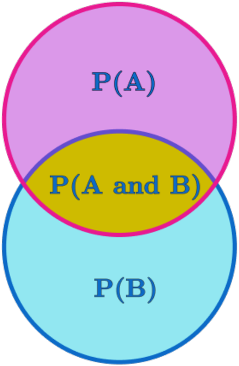
Examples of Markov-Chain Monte-Carlo Methods in Astronomy

Leandro Rocha Rímulo

Departamento de Física
Universidad de los Andes, Colombia

12 February 2019

Sets and probabilities



Example: twenty-sided dice

- whole set: $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$;
- even number: $X = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$;
- prime number: $Y = \{2, 3, 5, 7, 11, 13, 17, 19\}$;

Therefore,
$P(\Omega) = \frac{20}{20} = 1$
$P(X) = \frac{10}{20} = 0.5$
$P(Y) = \frac{8}{20} = 0.4$

Example: twenty-sided dice

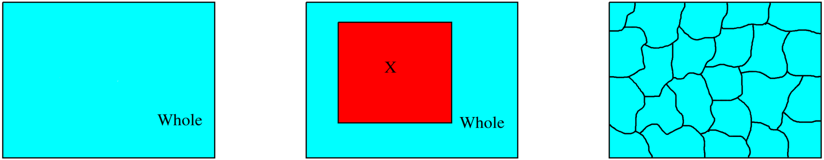
- whole set:
  $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$;
- even number: $X = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$;
- prime number: $Y = \{2, 3, 5, 7, 11, 13, 17, 19\}$;

- even number AND prime number: $X \cap Y = \{2\}$;
  $P(X \cap Y) = \frac{1}{20} = 0.05$
- even number OR prime number:
  $X \cup Y = \{2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20\}$
  $P(X \cup Y) = \frac{17}{20} = 0.85$
- odd number: $\Omega - X = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$;
  $P(\Omega - X) = \frac{10}{20} = 0.5$
- odd number that is NOT a prime number: $(\Omega - X) - Y = \{1, 9, 15\}$;
  $P((\Omega - X) - Y) = \frac{3}{20} = 0.15$

Sets and probabilities

The requirements for a function P to be a probability function are:

- $0 \leq P(X) \leq 1, \forall X \subseteq \Omega$
- $P(\Omega) = 1$
- $P\left(\bigcup_i X_i\right) = \sum_i P(X_i)$, where $X_i \cap X_j = \emptyset$



...where $\Omega \neq \emptyset$ is the set of all possible outcomes (the "ensemble"), and X is a set of zero or more possible outcomes.

Conditional probabilities:

Let there be two events X and W. We define the probability of "X given W" as

$$P\left(X \mid W\right) = \frac{P(X \cap W)}{P(W)} \tag{1}$$

Conditional probabilities:

Let there be two events X and W. We define the probability of "X given W" as

$$P\left(X \mid W\right) = \frac{P(X \cap W)}{P(W)} \tag{1}$$

Therefore,

- $0 \leq P(X \mid W) \leq 1, \forall X \subseteq \Omega, W \subseteq \Omega$
- $P(W \mid W) = 1$
- $P\left(\bigcup_{i=1}^{N} X_i \mid W\right) = \sum_{i=1}^{N} P(X_i \mid W)$, where $X_i \cap X_j = \emptyset$

Example: Coloured balls with numbers

- blue balls with prime numbers from 1 to 30:
  $B = \{b2, b3, b5, b7, b11, b13, b17, b19, b23, b29\}$.
- red balls with odd numbers from 1 to 30:
  $R = \{r1, r3, r5, r7, r9, r11, r13, r15, r17, r19, r21, r23, r25, r27, r29\}$.
- balls with number 11: $E = \{b11, r11\}$.

Red ball with number 11: $R \cap E = \{r11\}$.
$P(R \cap E) = \frac{1}{25}$.
But also,
$P(R \cap E) = P(R \mid E)P(E) = \frac{1}{2} \times \frac{2}{25} = \frac{1}{25}$
and
$P(R \cap E) = P(E \mid R)P(R) = \frac{1}{15} \times \frac{15}{25} = \frac{1}{25}$.

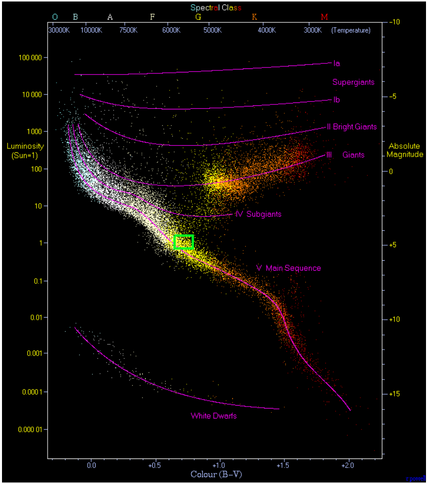
Example: Coloured balls with numbers

- blue balls with prime numbers from 1 to 30: $B = \{b2, b3, b5, b7, b11, b13, b17, b19, b23, b29\}$.
- red balls with odd numbers from 1 to 30: $R = \{r1, r3, r5, r7, r9, r11, r13, r15, r17, r19, r21, r23, r25, r27, r29\}$.
- balls with number 11: $E = \{b11, r11\}$.

