METALLICITY

OCTOBER 25, 2010

1. Mixture model

For each observation of $(L, \frac{\alpha}{\text{Fe}}, \frac{\text{Fe}}{\text{H}})$, let $\big\{(x_i, y_i)\big\}_{i=1}^{n}$ represent observed metallicities of stars drawn from one of m known model densities. We model the density of observations using the mixture model

$$f(x,y) = \sum_{j=1}^{m} \pi_j f_j(x,y) \qquad \sum_{j=1}^{m} \pi_j = 1 \qquad \pi_j \geq 0, j = 1, \dots, m$$

With a complete likelihood of

$$L(\boldsymbol{\pi}) = \prod_{i=1}^{n} f(x_i, y_i) = \prod_{i=1}^{n} \Big\{ \sum_{j=1}^{m} \pi_j f_j(x_i, y_i) \Big\}$$
$$l(\boldsymbol{\pi}) = \sum_{i=1}^{n} \log \Big(\sum_{j=1}^{m} \pi_j f_j(x_i, y_i) \Big)$$

Evaluation of $\partial l(\boldsymbol{\pi})/\partial \pi$ can be avoided by adding a latent indicator, z, to the observed data (x, y), representing the model group from which that observation was generated. Let G_j be the j^{th} model group, and let

$$z_{ij} = \mathbf{1}\big\{(x_i, y_i) \mapsto G_j\big\}$$

The incomplete data likelihood is defined over the complete data $\big\{(x_i, y_i, \mathbf{z}_i)\big\}_{i=1}^{n}$ as

$$L(\boldsymbol{\pi}) = \prod_{i=1}^{n} \prod_{j=1}^{m} \Big\{ f_j(x_i, y_i) \Big\}^{z_{ij}} \pi_j^{z_{ij}}$$
$$l(\boldsymbol{\pi}) = \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij} \log \big\{ \pi_j f_j(x_i, y_i) \big\}$$

2. EM

First we find the expected value of $l(\boldsymbol{\pi})$ conditional on the distribution of z. Since z_{ij} is an indicator function, its expected value is equal to the probability that data point i comes from model j.

2.1. Expected value of $l(\boldsymbol{\pi})$.

The expected value of $l(\boldsymbol{\pi})$ is

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\pi}}\Big[l(\boldsymbol{\pi})\big|\mathbf{x},\mathbf{y}\Big] &= \sum_{i=1}^{n} E(\boldsymbol{\pi}|x_i,y_i)l(\boldsymbol{\pi}|x_i,y_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} E_{\boldsymbol{\pi}}\big[z_{ij}|x_i,y_i\big]\big\{\log f_j(x_i,y_i) + \log \pi_j\big\}
\end{aligned}$$

where

$$\mathrm{E}_{\boldsymbol{\pi}}\Big[z_{ij}|x_i,y_i\Big] = \mathrm{Probability}\Big((x_i,y_i)\mapsto G_j\big|x_i,y_i\Big)$$

2.2. Expected value of $\pi|\mathbf{x},\mathbf{y}$.

The expected value of z_{ij} is the same as the expected value of π_j, given the data. This can be specified as

$$\mathrm{E}_{\boldsymbol{\pi}}\Big[z_j|x_i,y_i\Big] = P(z_j|x_i,y_i) = \frac{P(x_i,y_i|z_j=1)P(z_j=1)}{P(x_i,y_i)}$$

with constituent parts:

$$P(x_i,y_i|z_j=1) = f_j(x_i,y_i) \qquad P(x_i,y_i|z_j) = \prod_{j=1}^{m} f_j^{z_i} \qquad P(z_j) = \prod_{j=1}^{m} \pi_j^{z_j}$$

thus

$$\mathrm{E}_{\boldsymbol{\pi}}\Big[z_j|x_i,y_i\Big] = \frac{\pi_j f_j(x_i,y_i)}{\sum_{j=1}^{m}\pi_j f_j(x_i,y_i)}$$

Defining $w_{ij}^{(t)}$ as the expected value of z_{ij} at the t^{th} step, and $\pi_j^{(t)}$ as the MLE of π_j at the t^{th} step, yeilds

