Approximate Bayesian Computation for the Stellar Initial Mass Function

Jessi Cisewski
Department of Statistics
Yale University

SCMA6

Collaborators: Grant Weller (Savvysherpa), Chad Schafer (Carnegie Mellon), David Hogg (NYU)

Background: ABC

The posterior for θ given observed data x_{obs}:

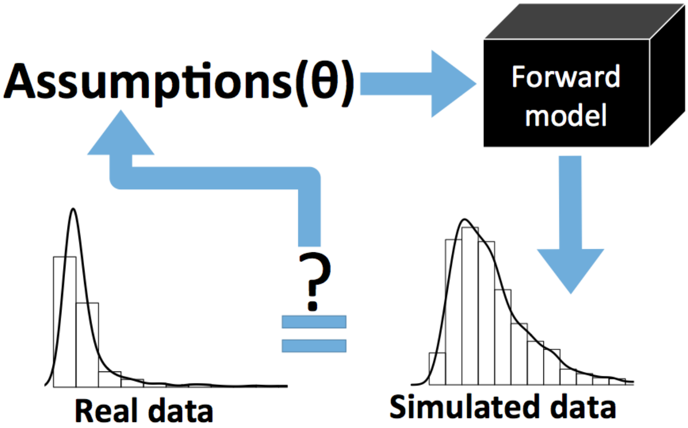
$$\pi(\theta \mid x_{\text{obs}}) = \frac{f(x_{\text{obs}} \mid \theta)\pi(\theta)}{\int f(x_{\text{obs}} \mid \theta)\pi(\theta)d\theta} = \frac{f(x_{\text{obs}} \mid \theta)\pi(\theta)}{f(x_{\text{obs}})}$$

Approximate Bayesian Computation

- "Likelihood-free" approach to approximating $\pi(\theta \mid x_{\text{obs}})$ ($f(x_{\text{obs}} \mid \theta)$ not specified)
- Proceeds via simulation of the forward process

Why would we not know $f(x_{\text{obs}} \mid \theta)$?

1. Physical model too complex to write as a closed form function of θ
2. Strong dependency in data
3. Observational limitations


Assumptions(θ)
Forward model
?
=
Real data
Simulated data

ABC for Astronomy

- cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation (Ishida et al., 2015)
- Approximate Bayesian Computation for Forward Modeling in Cosmology (Akeret et al., 2015)
- Likelihood-Free Cosmological Inference with Type Ia Supernovae: Approximate Bayesian Computation for a Complete Treatment of Uncertainty (Weyant et al., 2013)
- Likelihood - free inference in cosmology: potential for the estimation of luminosity functions (Schafer and Freeman, 2012)
- Approximate Bayesian Computation for Astronomical Model Analysis: A case study in galaxy demographics and morphological transformation at high redshift (Cameron and Pettitt, 2012)

Stellar Initial Mass Function: the distribution of star masses after a star formation event within a specified volume of space

Molecular cloud → Protostars → Stars

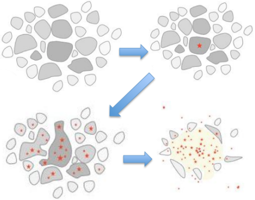


Image: adapted from

http://www.astro.ljmu.ac.uk

- Properties of the cluster formation could include, for example, core growth by accretion, interaction due to turbulent environment (Bate, 2012)
- Observation effects include aging, completeness, measurement error

Basic ABC algorithm

For the observed data x_{obs} and prior $\pi(\theta)$:

Algorithm*

1. Sample θ_{prop} from prior $\pi(\theta)$
2. Generate x_{prop} from forward process $f(x \mid \theta_{\text{prop}})$
3. Accept θ_{prop} if $x_{\text{obs}} = x_{\text{prop}}$
4. Return to step 1

*Introduced in Pritchard et al. (1999) (population genetics)

Step 3: Accept θ_{prop} if $x_{\text{obs}} = x_{\text{prop}}$

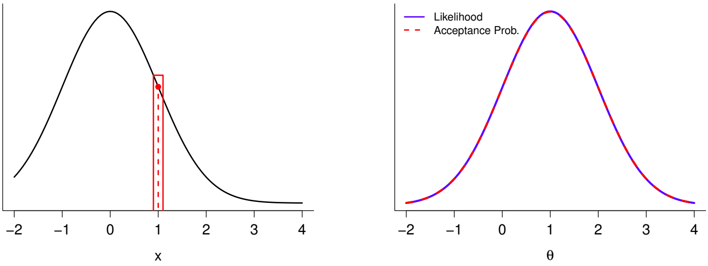
- Waiting for proposals such that $x_{\text{obs}} = x_{\text{prop}}$ would be computationally prohibitive
- Instead, accept proposals with $\Delta(x_{\text{obs}}, x_{\text{prop}}) \leq \epsilon$
  for some distance Δ and some tolerance threshold ϵ

An accepted θ is a draw from the posterior if

$$P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta) \propto f(x_{\text{obs}}|\theta) \quad \text{(the likelihood)}$$

Toy Example: Assume that x_{obs} is Gaussian N(θ,1).