Red ball with number 11: $R \cap E = \{r11\}$.
$P(R \cap E) = \frac{1}{25}$.
But also,
$P(R \cap E) = P(R \mid E)P(E) = \frac{1}{2} \times \frac{2}{25} = \frac{1}{25}$
and
$P(R \cap E) = P(E \mid R)P(R) = \frac{1}{15} \times \frac{15}{25} = \frac{1}{25}$.

Notice that,
$P(R)P(E) = \frac{15}{25} \times \frac{2}{25}$.
Therefore, $P(R \cap E) \neq P(R)P(E)$!

Probability density function (pdf):



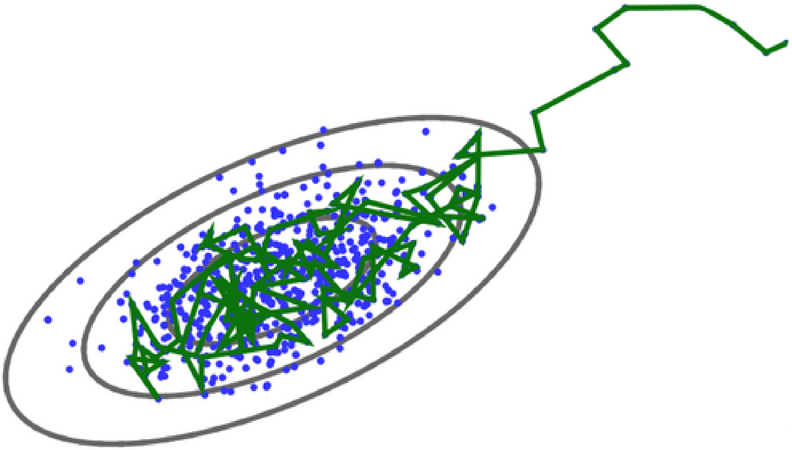
- $\boldsymbol{\theta}$ is an N-dimensional vector
- $\mathrm{d}W$ is an infinitesimal set that contains all elements $\boldsymbol{\theta}$ contained in the infinitesimal volume $\mathrm{d}V_{\boldsymbol{\theta}}$.

Therefore, the probability of $\mathrm{d}W$ is

$$P(\mathrm{d}W) = f(\boldsymbol{\theta})\mathrm{d}V_{\boldsymbol{\theta}} \,,$$

where f is the **pdf**.

Markov-Chain Monte Carlo (MCMC)



In statistics, the quantities of interest are usually integrals of functions ϕ, weighed by the probability density p, over a certain domain: $\int p(\boldsymbol{\theta})\phi(\boldsymbol{\theta})\mathrm{d}V_{\boldsymbol{\theta}}$.

Example:
Maxwell-Boltzmann distribution:

$$f(v) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{a^3} v^2 e^{-\frac{v^2}{2a^2}} \,, \text{ where } a = \left(\frac{kT}{m}\right)^{\frac{1}{2}}$$

- $\langle v \rangle = \int_0^\infty f(v) v \mathrm{d}v = \left(\frac{8}{\pi} a^2\right)^{\frac{1}{2}}$
- $v_{\mathrm{rms}} = \left(\int_0^\infty f(v) v^2 \mathrm{d}v\right)^{\frac{1}{2}} = \left(3a^2\right)^{\frac{1}{2}}$

Markov-Chain Monte Carlo (MCMC)

Some Markov chains have **stationary distributions** that are asymptotically reached by the states after a sufficiently large number of iterations, **regardless of the initial state of the chain**.

MCMC methods are a class of algorithms, which utilize Markov chains, for generating samples of states $\{\boldsymbol{\theta}_r\} |_{r=1}^{R}$ from **a probability distribution** $p(\boldsymbol{\theta})$ **which is given as input**.

The algorithms construct a Markov chain that has the distribution $p(\boldsymbol{\theta})$ as its stationary distribution (MacKay, 2003).

Markov-Chain Monte Carlo (MCMC)

The output of the MCMC algorithm is the sample of states $\{\boldsymbol{\theta}_r\} |_{r=1}^{R}$ and the stationary sample is such that $p(\boldsymbol{\theta})\mathrm{d}V_{\boldsymbol{\theta}} \to 1/R$, where $\mathrm{d}V_{\boldsymbol{\theta}}$ is a measure of the hypervolume occupied by a single state.

Therefore, **for the MCMC sample**:

$$\int p(\boldsymbol{\theta})\phi(\boldsymbol{\theta})\mathrm{d}V_{\boldsymbol{\theta}} \to \frac{1}{R}\sum_{i=1}^{R}\phi(\boldsymbol{\theta}_i)$$

Obs: the probability density $p(\boldsymbol{\theta})$ **does not need to be normalized**, because the algorithms deal only with ratios of these distributions.

In this work, we use `emcee` (Foreman-Mackey et al., 2013)[1], which is Python implementation of several emsemble samplers:

- the Affine Invariant MCMC Ensemble sampler of Goodman and Weare (2010).
- Parallel-tempering sampler.
- the Metropolis-Hastings sampler.

The states are referred to as walkers.

