A generative finite mixture model

hello

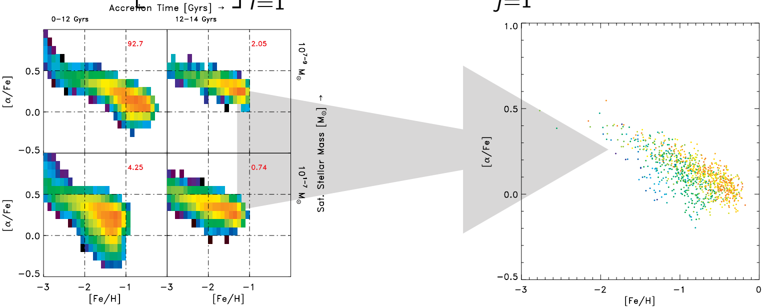
hello

- abc

Mixture components

$$\left[\frac{Fe}{H}, \frac{\alpha}{Fe}\right]_{i=1}^{N} \text{ i.i.d} \sim f(x, y) = \sum_{j=1}^{m} \pi_j f_j(x, y)$$

Observations



Rigid body dynamics

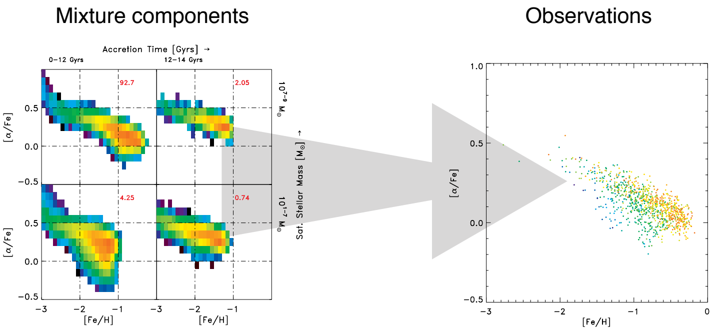
A modest attempt at using PGF/TikZ.
Green's Theorem:

- Curl
- Divergence

$$\oint_{\partial D} F \cdot ds = \int\int_{D} (\nabla \times F) \cdot k \, dA \tag{1}$$

- Boundary of Region
- Region

A generative finite mixture model



$$\left[\frac{Fe}{H}, \frac{\alpha}{Fe}\right]_{i=1}^{N} \text{i.i.d} \sim f(x,y) = \sum_{j=1}^{m} \pi_j f_j(x,y)$$

Where the mixing proportions, $\boldsymbol{\pi}$, give the formation history.

A generative finite mixture model

- each observed point comes from one of m mixture components (pictured)
- we propose a generative model in the form of a finite mixture model
- since each mixture component has an associated mass and accretion time range, the formation history is specified if we know what percentage of observations come from each mixture component
- our goal, then, is to determine the mixing proportions, $\boldsymbol{\pi}$

Model definition

$$\text{Let} \qquad x = \frac{\alpha}{Fe}, \qquad y = \frac{Fe}{H}$$

Given m mixture components, we propose that the density from which observations are generated is

$$f(x,y) = \sum_{j=1}^{m} \pi_j \, f_j(x,y) \tag{2}$$

- Mixing proportion
- Mixture component j

$$\text{where} \qquad \sum_{j=1}^{m} \pi_j = 1, \qquad \pi_j \geq 0, \qquad j = 1, \ldots, m$$

Definitions

- For notational simplicity, x$=\frac{\alpha}{Fe}$ and y$=\frac{Fe}{H}$
- Formally, given m mixture components, the density from which all observations are drawn is as shown
- The mixing proportions, $\boldsymbol{\pi}$ must be non-negative, and sum to 1

Estimating the mixing proportions $\boldsymbol{\pi}$

To estimate the mixing proportions, we can use a maximum likelihood approach

$$\hat{\boldsymbol{\pi}}_{\text{MLE}} = \underset{\boldsymbol{\pi}}{\operatorname{argmax}}\, L(\boldsymbol{\pi})$$

$$\text{where} \qquad \mathcal{L}(\boldsymbol{\pi}) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} \pi_j f_j(x_i, y_i)\right)$$

Unfortunately the standard MLE procedure for estimating $\boldsymbol{\pi}$ is intractable with this likelihood.

Expectation Maximization (EM) algorithm to the rescue!

