Ages of stellar populations from color-magnitude diagrams

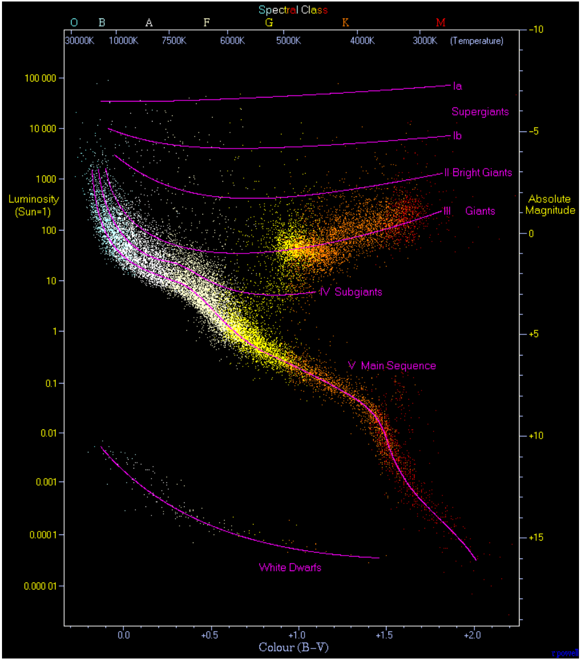
Paul Baines

Department of Statistics
Harvard University

September 30, 2008

Welcome!

Today we will look at using hierarchical Bayesian modeling to make inference about the properties of stars; most notably the age and mass of groups of stars. Complete with a brief dummies (statisticians) guide to the Astronomy behind it.


Spectral Class
O B A F G K M
30000K 10000K 7500K 6000K 5000K 4000K 3000K (Temperature)
Luminosity (Sun=1)
100 000
10 000
1000
100
10
1
0.1
0.01
0.001
0.0001
0.000 01
Absolute Magnitude
−10
−5
0
+5
+10
+15
Ia
Supergiants
Ib
II Bright Giants
III Giants
IV Subgiants
V Main Sequence
White Dwarfs
0.0
+0.5
+1.0
+1.5
+2.0
Colour (B−V)
cpowell

Isochrones For Dummies/Statisticians

Given the mass, age and metallicity of a star, we 'know' what its 'ideal' observation should be i.e., where it should be on the CMD.

The tables of these 'ideal' observations are called *isochrone tables*.

Why are they only 'ideal' colours/magnitudes?

Observational Error

Alas, as with every experiment there are observational errors and biases caused by the instruments.

Observational Error

Alas, as with every experiment there are observational errors and biases caused by the instruments.

1. These are relatively well understood – and can be considered to be Gaussian with *known* standard deviation.

Observational Error

Alas, as with every experiment there are observational errors and biases caused by the instruments.

1. These are relatively well understood – and can be considered to be Gaussian with *known* standard deviation.
2. Importantly, we can characterize the standard deviation as a function of the observed data.
   i.e., given $\mathbf{Y}_i = (Y_{iB}, Y_{iV}, Y_{iI})^T$ we have $\sigma_i = \sigma(\mathbf{Y}_i)$.

The Observed Data

We observe (depending on the experiment) p different colours/magnitudes for n stars.

Although it is equally straightforward to model colours $U - B, B - V$ etc., and magnitudes B, V, etc., we will stick with *magnitudes*.

The (known) standard deviations in each band are also recorded for each observation.

We also observe that we observe the n stars in the dataset and that we didn't observe any others!

The Likelihood I

$$y_i = \begin{pmatrix} \frac{1}{\sigma_i^{(B)}} B_i \\ \frac{1}{\sigma_i^{(V)}} V_i \\ \frac{1}{\sigma_i^{(I)}} I_i \end{pmatrix} \Bigg| A_i, M_i, Z \sim N\left(\tilde{f}_i, \mathbf{R}\right) \qquad i = 1, \ldots, n \qquad (1)$$

Where,

$$\tilde{f}_i = \begin{pmatrix} \frac{1}{\sigma_{Bi}} \cdot f_b(A_i, M_i, Z) \\ \frac{1}{\sigma_{Vi}} \cdot f_v(A_i, M_i, Z) \\ \frac{1}{\sigma_{Ii}} \cdot f_i(A_i, M_i, Z) \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} 1 & \rho^{(BV)} & \rho^{(BI)} \\ \rho^{(BV)} & 1 & \rho^{(VI)} \\ \rho^{(BI)} & \rho^{(VI)} & 1 \end{pmatrix}.$$

The Likelihood II

Let $S_i = 1$ if star i is observed, $S_i = 0$ otherwise.

$$S_i | \mathbf{Y}_i \sim \mathrm{Bernoulli}\left(p\left(\mathbf{Y}_i\right)\right) \tag{2}$$

where $p(\mathbf{Y}_i)$ is the probability of a star of a given magnitude being unobserved (provided by Astronomers).