[1] `http://dfm.io/emcee/current/`

Examples of 1-D distributions (`example000.py`):

Gaussian:

$$p(x) \propto e^{-\frac{x^2}{2\sigma^2}}\,, \text{ for all } x$$

Sinusoidal:

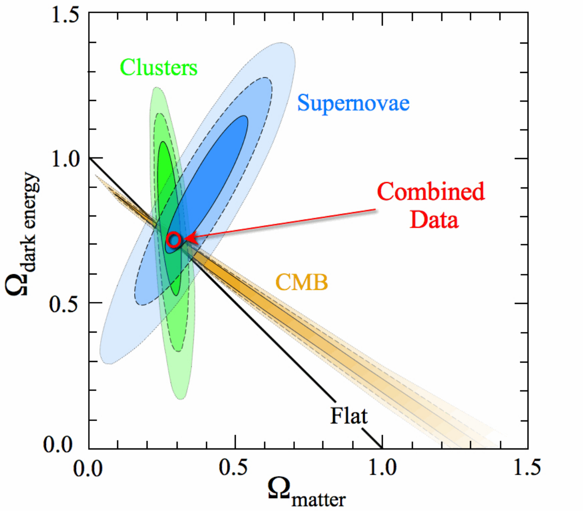
$$p(x) \propto A + \sin^2(3\pi x)\,, \text{ for } 0 \leq x \leq 1$$

Maxwell-Boltzmann:

$$f(v) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{a^3} v^2 e^{-\frac{v^2}{2a^2}}\,, \text{ for } x \geq 0$$

Examples of applications

Estimating model parameters (Bayesian inference)



Estimating model parameters

Question: What is the probability of a small subset of model parameters, given a small subset of data parameters? **Answer**: $P(\mathrm{d}M \mid \mathrm{d}D)$

$$
\begin{aligned}
P(\mathrm{d}M \mid \mathrm{d}D) &= \frac{P(\mathrm{d}M \cap \mathrm{d}D)}{P(\mathrm{d}D)} \\
&= \frac{P(\mathrm{d}D \mid \mathrm{d}M)P(\mathrm{d}M)}{P(\mathrm{d}D)}
\end{aligned}
$$

Estimating model parameters

Question: What is the probability of a small subset of model parameters, given a small subset of data parameters? **Answer**: $P(\mathrm{d}M \mid \mathrm{d}D)$

$$
\begin{aligned}
P(\mathrm{d}M \mid \mathrm{d}D) &= \frac{P(\mathrm{d}M \cap \mathrm{d}D)}{P(\mathrm{d}D)} \\
&= \frac{P(\mathrm{d}D \mid \mathrm{d}M)P(\mathrm{d}M)}{P(\mathrm{d}D)}
\end{aligned}
$$

Replacing the small probabilities by pdfs:

- Prior: $P(\mathrm{d}M) = \pi(\boldsymbol{\theta})\mathrm{d}V_{\boldsymbol{\theta}}$
- Likelihood: $P(\mathrm{d}D \mid \mathrm{d}M) = L(\boldsymbol{X} \mid \boldsymbol{\theta})\mathrm{d}V_{\boldsymbol{X}}$
- Posterior: $P(\mathrm{d}M \mid \mathrm{d}D) = p(\boldsymbol{\theta} \mid \boldsymbol{X})\mathrm{d}V_{\boldsymbol{\theta}}$
- Evidence: $P(\mathrm{d}D) = g(\boldsymbol{X})\mathrm{d}V_{\boldsymbol{X}}$

Therefore, the pdf to be samplped is

$$
p(\boldsymbol{\theta} \mid \boldsymbol{X}) = \frac{L(\boldsymbol{X} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{g(\boldsymbol{X})} \tag{2}
$$

The uv-model (my invention) (example_uv.py)

Let $\boldsymbol{\theta} = (u, v)$ be the model parameters,
such there are two observables: $\boldsymbol{X} = (x, y)$.

Therefore, the observables x and y depend on the model:
$x = x(u, v)$
$y = y(u, v)$.

The uv-model (my invention) (example_uv.py)

Let $\boldsymbol{\theta} = (u, v)$ be the model parameters,
such there are two observables: $\boldsymbol{X} = (x, y)$.

Therefore, the observables x and y depend on the model:
$x = x(u, v)$
$y = y(u, v)$.

Suppose each measurement has a gaussian error.
Therefore, the likelihoods of measurements of x and y alone are:

$$L_x(x \mid u, v) \propto \exp\left(-\frac{1}{2}\left(\frac{x(u,v) - x_{\text{obs}}}{\sigma_x}\right)^2\right)$$

$$L_y(y \mid u, v) \propto \exp\left(-\frac{1}{2}\left(\frac{y(u,v) - y_{\text{obs}}}{\sigma_y}\right)^2\right)$$

The uv-model (my invention) (example_uv.py)

Let $\boldsymbol{\theta} = (u, v)$ be the model parameters,
such there are two observables: $\boldsymbol{X} = (x, y)$.

Therefore, the observables x and y depend on the model:
$x = x(u, v)$
$y = y(u, v)$.