Estimating the mixing proportions $\boldsymbol{\pi}$

- hi

Expectation Maximization

Suppose we knew which mixture component f_j each observation came from:

$$z_{ij} = \mathbf{1}(x_i, y_i \sim f_j) = \begin{cases} 1 & (x_i, y_i) \sim f_j \\ 0 & \text{otherwise} \end{cases}$$

The log likelihood can then be expressed as

$$\ell(\boldsymbol{\pi}) = \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij} \log \left\{ \pi_j f_j(x_i, y_i) \right\}$$

The addition of the latent variable $\mathbf{z}$ actually makes things easier because it is easily differentiable in $\boldsymbol{\pi}$.

Expectation Maximization

- Suppose we knew which mixture component f_j each observation came from
- Then we could construct a latent indicator variable, z_{ij}, which is 1 if point i comes from mixture component j, and 0 otherwise
- The log like then becomes
- Since we're supposing that we know z_{ij}, it's trivial to differentiate this log likelihood with respect to $\hat{\boldsymbol{\pi}}$
-

Estimating $\hat{\boldsymbol{\pi}}$ using expectation maximization

We don't know $\mathbf{z}$, so we replace $\mathbf{z}$ with the expected value of $\mathbf{z}$, conditioned on the data and the last known $\hat{\boldsymbol{\pi}}$:

$$\hat{\boldsymbol{\pi}}^{(t)} = \underset{\boldsymbol{\pi}}{\operatorname{argmax}}\, \mathbb{E}\left[\ell(\boldsymbol{\pi}) \middle| \mathbf{x}, \mathbf{y}, \hat{\boldsymbol{\pi}}^{(t-1)}\right]$$

Starting with some random initial value for $\hat{\boldsymbol{\pi}}^{(0)}$, we iteratively

- Find the expected value of $\ell(\boldsymbol{\pi})$ using the current expected values of the latent variable $\mathbf{z}$
- Set $\hat{\boldsymbol{\pi}}^{(t)}$ to the $\underset{\boldsymbol{\pi}}{\operatorname{argmax}}$ of this expectation, which is simple to compute

And repeat until $\ell(\boldsymbol{\pi})$ stabilizes to a range $< 10^{-4}$

Estimating the mixing proportions π

- We don't know $\mathbf{z}$, so we replace $\mathbf{z}$ with the expected value of $\mathbf{z}$, conditioned on the data and the last known $\hat{\pi}$:
-
- the true likelihood is increasing in each iteration

Find the expected value of $L(\boldsymbol{\pi})$ using the current expected value of the latent variable

The expected value of $\ell(\boldsymbol{\pi})$, with respect to the conditional distribution of $\mathbf{z}$, given observed data and $\boldsymbol{\pi}^{(t-1)}$ is

$$\mathbb{E}_{\boldsymbol{\pi}}\left[\ell(\boldsymbol{\pi})\middle|\mathbf{x},\mathbf{y}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}_{\boldsymbol{\pi}}\left[z_{ij}|x_i,y_i\right]\left\{\log f_j(x_i,y_i)+\log \pi_j\right\}$$

Since z_{ij} is an indicator, its expected value is simply the probability that data point i comes from model j

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\pi}}\left[z_{ij}|x_i,y_i\right] &= \Pr_{\boldsymbol{\pi}}(z_{ij}|x_i,y_i) \\ &= \frac{p(x_i,y_i|z_{ij}=1)p(z_{ij}=1)}{p(x_i,y_i)} \\ &= \frac{\pi_j f_j(x_i,y_i)}{\sum_{j=1}^{m}\pi_j f_j(x_i,y_i)}\end{aligned}$$

Find the expected value of $L(\pi)$ using the current expected value of the latent variable

- cond prob

Find the $\underset{\boldsymbol{\pi}}{\operatorname{argmax}}$ of this expectation

Now that we have the expected value of $\ell(\boldsymbol{\pi})$ with respect to the conditional distribution of $\mathbf{z}$, we need only evaluate

$$\hat{\boldsymbol{\pi}}^{(t)} = \underset{\boldsymbol{\pi}}{\operatorname{argmax}}\, \mathbb{E}\left[\ell(\boldsymbol{\pi}) \middle| \mathbf{x}, \mathbf{y}, \hat{\boldsymbol{\pi}}^{(t-1)}\right]$$