Suppose $x_{\text{obs}} = 1$, $\epsilon = 0.1$.

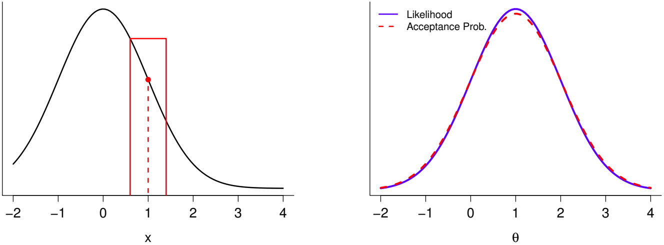


(left) Propose $\theta_{prop} = 0$; acceptance region for x_{prop} in red rectangle

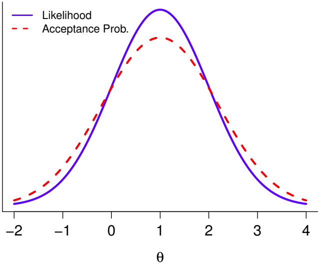
(right) Consider all possible θ and calculate acceptance probability

Illustration from Chad Schafer

$$x_{\text{obs}} = 1,\ \theta = 0,\ \epsilon = 0.4.$$



$x_{\text{obs}} = 1, \epsilon = 1 \rightarrow$



ABC Background

Comparing x_{prop} with x_{obs} is not feasible.

ABC Background

Comparing x_{prop} with x_{obs} is not feasible.

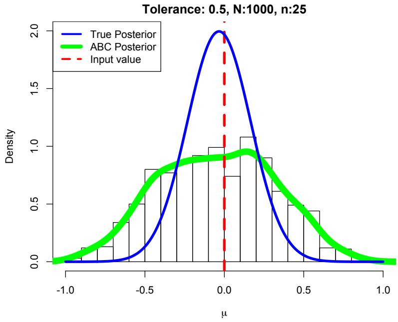
When x is high-dimensional, will have to make ϵ too large in order to keep acceptance probability reasonable.

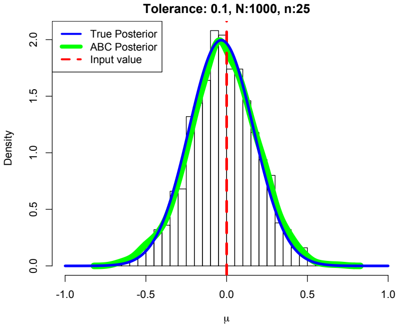
Instead, reduce the dimension by comparing summaries, $S(x_{\text{prop}})$ and $S(x_{\text{obs}})$.

Gaussian illustration

- Data x_{obs} consists of 25 iid draws from Normal$(\mu, 1)$
- Summary statistics $S(x) = \bar{x}$
- Distance function $\Delta(S(x_{\text{prop}}), S(x_{\text{obs}})) = |\bar{x}_{\text{prop}} - \bar{x}_{\text{obs}}|$
- Tolerance $\epsilon = 0.50$ and 0.10
- Prior $\pi(\mu) =$ Normal(0,10)

Gaussian illustration: posteriors for μ





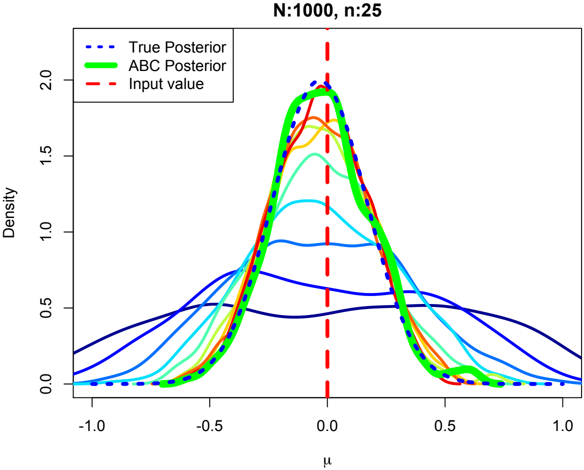
- How to pick a tolerance, ϵ?

  Instead of starting the ABC algorithm over with a smaller tolerance (ϵ), decrease the tolerance and use the already sampled particle system as a proposal distribution *rather* than drawing from the prior distribution.