Suppose each measurement has a gaussian error.
Therefore, the likelihoods of measurements of x and y alone are:

$$L_x(x \mid u, v) \propto \exp\left(-\frac{1}{2}\left(\frac{x(u,v) - x_{\mathrm{obs}}}{\sigma_x}\right)^2\right)$$

$$L_y(y \mid u, v) \propto \exp\left(-\frac{1}{2}\left(\frac{y(u,v) - y_{\mathrm{obs}}}{\sigma_y}\right)^2\right)$$

My invention:
$x(u, v) = \sqrt{u^2 + v^2}$
$y(u, v) = \frac{2}{\pi} \arctan\left(\frac{v}{u}\right)$

Estimating Be disk parameters from light curves (my PhD thesis)

The observable is the theoretical light curve:

$$M_X(\boldsymbol{\theta}) = M_{X_*}(\boldsymbol{\theta}) + \Delta X(\boldsymbol{\theta}) \,,$$

where the likelihood of each measurement is:

$$L\left(\boldsymbol{X} \mid \boldsymbol{\theta}\right) \propto e^{-\frac{1}{2}\chi^2} \,,$$

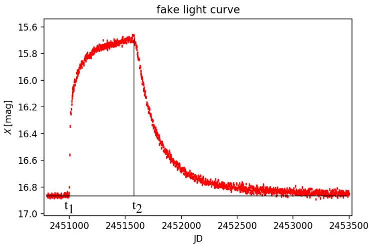
where

$$\chi^2 = \chi^2_{\rm diskless} + \chi^2_{\rm bump} \,,$$

$$\chi^2_{\rm diskless} = \sum_{\rm bands} \frac{(M_{X*}(\boldsymbol{\theta}) - M_{X*}^{\rm obs})^2}{\sigma^2(M_{X*}^{\rm obs})} \,,$$

$$\chi^2_{\rm bump} = \sum_{\rm bands} \sum_{\rm bumps} \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{(\Delta X_i(\boldsymbol{\theta}) - \Delta X_i^{\rm obs})^2}{\sigma^2(\Delta X_i^{\rm obs})} \,,$$

where N_t is the number of datapoints for a given bump at a given photometric band.



The prior knowledge of the distribution of Be stars:

$$\pi\left(\boldsymbol{\theta}\right) \propto M^{-2.3} f_{\rm Be}(M) e^{-\frac{(W - \langle W \rangle)^2}{2\sigma_W^2}}$$

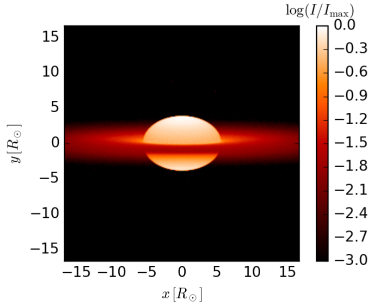
Radiative transfer models of bumps

Table: Representative grid of photometric bumps

Star	i [deg]	Σ_0 [g cm^{-2}]	$\tilde{\tau}_{\rm bu}$
Star 1	00.0	0.30	00.15
Star 2	21.8	0.41	00.45
Star 3	31.0	0.56	00.75
	38.2	0.75	01.50
	44.4	1.01	02.25
	50.0	1.37	03.00
	55.2	1.85	04.50
	60.0	2.50	06.00
	64.6		09.00
	69.1		15.00
	73.4		30.00
	77.6		
	81.8		
	85.9		
	90.0		

Star	Z	$M\,[M_\odot]$	W	$t/t_{\rm MS}$	$\alpha\tau$ [d]
Star 1	0.002	7	0.81	0.5	90.4
Star 2	0.002	11	0.81	0.5	103.3
Star 3	0.002	15	0.81	0.5	118.9

- MCRT code `HDUST`;
- Geneva tracks[a]: $Z, M, t/t_{\rm MS}, W \rightarrow R_{\rm eq}, T_{\rm eff}$, etc;



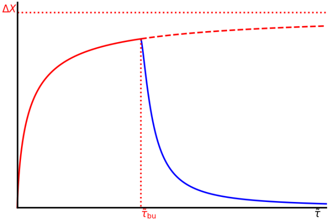
[a] Georgy et al. 2013A&A...553A..24G

From the analysis of the models, I found the following:

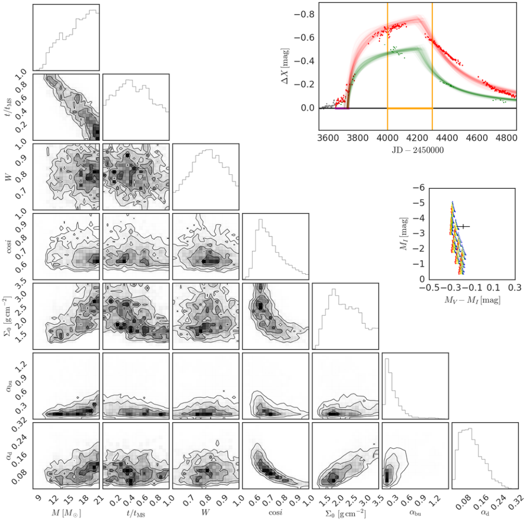
$$\Delta X_{\rm bu} = \Delta X_{\rm bu}^{\infty} \left[1 - \frac{1}{1 + (\xi_{\rm bu}\tilde{\tau})^{\eta_{\rm bu}}}\right]$$

$$\Delta X_{\rm d} = \Delta X_{\rm d}^{0} \left[\frac{1}{1 + (\xi_{\rm d}\,(\tilde{\tau} - \tilde{\tau}_{\rm bu}))^{\eta_{\rm d}}}\right]$$

$$\tilde{\tau} = \begin{cases} \alpha_{\rm bu}\frac{t-t_1}{\alpha\tau}\,, & t_1 \leq t < t_2 \\ \alpha_{\rm bu}\frac{t_2-t_1}{\alpha\tau} + \alpha_{\rm d}\frac{t-t_2}{\alpha\tau}\,, & t \geq t_2 \end{cases}\,,$$



