Enabling Data Science in the Gaia Mission Archive: IMF and SFR

Gaia CU9 Grand Challenge

Data Mining group, Spanish Virtual Observatory

30 de marzo de 2016

This presentation is based in a recent paper submitted to the Information Sciences journal: *Enabling Data Science in the Gaia Mission Archive: The Initial Mass Function and the Star Formation Rate* by D. Tapiador, A. Berihuete, L.M. Sarro, F. Julbe, E. Huedo.

This resentation is in the web, here only a brief summary

Goal: to stablish a probabilistic framework in terms of a **Hierarchical Bayesian Modelling**

- First step: independent models for IMF and SFR (PhD D. Tapiador)
- Second step (work in progress): a join probabilistic model (IMF,SFR)
- … towards a more realistic approach

It is **the first contextualization of the Gaia Grand Challenges**

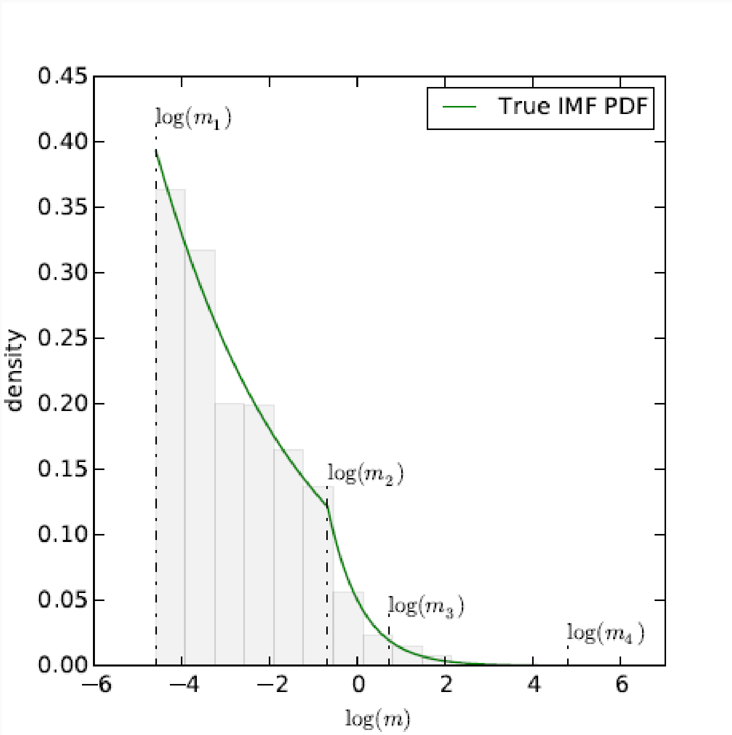
- Executed at Gaia Data Analytic Framework (GDAF)

Complex Hiearchical Bayesian Model (ask authors!!):

- The vector model parameter is itself treated as a random variable, the distribution of which is aimed to infer