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

2.3. **Solving for $\boldsymbol{\pi}$.**

$$\begin{aligned}0 &= \frac{\partial}{\partial \pi_k} \mathrm{E}_{\boldsymbol{\pi}}\Big[l(\boldsymbol{\pi})\Big|\mathbf{x},\mathbf{y}\Big] \\ &= \sum_{i=1}^{n}\Big\{w_{ij}^{(0)}\frac{1}{\pi_k} - w_{im}^{(0)}\frac{1}{1-\pi_1-\ldots-\pi_{m-1}}\Big\}, k=1,\ldots,m-1\end{aligned}$$

Therefore

$$\begin{aligned}\frac{1}{\pi_1}\sum_{i=1}^{n} w_{i1}^{(0)} = \ldots &= \frac{1}{\pi_{m-1}}\sum_{i=1}^{n} w_{i,m-1}^{(0)} = c \\ \hat{\pi}_k &= \frac{\sum_{i=1}^{n} w_{ik}^{(0)}}{c} \\ \pi_j^{(1)} &= \frac{\sum_{i=1}^{n} w_{ij}^{(0)}}{n}\end{aligned}$$

And in general,

$$\pi_j^{(t+1)} = \frac{\sum_{i=1}^{n} w_{ij}^{(t)}}{n}$$

3. Observed Information

As $\pi_m = 1 - \sum_{j=1}^{m} \pi_j$ there are only g free parameters. Thus let $g = m-1$, and $\boldsymbol{\pi}' = \boldsymbol{\pi}[1,\ldots,g]$. The observed information matrix, $I(\boldsymbol{\pi}'|\mathbf{x},\mathbf{y})$ is given by the $g\times g$ negative hessian:

$$l(\boldsymbol{\pi'}) = \sum_{i=1}^{n} \log\Big(\sum_{j=1}^{g} \pi_j f_j\Big)$$

$$\frac{\partial l(\boldsymbol{\pi'})}{\partial \pi_k} = \sum_{i=1}^{n} \frac{f_k}{\sum_{j=1}^{g} \pi_j f_j}$$

$$\frac{\partial^2 l(\boldsymbol{\pi'})}{\partial \pi_k \partial \pi_r} = -\sum_{i=1}^{n} \frac{f_k}{(\sum_{j=1}^{g} \pi_j f_j)^2 f_r}$$

$$I(\boldsymbol{\pi}|\mathbf{x},\mathbf{y}) = -\frac{\partial^2 l(\boldsymbol{\pi'})}{\partial \boldsymbol{\pi}' \partial \boldsymbol{\pi}'^T} = -\begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\pi'})}{\partial \pi_1^2} & \frac{\partial^2 l(\boldsymbol{\pi'})}{\partial \pi_1 \partial \pi_2} & \cdots & \frac{\partial^2 l(\boldsymbol{\pi'})}{\partial \pi_1 \partial \pi_g} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 l(\boldsymbol{\pi'})}{\partial \pi_g \partial \pi_1} & \frac{\partial^2 l(\boldsymbol{\pi'})}{\partial \pi_g \partial \pi_2} & \cdots & \frac{\partial^2 l(\boldsymbol{\pi'})}{\partial \pi_g^2} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \frac{1}{(\sum_{j=1}^{g} \pi_j f_j)^2} & \sum_{i=1}^{n} \frac{f_1}{(\sum_{j=1}^{g} \pi_j f_j)^2 f_2} & \cdots & \sum_{i=1}^{n} \frac{f_1}{(\sum_{j=1}^{g} \pi_j f_j)^2 f_g} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} \frac{f_g}{(\sum_{j=1}^{g} \pi_j f_j)^2 f_1} & \sum_{i=1}^{n} \frac{f_g}{(\sum_{j=1}^{g} \pi_j f_j)^2 f_2} & \cdots & \sum_{i=1}^{n} \frac{1}{(\sum_{j=1}^{g} \pi_j f_j)^2} \end{bmatrix}$$

*Excluding i where $(\sum_{j=1}^{g} \pi_j f_j)^2 f_i = 0$ to avoid dividing by zero.