- $\Delta X_{\rm bu}^{\infty} = \Delta X_{\rm bu}^{\infty}(\mathrm{Star}, \cos i, \Sigma_0)$
- $\xi_{\rm bu} = \xi_{\rm bu}(\mathrm{Star}, \cos i, \Sigma_0)$
- $\xi_{\rm d} = \xi_{\rm d}(\mathrm{Star}, \cos i, \Sigma_0, \tilde{\tau}_{\rm bu})$
- $\alpha\tau = \alpha\tau(\mathrm{Star})$



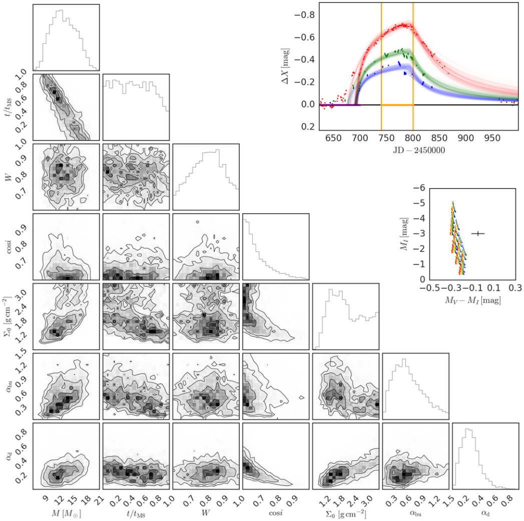
Example[a]

One bump of SMC_SC1 75701.

- $\Sigma_0 = 1.9^{+0.8}_{-0.4}\,\mathrm{g\,cm^{-2}}$
- $\alpha_{\mathrm{bu}} = 0.24^{+0.18}_{-0.08}$
- $\alpha_{\mathrm{d}} = 0.11^{+0.08}_{-0.05}$

- $-\dot{M}_{\mathrm{typ}} = 1.15^{+1.06}_{-0.52} \times 10^{-9}\, M_\odot\,\mathrm{yr}^{-1}$
- $-\dot{J}_{\mathrm{std}} = 2.47^{+2.62}_{-1.24} \times 10^{36}\,\mathrm{g\,cm^2\,s^{-2}}$
- $-\Delta J = 1.03^{+1.08}_{-0.52} \times 10^{44}\,\mathrm{g\,cm^2\,s^{-1}}$

[a] Rímulo et al. 2018MNRAS.476.3555R



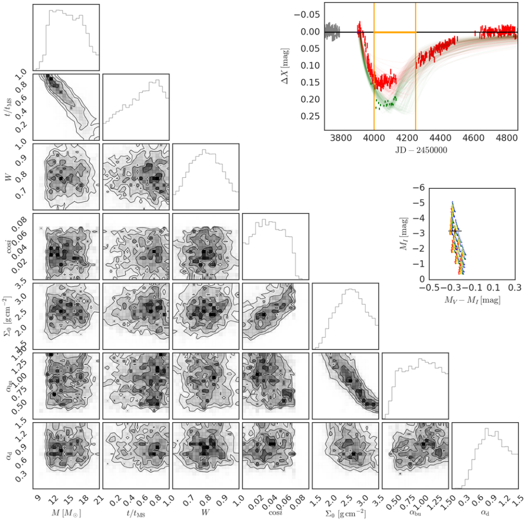
Example[a]:

One bump of SMC_SC5 65500.

- $\Sigma_0 = 1.7^{+0.6}_{-0.4}\,\mathrm{g\,cm^{-2}}$
- $\alpha_{\rm bu} = 0.7^{+0.38}_{-0.34}$
- $\alpha_{\rm d} = 0.26^{+0.13}_{-0.1}$

- $-\dot{M}_{\rm typ} = 2.07^{+1.66}_{-1.02} \times 10^{-9}\,M_\odot\,\mathrm{yr}^{-1}$
- $-\dot{J}_{\rm std} = 3.70^{+3.62}_{-2.05} \times 10^{36}\,\mathrm{g\,cm^2\,s^{-2}}$
- $-\Delta J = 0.32^{+0.31}_{-0.17} \times 10^{44}\,\mathrm{g\,cm^2\,s^{-1}}$

[a] Rímulo et al. 2018MNRAS.476.3555R



Example[a]:

One bump of SMC_SC1 92262.

- $\Sigma_0 = 2.4^{+0.4}_{-0.5}\,\mathrm{g\,cm^{-2}}$
- $\alpha_{\rm bu} = 0.99^{+0.26}_{-0.23}$
- $\alpha_{\rm d} = 0.94^{+0.35}_{-0.3}$

- $-\dot{M}_{\rm typ} = 4.42^{+2.72}_{-1.42} \times 10^{-9}\,M_\odot\,\mathrm{yr}^{-1}$
- $-\dot{J}_{\rm std} = 8.36^{+6.38}_{-3.24} \times 10^{36}\,\mathrm{g\,cm^2\,s^{-2}}$
- $-\Delta J = 1.51^{+1.01}_{-0.62} \times 10^{44}\,\mathrm{g\,cm^2\,s^{-1}}$

[a] Rímulo et al. 2018MNRAS.476.3555R

Membership probability in star clusters ("Bayesian Model Comparison")

Membership probability in star clusters

- Ω: The space of all possible configurations of stars.

