Link between curves and M theoretical distributions

(Duane will cover this part I think)
Partition the mass and accretion time into M combinations of $\mathcal{M}, \mathcal{T}$ where

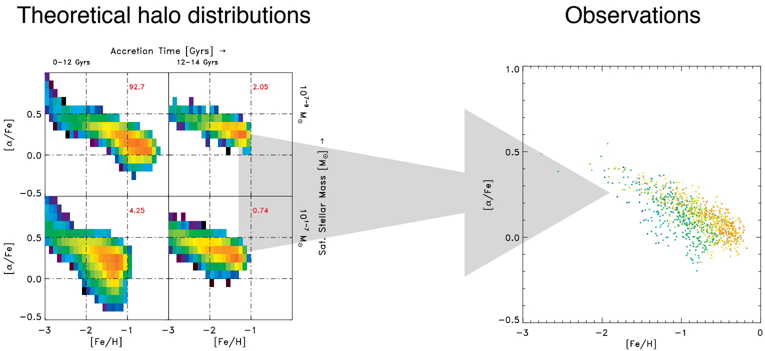
$$\text{Sat. stellar mass:} \quad \bigcup \mathcal{M}_j = [0, 10^9] M_{\odot}$$

$$\text{Accretion time:} \quad \bigcup \mathcal{T}_j = [0, 14]\text{Gyr}$$

$$f_j(x, y) = P(x, y | \text{Mass} \in \mathcal{M}_j, \text{Accretion time} \in \mathcal{T}_j)$$

Each observation is generated from one of these M theoretical distributions

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



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

Standard maximum likelihood estimates of $\boldsymbol{\pi}$ won't work

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

Suppose we knew which f_j each observation came from:

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

Then

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

Finding $\hat{\pi}$ using expectation maximization

- Find the expected value of the log likelihood, given the data
- Find the argmax_π of this expectation
- Repeat until $\log L(\pi)$ stabilizes

Find the expected value of the log likelihood, given the data

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

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

Find the argmax$_\pi$ of this expectation

$$\hat{\boldsymbol{\pi}}^{(t)} = \underset{\boldsymbol{\pi}}{\operatorname{argmax}} \, \mathrm{E}\Big[\log L(\boldsymbol{\pi}) \Big| \mathbf{x}, \mathbf{y}, \hat{\boldsymbol{\pi}}^{(t-1)} \Big]$$

Accounting for the $m-1$ free parameters of $\boldsymbol{\pi}$, differentiation proceeds, for $k = 1, \ldots, m-1$, as:

$$\frac{\partial}{\partial \pi_k} \mathrm{E}\Big[\log L(\boldsymbol{\pi}) \Big| \mathbf{x}, \mathbf{y} \Big] = \sum_{i=1}^{n} \left\{ w_{ik}^{(t-1)} \frac{1}{\pi_k} - w_{im}^{(t-1)} \frac{1}{1 - \pi_1 - \ldots - \pi_{m-1}} \right\}$$

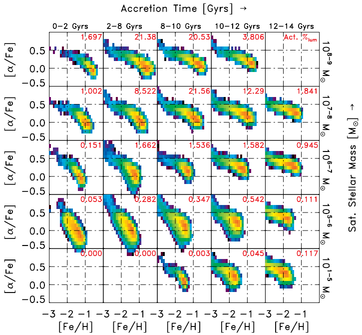
$$\frac{1}{\pi_k} \sum_{i=1}^{n} w_{ik}^{(t-1)} = \frac{1}{1 - \pi_1 - \ldots - \pi_{m-1}} \sum_{i=1}^{n} w_{im}^{(t-1)}$$