Tools: emcee algorithm in Spark platform coding in python

The initial mass function



Kroupa IMF

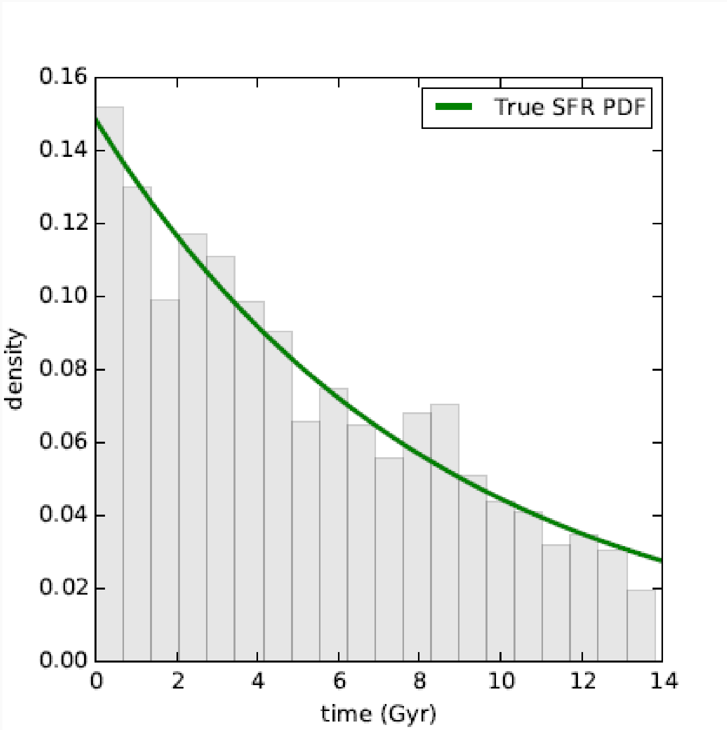
The Initial Mass Function (IMF) describes the distribution of initial true (**unknown**) masses for a population of stars. We establish the hypothesis that the IMF can be expressed as

$$\xi(m; \boldsymbol{\theta}) = c_j m^{-\theta_j},\ M_j < m \leq M_{j+1},$$

where, $j = 1, 2, 3$.

IMF PDF by setting $\boldsymbol{\theta} = (1.3, 2.3, 2.3)$

The star formation rate



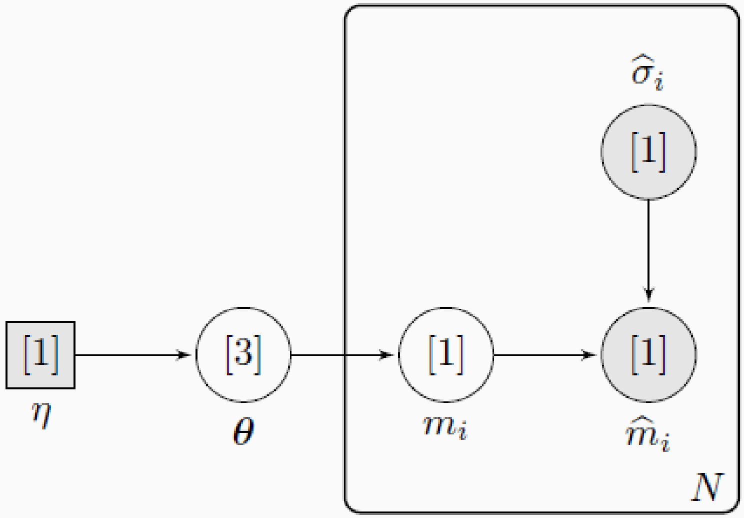
SFR PDF

Instead of $\zeta(t)$ which has units of solar masses per unit time, we will work with $\zeta'(t)$ defined as the fraction of stars (of whatever mass) created per unit time:

$$\zeta'(t) = \frac{\sum_k w_k \phi_k(t)}{\sum_k w_k \int_a^b \phi_k(t)\,\mathrm{d}t}.$$

Thus, the SFR defines a non-stationary Poisson process with intensity function $\zeta'(t)$, $t \geq 0$. Figure shows

