A Statistical Approach to Stellar Archaeology

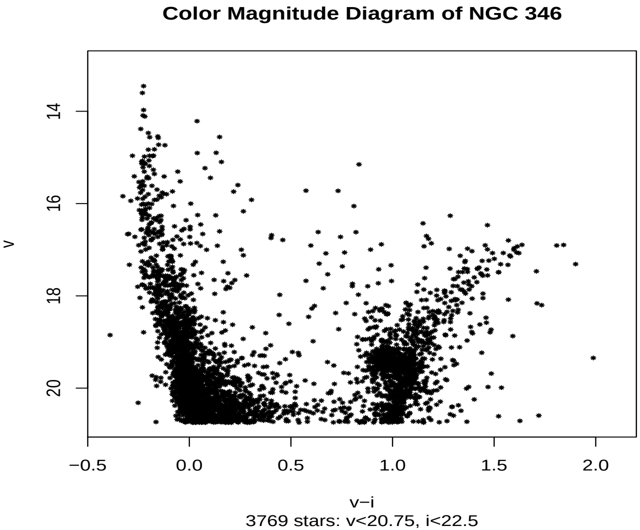
H. Lee, A. Zezas, and V. Kashyap

Harvard Smithsonian Center for Astrophysics

Aug. 1, 2007

Color-Magnitude Diagram (CMD)

The CMD of our target data set.



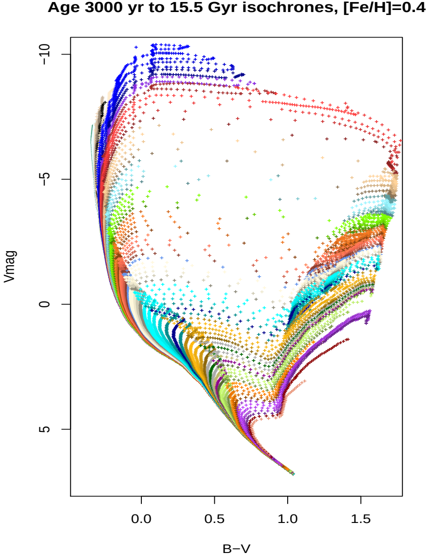
Isochrones - Reference Tables

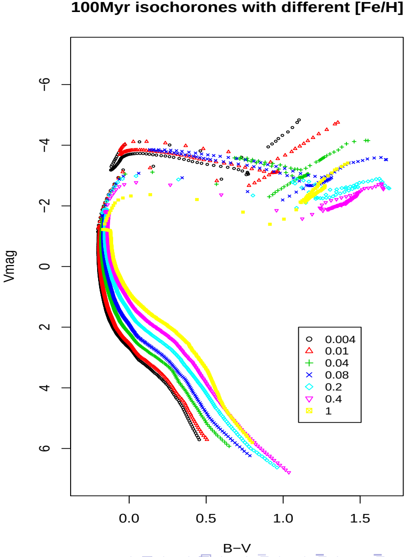
Isochrones reveal physical informations from CMD (Temp, mass, [Fe/H], age, class, etc.)

- Realization of theoretical models (quite complicated).
- Depends on so many input parameters.
- Tables tell the location of a star in the diagram (color vs magnitude) at the given age, metalicity, mass, etc.
- A typical inverse problem in astronomy

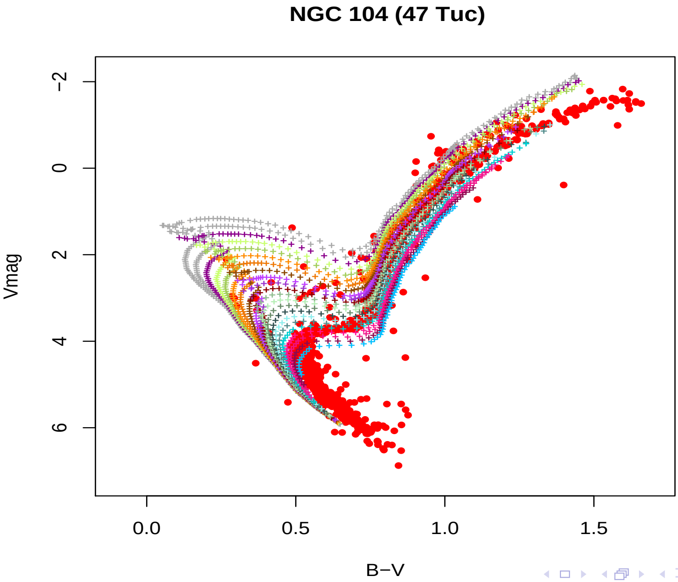
How these isochrones look like on CMD?