3.1. Variance and Covariance of $\boldsymbol{\pi}$.

The asymptotic covariance matrix of $\boldsymbol{\pi}$ can be approximated by the inverse of the observed information matrix, $I^{-1}(\hat{\boldsymbol{\pi}}|\mathbf{x},\mathbf{y})$, yielding

$$\mathrm{Cov}(\pi_p,\pi_q) = \begin{cases} I^{-1}(\hat{\boldsymbol{\pi}}) & p,q<m \\ \sum_{j=1}^{g} \mathrm{Cov}(\pi_j,\pi_q) & p=m, q<m \\ \sum_{j=1}^{g}\sum_{k=1}^{g} \mathrm{Cov}(\pi_j,\pi_k) & p,q=m \end{cases}$$

$$\mathrm{Var}(\pi_j) = \sigma_j^2 = \Big\{\big|\mathrm{Cov}(\boldsymbol{\pi})\big|\Big\}_{jj}$$

* There are many negative covariances; this is ok I think.

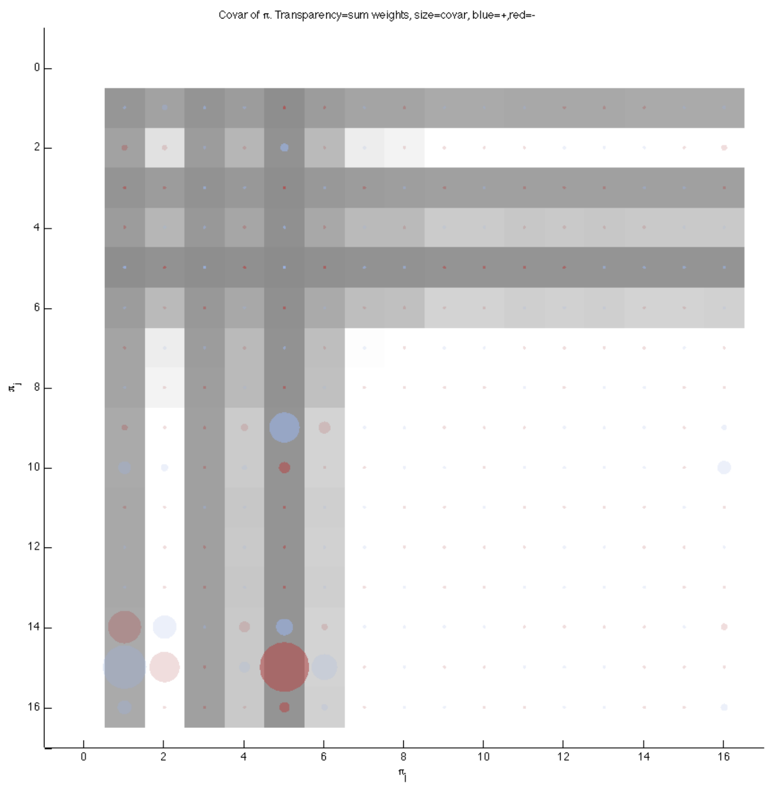
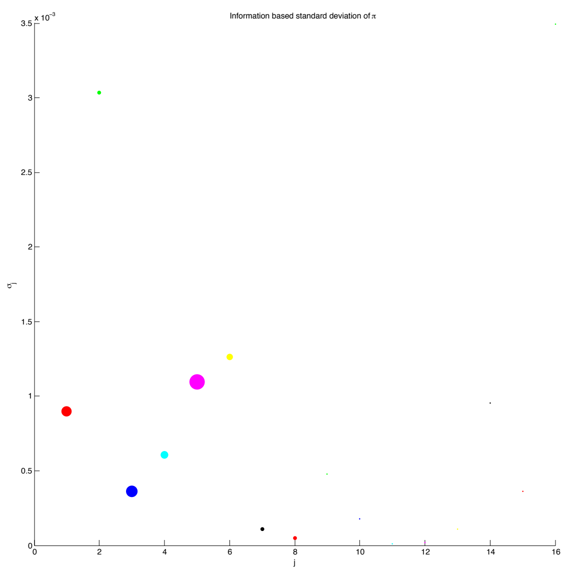


TABLE 1. Covariance, in percent, of $\hat{\boldsymbol{\pi}}$ overlaid on $\hat{\pi}_i + \hat{\pi}_j$. Larger dots are covariances; negative covariances are red.