$$\zeta'(t) = \frac{0.12}{(1 - \exp(-13.8 \cdot 0.12))} \exp(-0.12t)$$



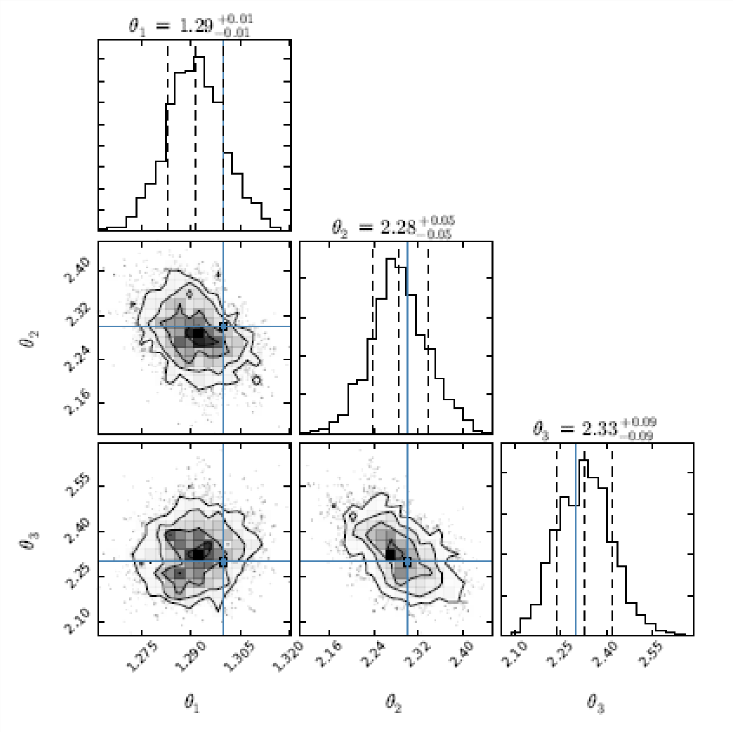
So what?

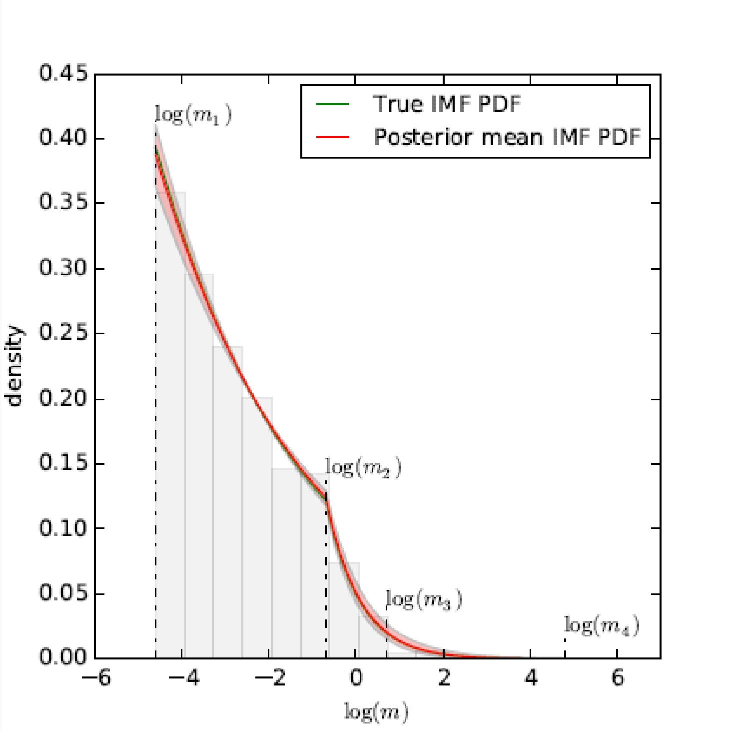
The main goal of astronomers is to move in the inverse way: from observations to model parameters. In our example, from the masses and their uncertainties, $\mathcal{D} = \{\hat{m}_i, \hat{\sigma}_i\}_{i=1}^{N}$, to the parameter $\boldsymbol{\theta}$.