The Likelihood II

Let $S_i = 1$ if star i is observed, $S_i = 0$ otherwise.

$$S_i | \mathbf{Y}_i \sim \mathrm{Bernoulli}\left(p\left(\mathbf{Y}_i\right)\right) \tag{2}$$

where $p(\mathbf{Y}_i)$ is the probability of a star of a given magnitude being unobserved (provided by Astronomers).

Note: We can also have $S_i = (S_{iB}, S_{iV}, S_{iI})^T$ and allow for some stars to be observed only in a subset of the bands.

The Parameters

Mass

Before we have any data, the prior distributions of mass and age are still not independent. We know *a priori* that old stars cannot have large mass, likewise for very young stars. Hence, we specify the prior on mass conditional on age:

$$p\left(M_i | A_i, M_{min}, M_{max}\left(A_i\right), \alpha\right) \propto \frac{1}{M_i^{\alpha}} \cdot \mathbf{1}_{\{M_i \in [M_{min}, M_{max}(A_i)]\}} \quad (3)$$

i.e. $M_i | A_i, M_{min}, M_{max}\left(A_i\right), \alpha \sim \text{Truncated-Pareto}.$

Age

For age we assume the following hierarchical structure:

$$A_i|\mu_A, \sigma_A^2 \overset{iid}{\sim} N\left(\mu_A, \sigma_A^2\right) \tag{4}$$

where $A_i = \log_{10}(\mathrm{Age})$, with μ_A and σ_A^2 hyperparameters...

Metallicity

Denoted by Z_i.

Assumed to be known and common to all stars i.e., $Z_i = Z = 4$

Hyperparameters

Next, we model the hyperparameters with the simple conjugate form:

$$\mu_A|\sigma_A^2 \sim N\left(\mu_0, \frac{\sigma_A^2}{\kappa_0}\right), \qquad \sigma_A^2 \sim Inv-\chi^2\left(\nu_0, \sigma_0^2\right) \qquad (5)$$

Where μ_0, κ_0, ν_0 and σ_0^2 are fixed by the user to represent prior knowledge (or lack of).

Correlation

We assume a uniform prior over the space of positive definite correlation matrices.

This isn't quite uniform on each of $\rho^{(BV)}, \rho^{(BI)}$ and $\rho^{(VI)}$, but it is very close.

Incompleteness

- Unfortunately, some dimmer stars may not be fully observed.
- This censoring can bias conclusions about the stellar cluster parameters.

Incompleteness

- Unfortunately, some dimmer stars may not be fully observed.
- This censoring can bias conclusions about the stellar cluster parameters.
- Since magnitudes are functions of photon arrivals, the censoring is stochastic.

Putting it all together

$$S_{ij}|\mathbf{Y}_i \sim \text{Bernoulli}\left(p\left(\mathbf{Y}_i\right)\right) \qquad i = 1, \ldots, n, n+1, \ldots, n+n_{mis} \quad j \in \{B, V, I\}$$

$$y_i = \left(\begin{array}{c} \frac{1}{\sigma_i^{(B)}} B_i \\ \frac{1}{\sigma_i^{(V)}} V_i \\ \frac{1}{\sigma_i^{(I)}} I_i \end{array} \right) \Bigg| A_i, M_i, Z \sim N\left(\tilde{f}_i, \mathbf{R}\right) \qquad i = 1, \ldots, n, n+1, \ldots, n+n_{mis}$$

$$M_i|A_i, M_{min}, \alpha \sim \text{Truncated-Pareto}\left(\alpha - 1, M_{min}, M_{max}\left(A_i\right)\right)$$

$$A_i|\mu_A, \sigma_A^2 \overset{iid}{\sim} N\left(\mu_A, \sigma_A^2\right)$$

$$\mu_A|\sigma_A^2 \sim N\left(\mu_0, \frac{\sigma_A^2}{\kappa_0}\right), \qquad \sigma_A^2 \sim Inv-\chi^2\left(\nu_0, \sigma_0^2\right)$$

$$p(\mathbf{R}) \propto \mathbf{1}_{\{\mathbf{R}\, p.d.\}}$$

Observed-Data Posterior

The product of the densities on the previous slide gives us the *complete-data posterior*. Alas, we don't observe all the stars, and n_{mis} is an unknown parameter. For now, lets just condition on n_{mis}. We have:

$$\begin{aligned}
\mathbf{W}_{obs} &= \left\{ n, \mathbf{y}_{[1:n]} = (\mathbf{y}_1, \ldots, \mathbf{y}_n), \mathbf{S} = \{1, \ldots, 1, 0, \ldots, 0)\} \right\} \\
\mathbf{W}_{mis} &= \left\{ m, \mathbf{Y}_{[(n+1):(n+m)]}, \mathbf{M}_{[(n+1):(n+m)]}, \mathbf{A}_{[(n+1):(n+m)]} \right\} \\
\Theta &= \left\{ \mathbf{M}_{[1:n]}, \mathbf{A}_{[1:n]}, \mu_A, \sigma_A^2, \mathbf{R} \right\}
\end{aligned}$$

where $\mathbf{X}_{a:b}$ denotes the vector $(X_a, X_{a+1}, \ldots, X_b)$

Observed-Data Posterior

We want $p(\Theta|\mathbf{W}_{obs})$ but so far we have $p(\Theta, \mathbf{W}_{mis}|\mathbf{W}_{obs})$. So, we integrate out the missing data:

$$p(\Theta|\mathbf{W}_{obs}) = \int p(\Theta, \mathbf{W}_{mis}|\mathbf{W}_{obs})\, d\mathbf{W}_{mis} \tag{6}$$

In practice, this integration is done by sampling from $p(\Theta, \mathbf{W}_{mis}|\mathbf{W}_{obs})$ and retaining only the samples of Θ.

Observed-Data Posterior

We form a Gibbs sampler to sample from $p\left(\Theta, \mathbf{W}_{mis} | \mathbf{W}_{obs}\right)$. Given a current state of our Markov Chain, $\Theta = \Theta^{(t)}$ and $\mathbf{W}_{mis} = \mathbf{W}_{mis}^{(t)}$:

1. Draw $\Theta^{(t+1)}$ from $p\left(\Theta | \mathbf{W}_{mis}^{(t)}, \mathbf{W}_{obs}\right)$ (as before)
2. Draw $\mathbf{W}_{mis}^{(t+1)}$ from $p\left(\mathbf{W}_{mis} | \Theta^{(t+1)}, \mathbf{W}_{obs}\right)$ (new)

SAMPLING $\mathbf{W}_{mis}$

At each iteration of the Gibbs sampler we need to draw the missing data from the appropriate distribution.

In other words, given a bunch of masses, ages, and metallicities of n_{mis} missing stars, find a bunch of $\mathbf{Y}_i$'s that are consistent with that:

$$p_i\left(Y_i|\mathbf{Y}_{[1:n]},\mathbf{M},\mathbf{A},\mu_A,\sigma_A^2\right)\propto[1-\pi\left(\mathbf{Y}_i\right)]\cdot \tag{7}$$

$$\exp\left\{-\frac{1}{2}\left(\mathbf{Y}_i-\tilde{f}(\mathbf{Y}_i;M_i,A_i,Z)\right)^T R^{-1}\left(\mathbf{Y}_i-\tilde{f}(\mathbf{Y}_i;M_i,A_i,Z)\right)\right\} \tag{8}$$

for $i=n+1,\ldots,n+m$.

SAMPLING $\mathbf{W}_{mis}$

Once we have sampled a new set of $\mathbf{Y}_{mis}$, we need to sample the standard deviation of the Gaussian error for those stars.

Here we assume this is a deterministic mapping: $\sigma = \sigma(\mathbf{Y}_i)$.

Sampling from the posterior

Some notes:

Sampling from the posterior

Some notes:

1. We have our model – what does our posterior look like?

Sampling from the posterior

Some notes:

1. We have our model – what does our posterior look like?
2. Ugly. No chance of working with it analytically ⇒ MCMC!

Sampling from the posterior

Some notes:

1. We have our model – what does our posterior look like?
2. Ugly. No chance of working with it analytically ⇒ MCMC!
3. Going to have to be a Gibbs sampler. How to break it up?

Sampling from the posterior

Some notes:

1. We have our model – what does our posterior look like?
2. Ugly. No chance of working with it analytically $\Rightarrow$ MCMC!
3. Going to have to be a Gibbs sampler. How to break it up?
4. Mass and Age are going to be extremely highly correlated (i.e., sample jointly)

Sampling from the posterior

Some notes:

1. We have our model – what does our posterior look like?
2. Ugly. No chance of working with it analytically ⇒ MCMC!
3. Going to have to be a Gibbs sampler. How to break it up?
4. Mass and Age are going to be extremely highly correlated (i.e., sample jointly)
5. No analytic simplification for terms in M_i, A_i because of f

Sampling from the posterior

Some notes:

1. We have our model – what does our posterior look like?
2. Ugly. No chance of working with it analytically ⇒ MCMC!
3. Going to have to be a Gibbs sampler. How to break it up?
4. Mass and Age are going to be extremely highly correlated (i.e., sample jointly)
5. No analytic simplification for terms in M_i, A_i because of f
6. High dimensional multi-modal, so we also use parallel tempering.