Membership probability in star clusters

- Ω: The space of all possible configurations of stars.
- All possible stars divide in 2 groups:
  stars in the cluster ($\mathcal{H}_0$) and
  stars in the field ($\mathcal{H}_1$).
  Therefore, $\Omega = \mathcal{H}_0 \cup \mathcal{H}_1$.

Membership probability in star clusters

- Ω: The space of all possible configurations of stars.
- All possible stars divide in 2 groups:
  stars in the cluster ($\mathcal{H}_0$) and
  stars in the field ($\mathcal{H}_1$).
  Therefore, $\Omega = \mathcal{H}_0 \cup \mathcal{H}_1$.
- Stars may belong to very different models. Let $\mathrm{d}M$ be a very small subset of the whole, such that the stars in it have model parameters $\boldsymbol{\theta}$ in the very small volume $\mathrm{d}V_\theta$.

  Therefore, it follows also that the whole is the union of all those very small sets:
  $\Omega = \bigcup_i \mathrm{d}M_i$.

Membership probability in star clusters

- Ω: The space of all possible configurations of stars.
- All possible stars divide in 2 groups:
  stars in the cluster ($\mathcal{H}_0$) and
  stars in the field ($\mathcal{H}_1$).
  Therefore, $\Omega = \mathcal{H}_0 \cup \mathcal{H}_1$.
- Stars may belong to very different models. Let $\mathrm{d}M$ be a very small subset of the whole, such that the stars in it have model parameters $\boldsymbol{\theta}$ in the very small volume $\mathrm{d}V_\theta$.

  Therefore, it follows also that the whole is the union of all those very small sets:
  $\Omega = \bigcup_i \mathrm{d}M_i$.
- Let us consider a sequence of N stars. Let $\mathrm{d}D_k$ be the subset of sequences of N stars such that the k star of the sequence has data parameters $\boldsymbol{X}$ in the very small volume $\mathrm{d}V_X$. We, therefore, define
  $\mathrm{d}D = \bigcap_{k=1}^{N} \mathrm{d}D_k$.

Membership probability in star clusters

Question: What is the probability of a small subset of cluster+field parameters, given a small subset of data of N stars (from the field or the cluster)? **Answer**: $P(\mathrm{d}M \mid \mathrm{d}D)$.

Membership probability in star clusters

Question: What is the probability of a small subset of cluster+field parameters, given a small subset of data of N stars (from the field or the cluster)? **Answer**: $P(\mathrm{d}M \mid \mathrm{d}D)$.

$$\begin{aligned}
P(\mathrm{d}M \mid \mathrm{d}D) &= \frac{P(\mathrm{d}M \cap \mathrm{d}D)}{P(\mathrm{d}D)} \\
&= \frac{P(\mathrm{d}M \cap \Omega \cap \mathrm{d}D)}{P(\mathrm{d}D)} \\
&= \frac{P(\mathrm{d}M \cap \mathcal{H}_0 \cap \mathrm{d}D) + P(\mathrm{d}M \cap \mathcal{H}_1 \cap \mathrm{d}D)}{P(\mathrm{d}D)}
\end{aligned}$$

Membership probability in star clusters

Question: What is the probability of a small subset of cluster+field parameters, given a small subset of data of N stars (from the field or the cluster)? **Answer**: $P(\mathrm{d}M \mid \mathrm{d}D)$.

$$
\begin{aligned}
P(\mathrm{d}M \mid \mathrm{d}D) &= \frac{P(\mathrm{d}M \cap \mathrm{d}D)}{P(\mathrm{d}D)} \\
&= \frac{P(\mathrm{d}M \cap \Omega \cap \mathrm{d}D)}{P(\mathrm{d}D)} \\
&= \frac{P(\mathrm{d}M \cap \mathcal{H}_0 \cap \mathrm{d}D) + P(\mathrm{d}M \cap \mathcal{H}_1 \cap \mathrm{d}D)}{P(\mathrm{d}D)}
\end{aligned}
$$

$$
\begin{aligned}
P((\mathrm{d}M \cap \mathcal{H}_i) \cap \mathrm{d}D) &= P(\mathrm{d}D \mid (\mathrm{d}M \cap \mathcal{H}_i))P(\mathrm{d}M \cap \mathcal{H}_i) \\
&= P\left(\bigcap_{k=1}^{N} \mathrm{d}D_k \mid (\mathrm{d}M \cap \mathcal{H}_i)\right) P(\mathrm{d}M \cap \mathcal{H}_i) \\
&= \left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid (\mathrm{d}M \cap \mathcal{H}_i)\right)\right) P(\mathrm{d}M \cap \mathcal{H}_i)\,.
\end{aligned}
$$

Membership probability in star clusters

Therefore,

$$P(\mathrm{d}M \mid \mathrm{d}D) = \frac{1}{P(\mathrm{d}D)} \sum_{i=0}^{1} \left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid (\mathrm{d}M \cap \mathcal{H}_i)\right) \right) P(\mathrm{d}M \cap \mathcal{H}_i)$$