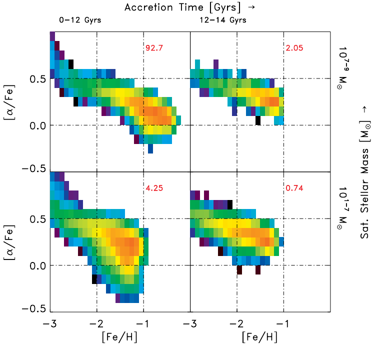
Consequently

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

$$\hat{\pi}_m^{(t)} = 1 - \pi_1 - \ldots - \pi_{m-1}$$

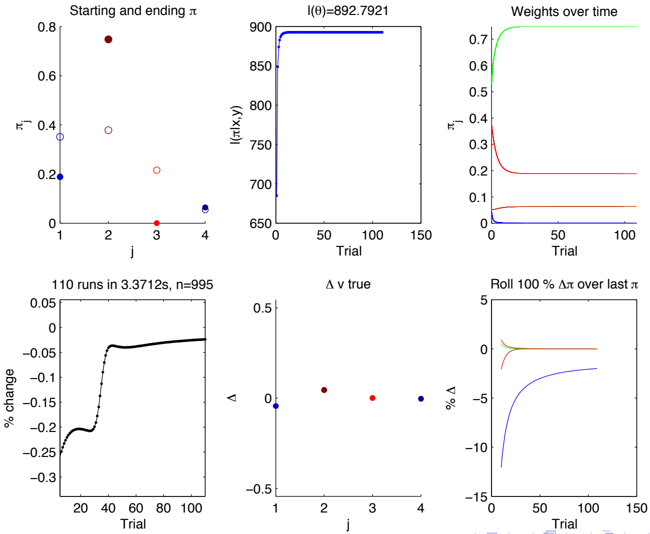
We used a 5x5 and a 2x2 set of theoretical distributions



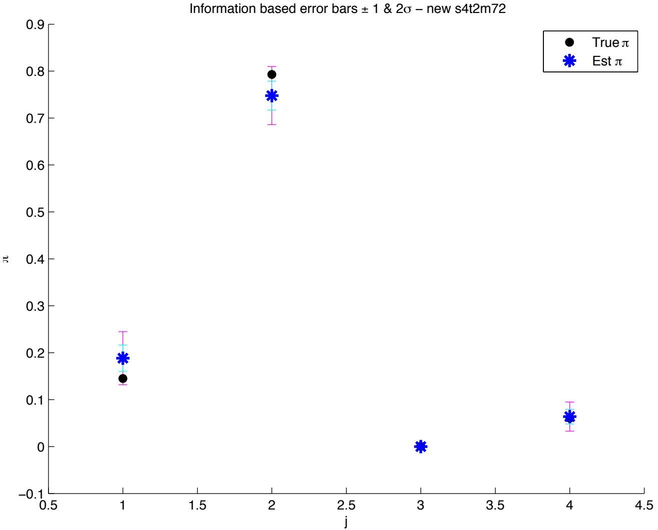


Simulation results

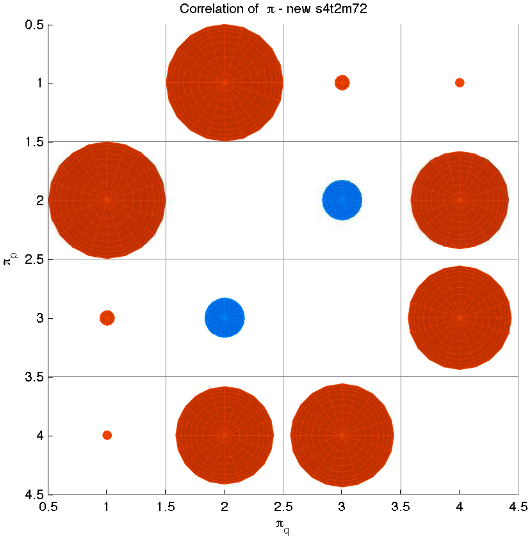
Works starting at about 1,000 observations, although larger π values are found with smaller data sets.



Confidence intervals



Correlation between π



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