Which can be analytically specified, at each time t, as:

$$\hat{\pi}_k^{(t)} = \frac{\sum_{i=1}^{n} w_{ij}^{(t-1)}}{n}$$

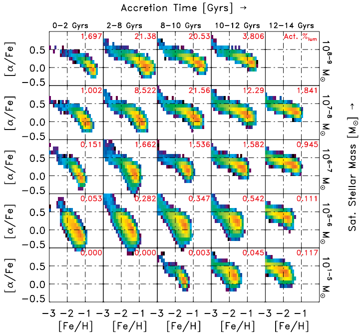
where $$w_{ij}^{(t+1)} = \frac{\pi_j^{(t)} f_j(x_i, y_i)}{\sum_{k=1}^{m} \pi_k^{(t)} f_k(x_i, y_i)}$$

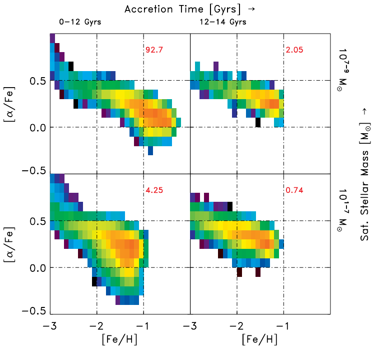
Find the $\arg\max_{\pi}$ of this expectation

- Note how simple this is to compute

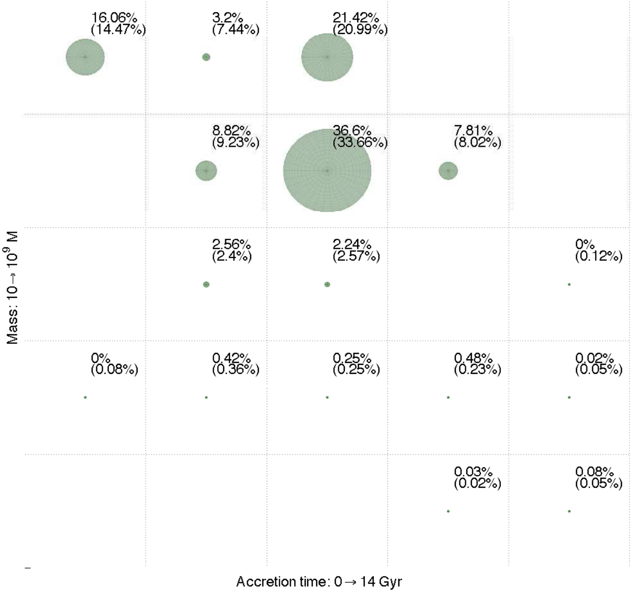
Simulations

- Generated observations from one halo and multiple halos
- Used a 5x5 grid ($m = 25$), and several 2x2 grids ($m = 4$)
  - 5x5 grid did not work for some halo realizations
  - 2x2 grid reliably converged on the correct mixing proportions