Isochrones continued- How they look?





47 Tuc: 11.2 Gyr Old Globular Cluster



Motivation

CMD (data) and Isochrones (model) are available and we like to know the age distribution of stellar clusters in a statistical fashion.

Statistical Modeling- Simplest

Bayes Rule: the age (τ) posterior distribution is proportional to the likelihood times a prior. For a single star,

$$p(\tau|M_i, C_i) \propto l_i(M_i, C_i|\tau)p(\tau),$$

where M_i is magnitude (M_v) and C_i color ($B - V$). Therefore,

$$p(\tau|\{M_i\}, \{C_i\}) \propto p^n(\tau) \prod_{i=1}^{n} l_i(M_i, C_i|\tau)$$

We like to focus on the likelihood,

$$\prod_{i=1}^{n} l_i(M_i, C_i|\tau).$$

How can we estimate this likelihood?

Information Theory: K-L distance

The Kullback-Leibler distance is defined to be

$$D(f_\tau(x); g(x|\tau)) = \int \log f_\tau(x) f_\tau(x) dx - \int \log g(x|\tau) f_\tau(x) dx \geq 0$$

We do not know the true age density $f_\tau(x)$ but introducing $g(x|\tau)$ and maximizing $E_\tau[(\log g(X|\tau)]$ provides the best τ **(age)** for a given $g(x|\tau)$, where x is observed and its random variable is denoted by X.

$$\hat{\tau} = \arg \max_{\tau \in \{\tau_j\}} E_\tau[\log g(X|\tau)]$$

To estimate $E_\tau[\log g(X|\tau)]$, we used the empirical mean of the log likelihood.

$$E_\tau[\log g(X|\tau)] = \frac{1}{n}\sum_{i=1}^{n} \log g(x_i|\tau) + \frac{b}{n},$$

where b is a bias term such as the penalty terms in AIC and BIC. We assume $\frac{b}{n} \to 0$. Finding τ that maximizes $\frac{1}{n}\sum_{i=1}^{n} \log g(x_i|\tau)$ leads to the best guess for the age of the stellar cluster.

Then, what would be the likelihood?

Likelihood

We took **Multivariate Normal** to establish the likelihood based on additional informations from data:

- Errors on each observation (σ_i are known)
- Independence among color bands (U,B,V,I,R,etc)
- **Multivariate Normal** assumption is quite reasonable

For a star i, the recorded value is apparent magnitudes (v_i, b_i), corrected by the appropriate distance modulus and the extinction laws (these corrections have their own uncertainties but we ignore at the moment).

Likelihood - continued

By denoting this corrected value as $(M_i, C_i) = x_i$, the likelihood of a star is

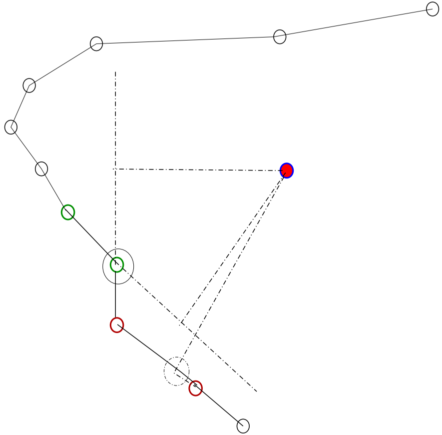
$$l_i(M_i, C_i|\tau) = \frac{1}{2\pi|\Sigma_i|^{1/2}} exp(-\frac{1}{2}(x_i - \mu_i)^T \Sigma_i^{-1}(x_i - \mu_i)),$$

where Σ_i is a covariance matrix for star i and $\mu_i = \mu_i(\tau, [Fe/H], mass, class)$ although the function μ_i is unknown. As indicated, the best tactic is maximizing the likelihood and this is achieved by finding μ_i that minimize the distance to x_i. Given τ_j ($j = 1, ..., 91$, Geneva models provide 91 age grids), $\mu_i(h_j)$ minimizes the distance to x_i.

But how to minimize the distance?

