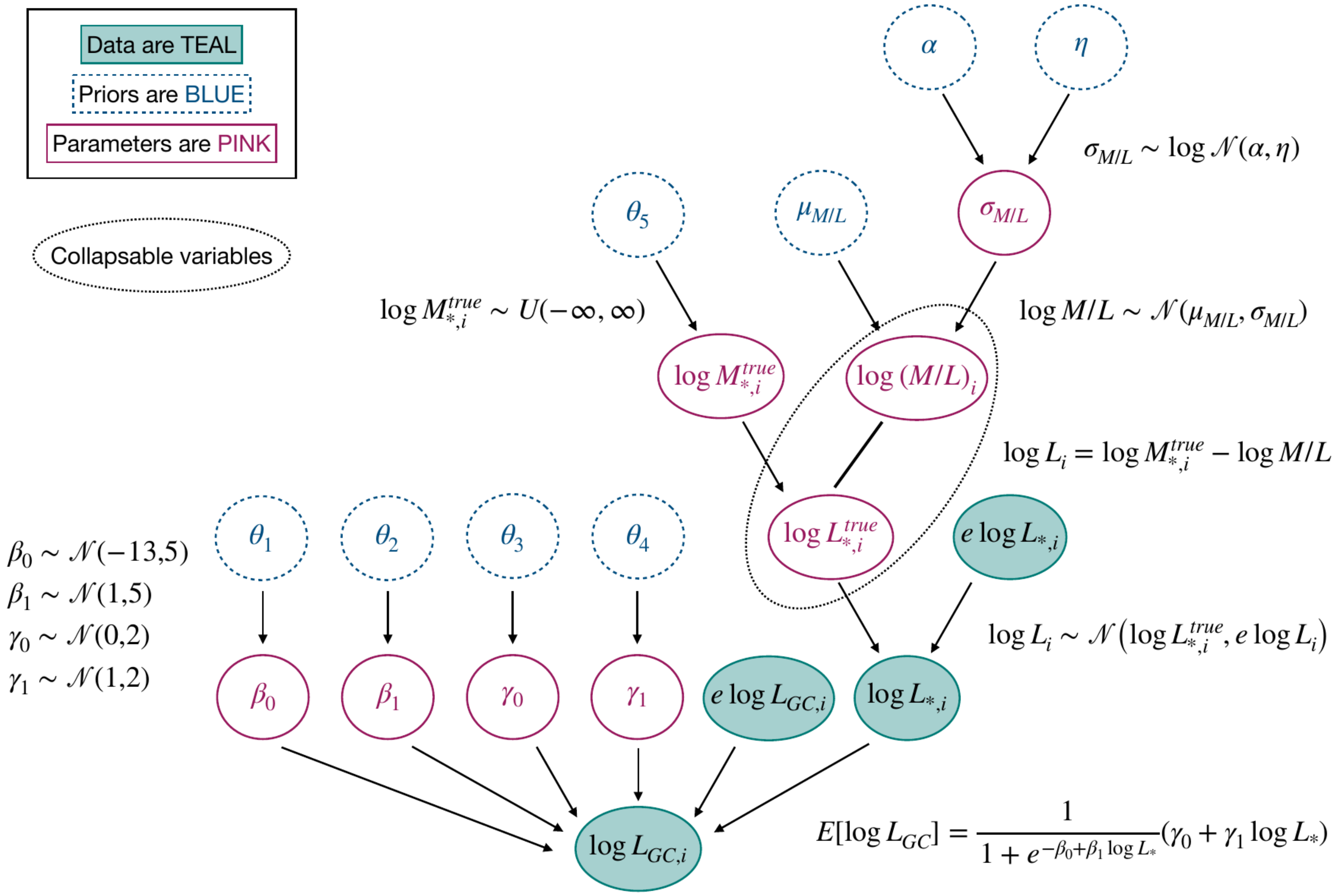
Using a hierarchical errors-in-variables model

What we observe in the local universe:

Large galaxies have GCs, but some smaller (dwarf) galaxies do not.

In between these extremes, there is a transition.
How large is the transition region?