  Particle system: (1) retained sampled values, (2) importance weights

Beaumont et al. (2009); Moral et al. (2011); Bonassi and West (2004)

Gaussian illustration: sequential posteriors



Tolerance sequence, $\epsilon_{1:10}$:

1.00 0.75 0.53 0.38 0.27 0.19 0.15 0.11 0.08 0.06

We propose an ABC algorithm for inference on the stellar IMF

Examples of IMF models

- **Power-law: Salpeter (1955)**
  - Used a power law with $\alpha = 2.35$
- **Broken power-law: Kroupa (2001)**

$$\Phi(M) \propto M^{-\alpha_i}, M_{1i} \leq M \leq M_{2i}$$

$$\begin{array}{ll} \alpha_1 = 0.3 & \text{for } 0.01 \leq M/M^*_{Sun} \leq 0.08 \text{ [Sub-stellar]} \\ \alpha_2 = 1.3 & \text{for } 0.08 \leq M/M_{Sun} \leq 0.50 \\ \alpha_3 = 2.3 & \text{for } 0.50 \leq M/M_{Sun} \leq M_{\max} \end{array}$$

- **Log-Normal model: Chabrier (2003)**

$$\xi(\log m) = \frac{dn}{d\log m} = 0.158 \times \exp\left(-\frac{(\log m - \log 0.08)^2}{2(0.69)^2}\right)$$

*1 M_{Sun} = 1 Solar Mass (the mass of our Sun)

ABC for the stellar initial mass function

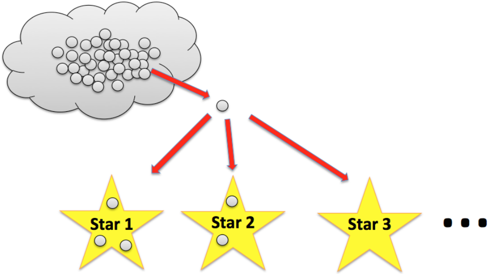
- We propose an ABC algorithm with a new data-generating model
  - Can account for various observational limitations and uncertainties
  - Goal is to capture information about the cluster formation process
- Use ideas of preferential attachment ["Rich get richer" (Yule, 1925; Simon, 1955)]

  Applications: wealth distribution, evolution of citation networks, number of internet page links

  Often considered underlying mechanism for data exhibiting power-law behavior (D'Souza et al., 2007)

Proposed Model Generation Process

1. Fix cloud mass that forms stars: M_{ecl}
2. $m_t \sim Exponential(\lambda)$ enters the system of stars, $t = 1, 2, \ldots$
3. The mass quantity m_t does one of the following:
    - Starts a new star
    - Joins an existing star
4. Process is complete when the total mass reaches M_{ecl}



Proposed Model Form

A mass, m_t, entering the system will

- Form a new star with probability

$$\pi_t = \min(1, \alpha)$$

- Join existing star k, $k = 1, \ldots, n_t$, with probability

$$\pi_{kt} \propto M_{k,t}^{\gamma}$$

α = probability of entering mass forming a new star

γ = growth component ($\gamma = 1$ is linear growth)

The generating process is complete when the total mass of formed stars reaches M_{ecl}. The possible ranges of parameters are $\lambda > 0$, $\alpha \in [0, 1)$, and $\gamma > 0$.

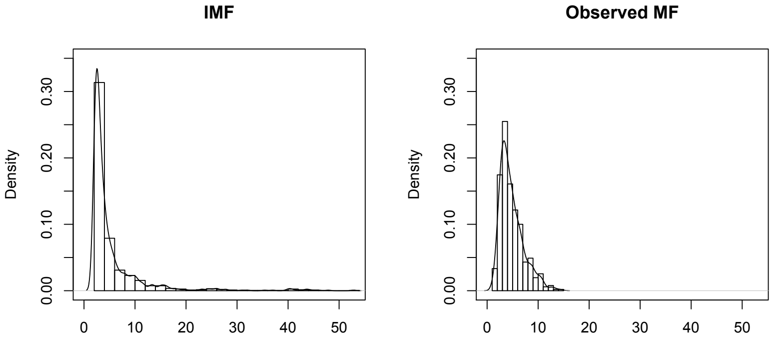
Selecting summary statistics

- Use simplified model: broken power-law, assume independent draws
- Account for the following observation effects:
  - Aging $\longrightarrow$ more massive stars die faster
  - Completeness function:

$$P(\text{observing star} \mid m) = \begin{cases} 0, & m < C_{\text{min}} \\ \frac{m - C_{\text{min}}}{C_{\text{max}} - C_{\text{min}}}, & m \in [C_{\text{min}}, C_{\text{max}}] \\ 1, & m > C_{\text{max}} \end{cases}$$