Defining a point of min. distance

Finding a point of minimum distance, associated with a complicated curve, only represented by a set of points.
→ **piece-wise Euclidean distance**



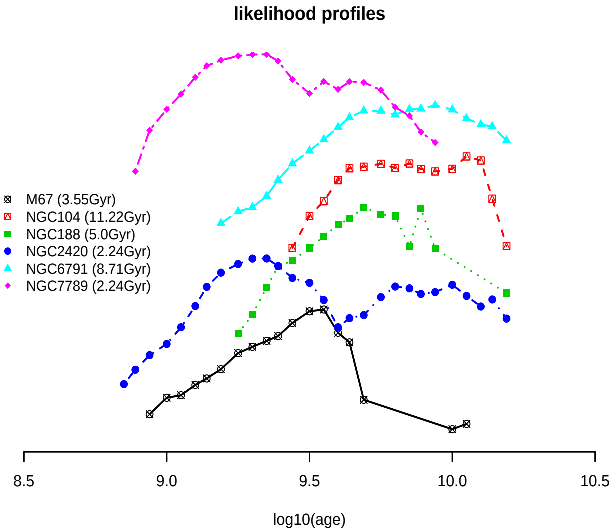
Applying the model selection method

Real Data Application:

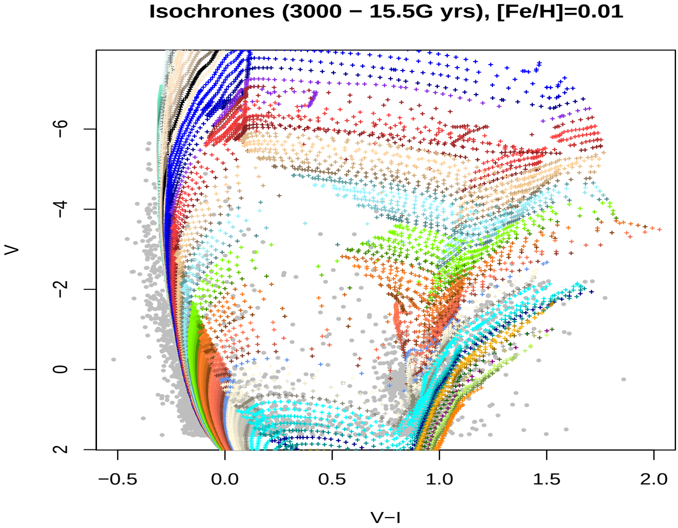
NGC	[Fe/H]	$(m-M)_o$	$E(B-V)$	Age[1](Gyr)	Est. Age
104 (Tuc47)	0.004	13.33	0.04	10.9±1.4	11.22
188	0.02	11.17	0.09	6.3±0.8	5.0
2420	0.007	11.94	0.05	2.2±0.3	2.24
2682 (M67)	0.02	9.59	0.04	4.3±0.5	3.55
6791	0.050	12.96	0.15	10.2±1.2	8.71
7789	0.014	11.22	0.29	1.80±0.3	2.24

[1]from Salaris et.al. (2004)

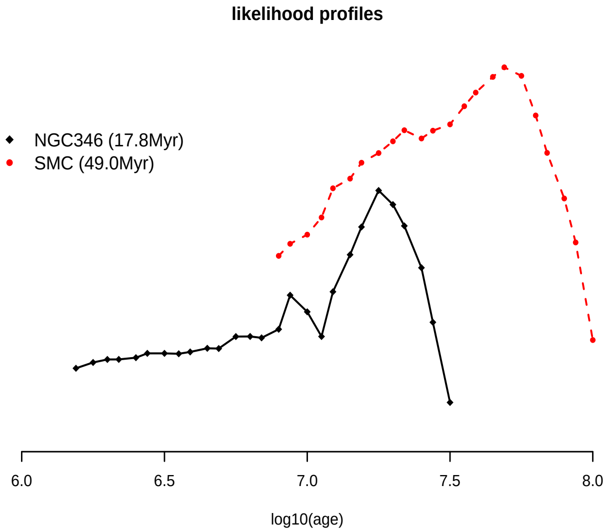
Log Likelihood Profiles of Some Stellar Clusters



Age of NGC 346 - part of SMC



Log Likelihood Profiles of NGC 346 and SMC



Discussion

The model selection by maximizing likelihoods is

- **empirical method:** the correctness of ages highly depends on data. (data processing affects results: e.g. foreground stars and covariance matrix).
- **easy, quick, simple, heuristic**, and **diagnostic.**

Requires a fine tuning: **Extention to Bayesian approaches**

- posterior distribution from marginalizing the initial mass function (IMF)

$$\begin{aligned} p(\tau|\{M_i\},\{C_i\}) &= \int p(\tau, m|\{M_i\},\{C_i\})dm \\ &\propto \int \prod_{i=1}^{n} l_i(M_i, C_i|\tau, m)p(\tau, m)dm \end{aligned}$$

- developing **hierarchical models** to incorporate not only IMF but **metalicity, classes, completeness, and uncertainties, etc...**

Thank you!

Please visit AstroStatistics blog at
http://groundtruth.info/AstroStat/slog/