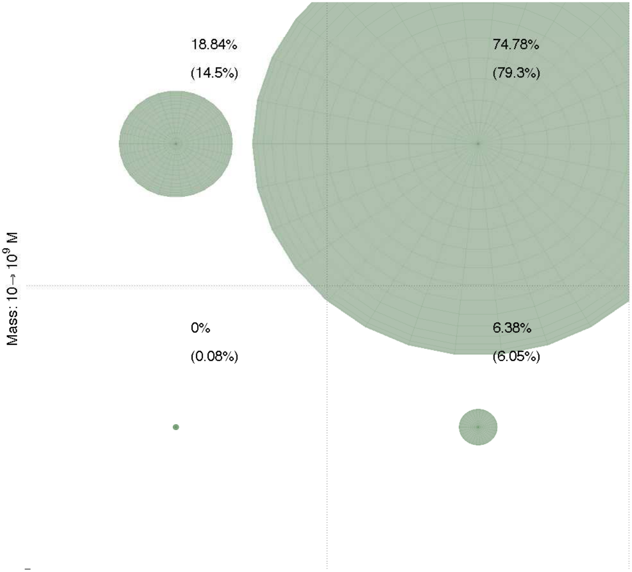




EM formation history for m=25

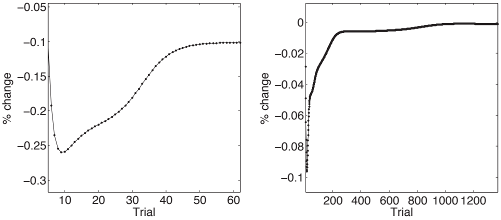


EM formation history for m=4

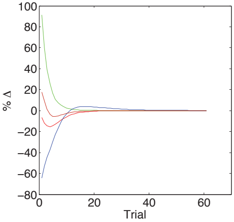


EM overview

Log likelihood % change

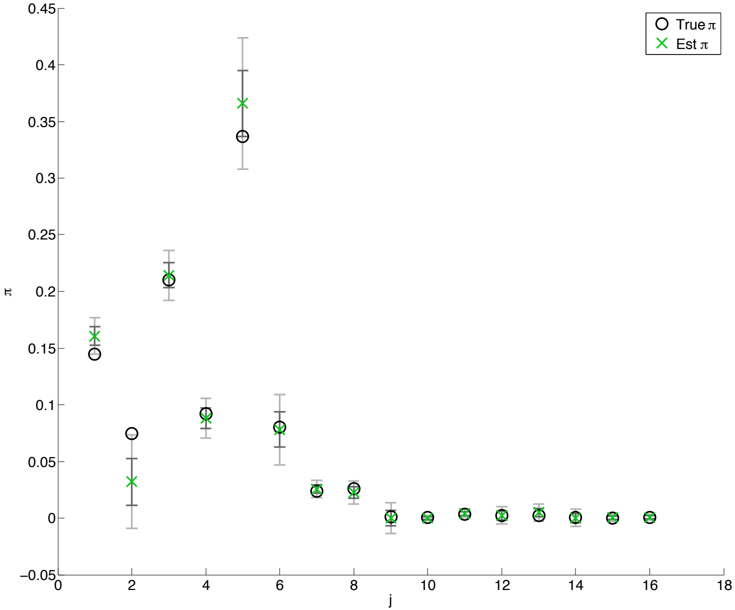


π % change

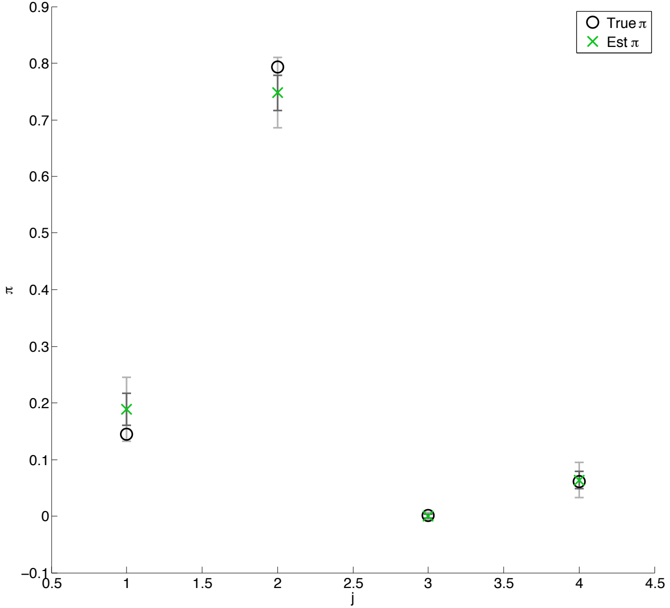


- Works with as few as 1,000 observations
- Insensitive to initialization of π
- Always converges
- Large weights identified after 10 iterations
- $\ell(\pi)$ stops changing appreciably after 60 (m=4) or 600 (m=25) iterations

Confidence intervals from observed Fisher Information

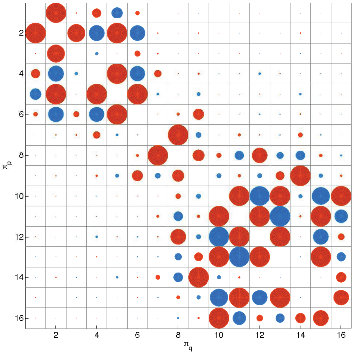


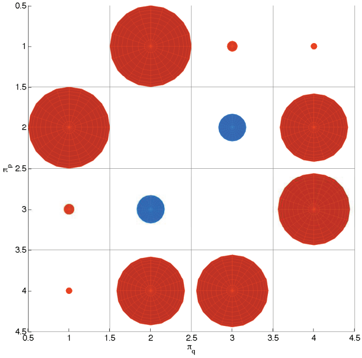
Confidence intervals from observed Fisher Information



Correlation between π

$$I(\pi'|\mathbf{x}, \mathbf{y}) = -\frac{\partial^2 \ell(\pi')}{\partial \pi' \partial \pi'^T}$$





Conclusion

Worked

- 2x2
- EM
- 5x5 in a few cases
- M-of-n bootstrapped errors

Did not work

- 5x5
- Parametric bootstrapped errors

Future improvements

- Non-arbitrary gridding
- Smoothing of f_j