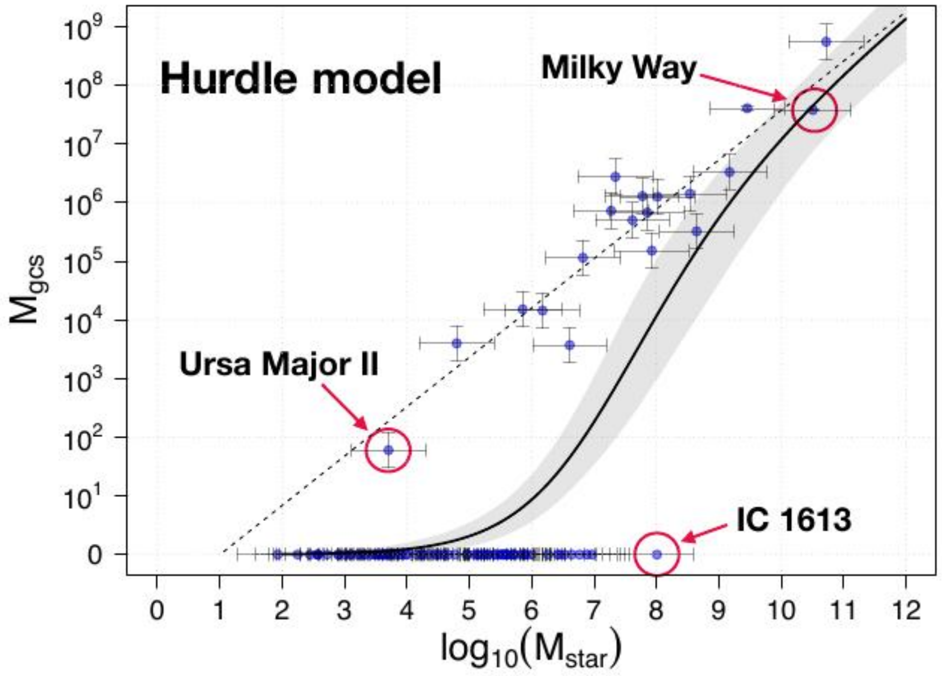
Hierarchical Errors-in-Variables Bayesian Log-Normal Hurdle Model

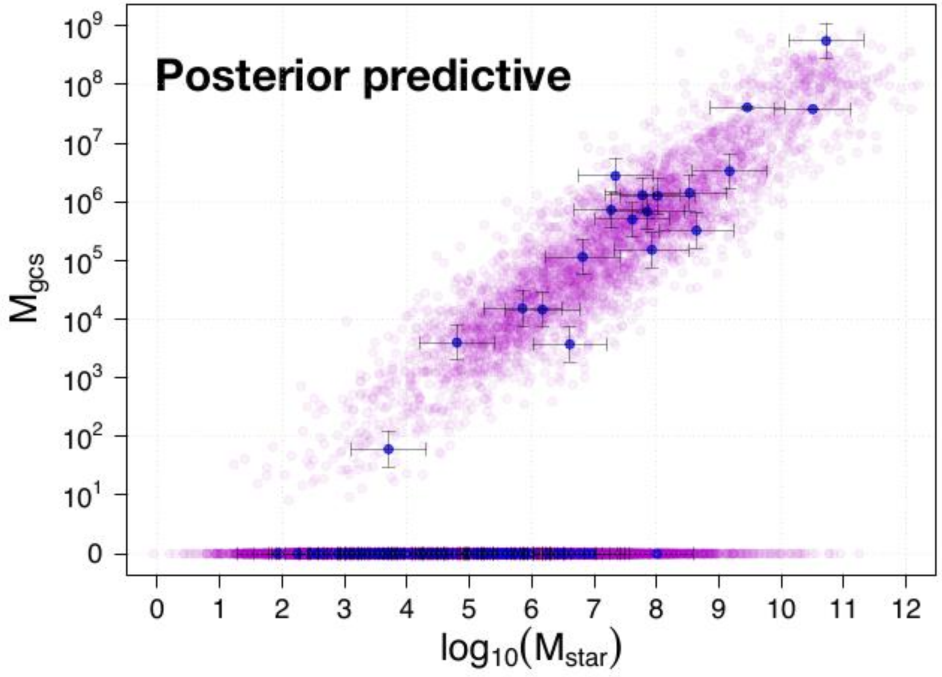


Sam Berek
PhD Candidate

Berek, **Eadie,** Speagle, & Harris (2023), ApJ 955(1), 22.

HERBAL model





Sam Berek
PhD Candidate

Berek, **Eadie,** Speagle, & Harris (2023), ApJ 955(1), 22.