Multivariate Evolutionary Classification in Astrophysics

(Astrocladistics)

Didier Fraix-Burnet

Biologists  P. Choler, E. Douzery

Astrophysicists  E. Davoust, T. Chattopadhyay, C. Charbonnel, F. Lamareille

Statisticians  A.K. Chattopadhyay, (T. Chattopadhyay)

« Soft-Computing »  M. Thuillard

Students  A. Verhamme, M. Dugué

Outline

- Why classification?
- Why be multivariate?
- Why be evolutive?
- Why cladistics?
- Classification = clustering + taxonomy
- Confounding correlations
- Cladistics with continuous characters
- Spectra
- Classification and dimensionality reduction
- Astrostatistics: what's next?
Stars

Dust

Gas

Interacting or merging galaxy

Dwarf galaxy
Why classification?
Why be multivariate?

Globular clusters of our Galaxy

<table>
<thead>
<tr>
<th></th>
<th>MRC</th>
<th>MPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>All [Fe/H] &gt; -1</td>
<td>14</td>
<td>40</td>
</tr>
<tr>
<td>R_{gc} = 0 - 4 kpc</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>R_{gc} = 4 - 9 kpc</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>All [Fe/H] &lt; -1</td>
<td>40</td>
<td>23</td>
</tr>
<tr>
<td>R_{gc} = 0 - 4 kpc</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>R_{gc} = 4 - 8 kpc</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>R_{gc} = 8 - 12 kpc</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>R_{gc} = 12 - 20 kpc</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>R_{gc} &gt; 20 kpc</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>-2.30 &lt; [Fe/H] ≤ -1.85</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>-1.85 &lt; [Fe/H] ≤ -1.65</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>-1.65 &lt; [Fe/H] ≤ -1.50</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>-1.50 &lt; [Fe/H] ≤ -1.32</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>-1.32 &lt; [Fe/H] ≤ -1.00</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>All [Fe/H] &lt; -1.70</td>
<td>27</td>
<td>20</td>
</tr>
<tr>
<td>HBR &gt; 0, R_{gc} &gt; 8 kpc</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>HBR &lt; 0, R_{gc} &gt; 8 kpc</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>HBR &lt; 0</td>
<td>18</td>
<td>6</td>
</tr>
</tbody>
</table>

Metallicity Fe/H

R (distance from galactic center)

Horizontal Branch
Why be multivariate and evolutive?

Simple resemblance is not enough to describe and to understand
Classification, complexity, evolution

<table>
<thead>
<tr>
<th>Appearance</th>
<th>Few parameters</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global similarity</td>
<td>All parameters</td>
<td>Distance cluster analyses</td>
</tr>
<tr>
<td>Common history</td>
<td>Evolutive characters</td>
<td>Cladistics</td>
</tr>
</tbody>
</table>

Global similarity (distance based)

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Evolved states: 0 | 1 | 2

Common history (character based)

B <-> A <-> C
Astrocladistic analysis of Globular Clusters of our Galaxy

4 parameters:

LogTe
Fe/H
Mv
Age (1/2)

Fraix-Burnet, Davoust, Charbonnel, MNRAS 2009
The fundamental plane of galaxies

699 galaxies, 4 characters: $\sigma, \mu_e, R_e, M_{g2}$

Cluster analysis (K-means)

Cladistics

Fraix-Burnet, Dugué, Chattopadhyay, Chattopadhyay, Davoust
2010 MNRAS
Confounding Correlations

Evolution as a confounding parameter

\[
\begin{align*}
  r_e &= A_1 \tilde{X}^p \\
  \sigma &= A_2 \tilde{X}^s \\
  L &= A_3 \tilde{X}^t
\end{align*}
\]

For the observed fundamental plane:

\[
\log r_e = a \log \sigma + b \mu_e + c
\]

\[\Rightarrow p = sa + (-2.5t + 5p)b\]

X can be fraction of starburst, black hole mass, anything related to global evolution

If \[2s + p = 1\] then « virial plane ».
Minimum Contradiction Analysis

Continuous characters in cladistics

Tree
Split network

Perfect order

Contradiction = distance to perfect order

Kalmanson inequalities

\[ Y_{i,j}^n \geq Y_{i,i}^n, \quad Y_{k,j}^n \geq Y_{k,k}^n \quad (i \leq j \leq k) \text{ with } Y_{i,j} = 1/2 \cdot (d_{i,n} + d_{j,n} - d_{i,j}) \]

\[ C = \sum_{k \geq j \geq i} \left( \max \left( \left( Y_{i,k}^n - Y_{i,j}^n \right), 0 \right) \right)^\beta + \sum_{k \geq j > i} \left( \max \left( \left( Y_{i,k}^n - Y_{j,k}^n \right), 0 \right) \right)^\beta \]

Thuillard, Fraix-Burnet, *Evolutionary Bioinformatics* 2009

Convex hull

Conditions on parameters
Spectra

K-means on 700,000 spectra → classes

Sanchez-Almeida et al 2010
Gosh, Chattopadhyay, Fraix-Burnet

Minimum Spanning Tree or Cladistics

Ascasibar & Sanchez-Almeida 2011
Object Space

Classification

Model Fitting

Parameter Space

Sub Parameter Space

Dimensionality Reduction

reduce exemplars

test geometry

return unsuitable parameters

test geometry

preserve geometry

simplify description

simplify understanding
Astrostatistics : what's next ?

Training and culture:
• Statistical tools
• Statistical inference ≠ astrophysical inference

Visibility/recognition of astrostatistics
• Many works
• Many needs
• Necessity of an official structure

Animation
• Focussed and practical workshops
• Re-use methodologies (not always developments)