Data set available from Gaia: the posterior samples of mass and ages produced by the FLAME module

1 million sources have been simulated, with 50 masses and ages for each source (realistic errors have been assumed)

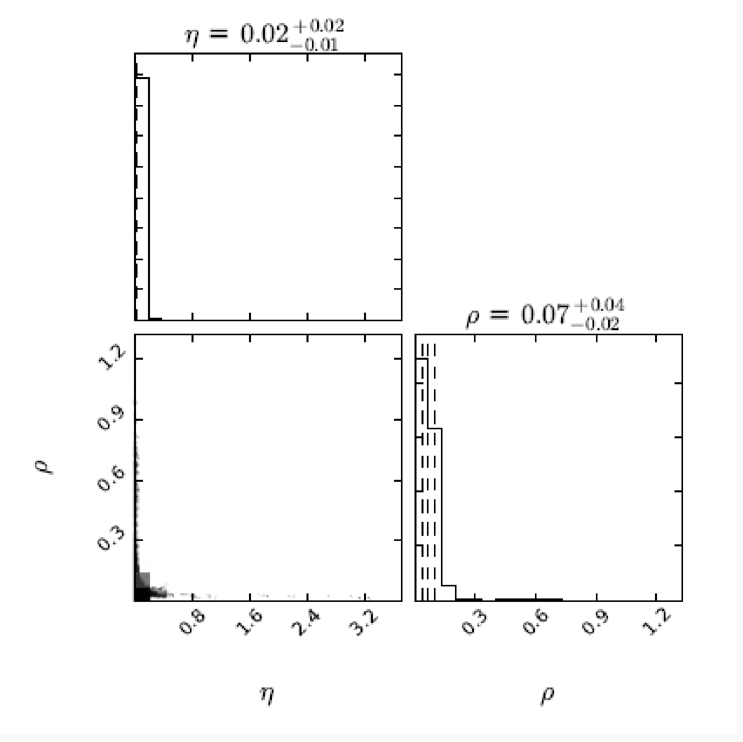
The initial mass function

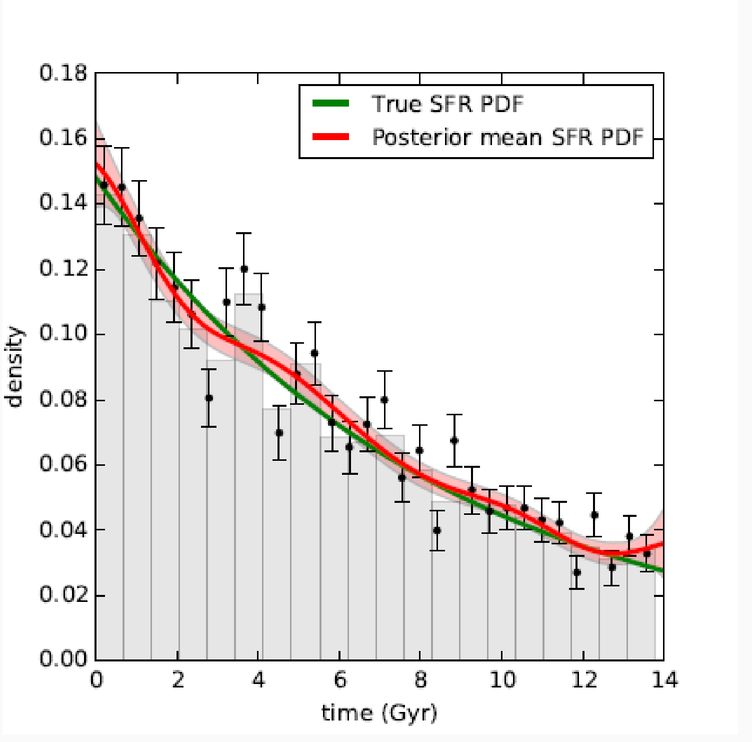




The true value was $\boldsymbol{\theta} = (1.3, 2.3, 2.3)$, and the estimated value $\hat{\boldsymbol{\theta}} = (1.29^{+0.01}_{-0.01}, 2.28^{+0.05}_{-0.05}, 2.33^{+0.09}_{-0.09})$ obtained by calculating the quantiles 0.16, 0.50 and 0.84 of the posterior samples.

The star formation rate





The true SFR PDF (green line) was

$$\zeta'(t) = \frac{0.12}{(1 - \exp(-13.8 \cdot 0.12))} \exp(-0.12t)$$