Membership probability in star clusters

Therefore,

$$P(\mathrm{d}M \mid \mathrm{d}D) = \frac{1}{P(\mathrm{d}D)} \sum_{i=0}^{1} \left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid (\mathrm{d}M \cap \mathcal{H}_i)\right) \right) P(\mathrm{d}M \cap \mathcal{H}_i)$$

Replacing the probabilities of small subsets by pdfs:

$$p(\boldsymbol{\theta} \mid \boldsymbol{X}) = \frac{1}{g(\boldsymbol{X})} \sum_{i=0}^{1} \left(\prod_{k=1}^{N} \psi_i\left(\boldsymbol{X}_k \mid \boldsymbol{\theta}_i\right) \right) \pi_i(\boldsymbol{\theta}_i)$$

Membership probability in star clusters

Therefore,

$$P(\mathrm{d}M \mid \mathrm{d}D) = \frac{1}{P(\mathrm{d}D)} \sum_{i=0}^{1} \left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid (\mathrm{d}M \cap \mathcal{H}_i)\right) \right) P(\mathrm{d}M \cap \mathcal{H}_i)$$

Replacing the probabilities of small subsets by pdfs:

$$p(\boldsymbol{\theta} \mid \boldsymbol{X}) = \frac{1}{g(\boldsymbol{X})} \sum_{i=0}^{1} \left(\prod_{k=1}^{N} \psi_i\left(\boldsymbol{X}_k \mid \boldsymbol{\theta}_i\right) \right) \pi_i(\boldsymbol{\theta}_i)$$

From the theory:

$$\psi_i\left(\boldsymbol{\mu}, \boldsymbol{r}, M \mid \boldsymbol{\theta}_i\right) = \xi_i(M \mid \boldsymbol{\theta}_i) f_i(\boldsymbol{r} \mid \boldsymbol{\theta}_i) \frac{\exp\left(-\frac{1}{2(1-\rho_i^2)} E\right)}{2\pi\sigma_{x_i}\sigma_{y_i}(1-\rho_i^2)^{\frac{1}{2}}}$$

where

$$E = \frac{(\mu_x - \mu_{x_i})^2}{\sigma_{x_i}^2} + \frac{(\mu_y - \mu_{y_i})^2}{\sigma_{y_i}^2} - \frac{2\rho_i(\mu_x - \mu_{x_i})(\mu_y - \mu_{y_i})}{\sigma_{x_i}\sigma_{y_i}}$$

`example_steven.py`: Part 1: Generating fakes cluster and field:

$$p(\boldsymbol{\theta} \mid \boldsymbol{X}) \propto \left(\prod_{k=1}^{N} \psi_0\left(\boldsymbol{X}_k \mid \boldsymbol{\theta}_0\right)\right) \pi_0(\boldsymbol{\theta}_0) + \left(\prod_{k=1}^{N} \psi_1\left(\boldsymbol{X}_k \mid \boldsymbol{\theta}_1\right)\right) \pi_1(\boldsymbol{\theta}_1)$$

From the theory:

$$\psi_i\left(\boldsymbol{\mu}, \boldsymbol{r}, M \mid \boldsymbol{\theta}_i\right) = \xi_i(M \mid \boldsymbol{\theta}_i) f_i(\boldsymbol{r} \mid \boldsymbol{\theta}_i) \frac{\exp\left(-\frac{1}{2(1-\rho_i^2)} E\right)}{2\pi\sigma_{x_i}\sigma_{y_i}(1-\rho_i^2)^{\frac{1}{2}}}$$

where

$$E = \frac{(\mu_x - \mu_{x_i})^2}{\sigma_{x_i}^2} + \frac{(\mu_y - \mu_{y_i})^2}{\sigma_{y_i}^2} - \frac{2\rho_i(\mu_x - \mu_{x_i})(\mu_y - \mu_{y_i})}{\sigma_{x_i}\sigma_{y_i}}$$

and

- $f_0(\boldsymbol{r} \mid \boldsymbol{\theta}_0)$ is bivariate gaussian too. (My invention!)
- $\xi_0(M \mid \boldsymbol{\theta}_0) \propto M^{-2.3}$ (Well known IMF)
- $f_1(\boldsymbol{r} \mid \boldsymbol{\theta}_1) = \text{constant}$
- $\xi_1(M \mid \boldsymbol{\theta}_1) \propto M^{-5.3}$ (because $t_{\mathrm{MS}} \propto M^{-3}$)

`example_steven.py`: **Part 2: Fitting pdf to stars from the cluster+field:**

$$p(\boldsymbol{\theta} \mid \boldsymbol{X}) \propto \left(\prod_{k=1}^{N} \psi_0\left(\boldsymbol{X}_k \mid \boldsymbol{\theta}_0\right)\right) \pi_0(\boldsymbol{\theta}_0) + \left(\prod_{k=1}^{N} \psi_1\left(\boldsymbol{X}_k \mid \boldsymbol{\theta}_1\right)\right) \pi_1(\boldsymbol{\theta}_1)$$

From the theory:

$$\psi_i\left(\boldsymbol{\mu}, \boldsymbol{r}, M \mid \boldsymbol{\theta}_i\right) = n_i \frac{\exp\left(-\frac{1}{2(1-\rho_i^2)} E\right)}{2\pi\sigma_{x_i}\sigma_{y_i}(1-\rho_i^2)^{\frac{1}{2}}}$$

where

$$E = \frac{(\mu_x - \mu_{x_i})^2}{\sigma_{x_i}^2} + \frac{(\mu_y - \mu_{y_i})^2}{\sigma_{y_i}^2} - \frac{2\rho_i(\mu_x - \mu_{x_i})(\mu_y - \mu_{y_i})}{\sigma_{x_i}\sigma_{y_i}}$$

Membership probability in star clusters

Question: What is the probability of a new star (star $N+1$) belong to the cluster? **Answer**: $P\left(\mathcal{H}_i \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right)$.