0.0001	0.001	-	-	-0.0004	-0.0001	-	-	0.0001	-	-	-	-	-	-	0.0007
-0.001	-0.0009	-	-0.0002	0.0015	-0.0004	-	-	-	-	-	-	-	-	-	-0.001
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-0.0002	0.0003	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-0.0001	-	-	0.0001	-	-	-	-	-	-	-	-	-	-	-
0.0003	-0.0001	-	0.0001	-0.0004	0.0002	-	-	-	-	-	-	-	-	-	-
-0.0001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-0.0011	-0.0001	-	-0.0013	0.0053	-0.0021	-	-	-	-	-	-	-	0.0001	-	0.001
0.0023	0.0013	-0.0001	0.0008	-0.002	-	-0.0001	0.0001	-	-	-	-	-	-	-	0.0023
-	-	-	-	-0.0001	0.0001	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.0003	-0.0003	-	-	-0.0001	0.0001	-	-	-	-	-	-	-	-	-	-
-0.0058	0.0041	0.0001	-0.0019	0.003	-0.0011	0.0002	0.0001	0.0003	-	-	-0.0001	-0.0001	-0.0001	-	-0.0012
0.0076	-0.0053	-0.0001	0.0019	-0.0086	0.0045	-0.0003	0.0002	-0.0003	-	-	-	-	-0.0001	-	-0.0005
0.0025	-0.0001	-0.0001	-0.0005	-0.0018	0.0011	-0.0001	0.0003	-	-	-	-	-	-	-	0.0012



(A) Larger dots represent more weight accounted for by that $\hat{\pi}_j$.

	$\hat{\pi}_j$	π_j	$\sigma_{\hat{\pi}_j}$	Z-score
1	16.04	14.467	0.0898	17.509
2	3.320	7.435	0.3033	-13.564
3	21.392	20.991	0.0361	11.067
4	8.846	9.234	0.0608	-6.369
5	36.443	33.656	0.1094	25.479
6	7.882	8.019	0.1264	-1.080
7	2.560	2.400	0.0108	14.793
8	2.239	2.573	0.0047	-70.08
9	0	0.118	0.0480	-2.474
10	0	0.076	0.0178	-4.317
11	0.421	0.358	0.0009	65.245
12	0.246	0.253	0.0021	-2.959
13	0.485	0.225	0.0110	23.56
14	0.008	0.048	0.0954	-0.420
15	0.030	0.024	0.0362	0.183
16	0.083	0.046	0.3494	0.103

(B)

FIGURE 1. Variance of $\hat{\boldsymbol{\pi}}$, true $\boldsymbol{\pi}$, in percent, and z-scores.

$$\text{Z-score} = \frac{\hat{\pi}_j - \pi}{\sigma_{\hat{\pi}_j}}$$

4. Likelihood ratio test

Given certain conditions

$$H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_{\text{true}}$$
$$H_1 : \boldsymbol{\pi} \neq \boldsymbol{\pi}_{\text{true}}$$

$$\Lambda = -2 \log \frac{\sup_{\boldsymbol{\pi}=\boldsymbol{\pi}_{\text{true}}} l(\boldsymbol{\pi})}{\sup_{\boldsymbol{\pi}} l(\boldsymbol{\pi})} = -2\big\{ l(\boldsymbol{\pi}_{\text{true}}) - l(\hat{\boldsymbol{\pi}})\big\} \sim \chi^2_{m-1}$$

$$\Lambda_{\text{Halo 3}} = 25.025 \sim \chi^2_{15}$$
$$\text{p-value 10k} = 4.961\%$$
$$\text{p-value 30k} = 1.3 \times 10^{-5}\%$$
$$\text{p-value 50k} = 1.4 \times 10^{-12}\%$$

Thus we accept H_0 when requiring 95% or less confidence; there is only a 4.961% chance we would see a value this extreme or more given H_0 is true. This holds for 400 to 1600 EM iterations.