Parallel Tempering

A brief overview of parallel tempering:

The parallel tempering framework involves sampling N chains, with the i^{th} chain of the form:

$$p_i(\theta) = p(\theta|\mathbf{y})^{1/t_i} \propto \exp\left\{-\frac{H(\theta)}{t_i}\right\} \tag{9}$$

As t_i increases the target distributions become flatter.

Simulation Results

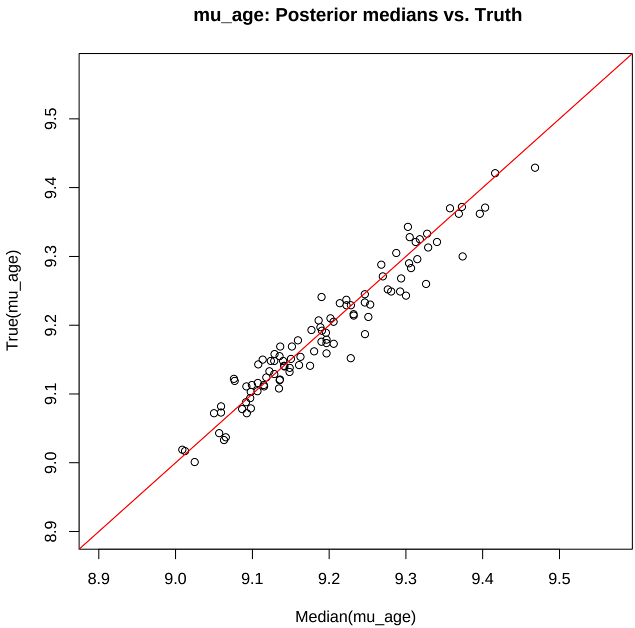
We simulate 100 datasets from the model with $n = 100$:

$$\mu_A = 9.2 \ \sigma_A^2 = 0.01^2$$

$$M_{(min)} = 0.8 \ \alpha = 2.5$$

$$\mathbf{R} = \mathbf{I}$$

$$(\sigma_{B_i}, \sigma_{V_i}, \sigma_{I_i}) \in (0.03, 0.12)$$

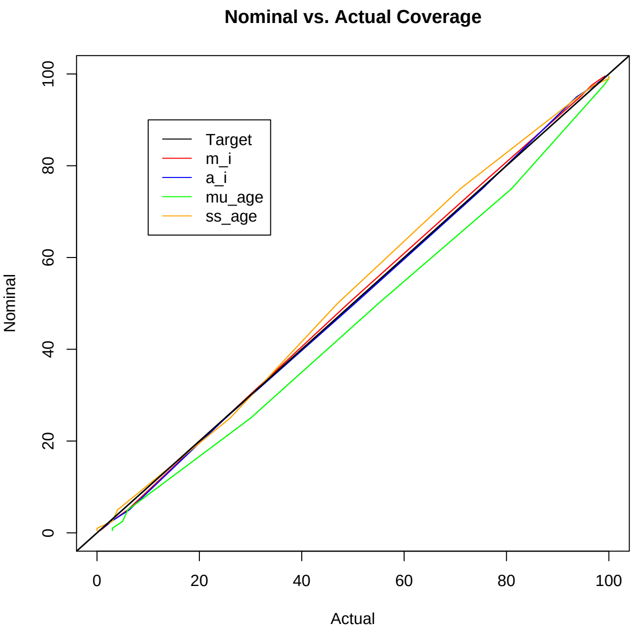

mu_age: Posterior medians vs. Truth
True(mu_age)
Median(mu_age)
8.9
9.0
9.1
9.2
9.3
9.4
9.5

Does it work?

Post_p	0.5	1.0	2.5	5.0	25.0	50.0
m_cover	0.6	1.2	2.8	6.0	25.0	49.1
a_cover	0.4	1.1	2.8	6.3	25.1	50.4
mu_age	3.0	3.0	5.0	6.0	30.0	55.0
ss_age	0.0	0.0	3.0	4.0	26.0	47.0

Does it work?

Post_p	50.0	75.0	95.0	97.5	99.0	99.5
m_cover	49.1	74.1	94.4	96.6	98.5	99.2
a_cover	50.4	75.3	93.7	97.2	99.0	99.3
mu_age	55.0	81.0	97.0	99.0	100.0	100.0
ss_age	47.0	71.0	94.0	97.0	100.0	100.0


Nominal vs. Actual Coverage
Target
m_i
a_i
mu_age
ss_age
Nominal
Actual
0
20
40
60
80
100

Future Work

Some important things still need to be built into the model before it is fit for purpose:

- **Extinction/Absorption:** Shift in observed data
- **Multi-Cluster Models:** Allow for multiple stellar clusters