Membership probability in star clusters

Question: What is the probability of a new star (star $N+1$) belong to the cluster? **Answer**: $P\left(\mathcal{H}_i \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right)$.

$$
\begin{aligned}
P\left(\mathcal{H}_0 \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right) &= P\left(\Omega \cap \mathcal{H}_0 \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right) \\
&= P\left(\bigcup_j \mathrm{d}M_j \cap \mathcal{H}_0 \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right) \\
&= \sum_j P\left(\mathrm{d}M_j \cap \mathcal{H}_0 \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right) \\
&= \sum_j \frac{P\left(\mathrm{d}M_j \cap \mathcal{H}_0 \cap \mathrm{d}D \cap \mathrm{d}D_{N+1}\right)}{P(\mathrm{d}D \cap \mathrm{d}D_{N+1})} \\
&= \sum_j \frac{P\left(\mathrm{d}D \cap \mathrm{d}D_{N+1} \mid \mathrm{d}M_j \cap \mathcal{H}_0\right) P(\mathrm{d}M_j \cap \mathcal{H}_0)}{P(\mathrm{d}D \cap \mathrm{d}D_{N+1})}
\end{aligned}
$$

Therefore,

$$
P\left(\mathcal{H}_0 \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right) = \sum_j \frac{\left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid \mathrm{d}M_j \cap \mathcal{H}_0\right)\right) P\left(\mathrm{d}D_{N+1} \mid \mathrm{d}M_j \cap \mathcal{H}_0\right) P(\mathrm{d}M_j \cap \mathcal{H}_0)}{P(\mathrm{d}D \cap \mathrm{d}D_{N+1})}
$$

example_steven.py: Part 3: Membership probability of star clusters:

From the Bayes' theorem, the above equation becomes:

$$P\left(\mathcal{H}_0 \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right) = \frac{\sum_j \left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid \mathrm{d}M_j \cap \mathcal{H}_0\right)\right) P\left(\mathrm{d}D_{N+1} \mid \mathrm{d}M_j \cap \mathcal{H}_0\right) P(\mathrm{d}M_j \cap \mathcal{H}_0)}{\sum_{i=0}^{1} \sum_j \left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid \mathrm{d}M_j \cap \mathcal{H}_i\right)\right) P\left(\mathrm{d}D_{N+1} \mid \mathrm{d}M_j \cap \mathcal{H}_i\right) P(\mathrm{d}M_j \cap \mathcal{H}_i}$$

example_steven.py: Part 3: Membership probability of star clusters:

From the Bayes' theorem, the above equation becomes:

$$P\left(\mathcal{H}_0 \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right) = \frac{\sum_j \left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid \mathrm{d}M_j \cap \mathcal{H}_0\right)\right) P\left(\mathrm{d}D_{N+1} \mid \mathrm{d}M_j \cap \mathcal{H}_0\right) P(\mathrm{d}M_j \cap \mathcal{H}_0)}{\sum_{i=0}^{1}\sum_j \left(\prod_{k=1}^{N} P\left(\mathrm{d}D_k \mid \mathrm{d}M_j \cap \mathcal{H}_i\right)\right) P\left(\mathrm{d}D_{N+1} \mid \mathrm{d}M_j \cap \mathcal{H}_i\right) P(\mathrm{d}M_j \cap \mathcal{H}_i)}$$

Replacing the Probabilities of small subsets by pdfs:

$$\begin{aligned} P\left(\mathcal{H}_0 \mid \mathrm{d}D \cap \mathrm{d}D_{N+1}\right) &= \frac{\int \left(\prod_{k=1}^{N} \psi_0\left(\boldsymbol{X}_k \mid \boldsymbol{\theta}_0\right)\pi_0(\boldsymbol{\theta}_0)\right)\psi_0\left(\boldsymbol{X}_{N+1} \mid \boldsymbol{\theta}_0\right)\mathrm{d}V_{\boldsymbol{\theta}_0}}{\sum_{i=0}^{1}\int \left(\prod_{k=1}^{N} \psi_i\left(\boldsymbol{X}_k \mid \boldsymbol{\theta}_i\right)\pi_i(\boldsymbol{\theta}_i)\right)\psi_i\left(\boldsymbol{X}_{N+1} \mid \boldsymbol{\theta}_i\right)\mathrm{d}V_{\boldsymbol{\theta}_i}} \\ &\approx \frac{\psi_0\left(\boldsymbol{X}_{N+1} \mid \hat{\boldsymbol{\theta}}_0\right)}{\psi_0\left(\boldsymbol{X}_{N+1} \mid \hat{\boldsymbol{\theta}}_0\right) + \psi_1\left(\boldsymbol{X}_{N+1} \mid \hat{\boldsymbol{\theta}}_1\right)} \end{aligned}$$

Thank you!