  - Measurement error

$$L(\alpha \mid m_{1:n_{obs}}, n_{tot}) =$$
$$\left(P(M > T_{age}) + \left(\frac{1-\alpha}{M_{\text{max}}^{1-\alpha} - M_{\text{min}}^{1-\alpha}} \right) \int_{C_{\text{min}}}^{C_{\text{max}}} M^{-\alpha} \times \left(1 - \frac{M - C_{\text{min}}}{C_{\text{max}} - C_{\text{min}}} \right) dM \right)^{n_{tot} - n_{obs}}$$
$$\times \prod_{i=1}^{n_{obs}} \left\{ \int_2^{T_{age}} (2\pi\sigma^2)^{-\frac{1}{2}} m_i^{-1} e^{-\frac{1}{2\sigma^2}(\log(m_i) - \log(M))^2} \left(\frac{1-\alpha}{M_{\text{max}}^{1-\alpha} - M_{\text{min}}^{1-\alpha}} \right) M^{-\alpha} \right.$$
$$\left. \times \left(I\{M > C_{\text{max}}\} + \left(\frac{M - C_{\text{min}}}{C_{\text{max}} - C_{\text{min}}} \right) I\{C_{\text{min}} \leq M \leq C_{\text{max}}\} \right) dM \right\}$$



Sample size $= 1000$ stars, $[C_{\mathrm{min}}, C_{\mathrm{max}}] = [2, 4]$, $\sigma = 0.25$

Summary statistics

We want to account for the following with our summary statistics and distance functions:

1. Shape of the observed Mass Function

$$\rho_1(m_{prop}, m_{obs}) = \left[\int \left\{ \hat{f}_{\log m_{prop}}(x) - \hat{f}_{\log m_{obs}}(x) \right\}^2 dx \right]^{1/2}$$

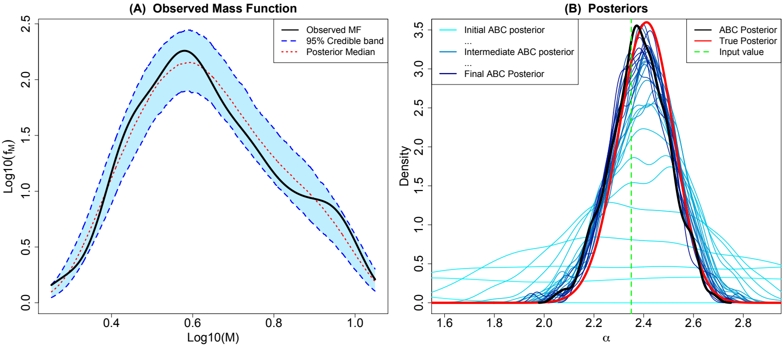
2. Number of stars observed

$$\rho_2(m_{prop}, m_{obs}) = |1 - n_{prop}/n_{obs}|$$

m_{prop} = masses of the stars simulated from the forward model
m_{obs} = masses of observed stars
n_{prop} = number of stars simulated from the forward model
n_{obs} = number of observed stars

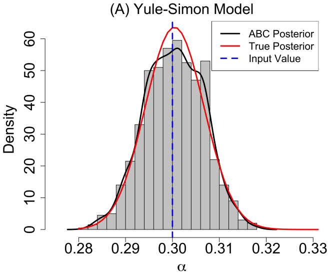
Simulation Study with simpler model

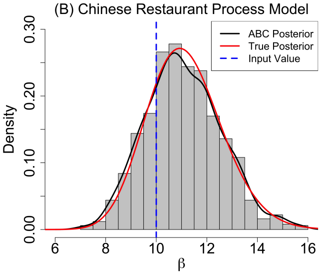
1. Draw $n = 10^3$ stars
2. IMF slope $\alpha = 2.35$ with $M_{min} = 2$ and $M_{max} = 60$
3. $N = 10^3$ particles
4. $T = 30$ sequential time steps



Special cases of proposed forward model

- **Yule-Simon model**: m_t fixed, $\gamma = 1$, $\alpha = 0.30$
- **Chinese restaurant process**: m_t fixed, $\gamma = 1$, $\alpha = 0$, modified π_t (prob of forming a new star)



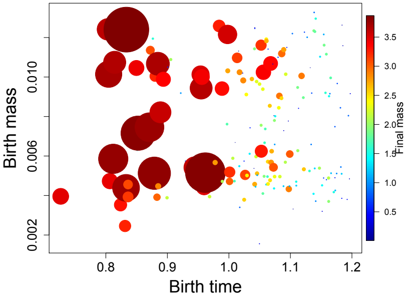


Mon. Not. R. Astron. Soc. **419,** 3115–3146 (2012) doi:10.1111/j.1365-2966

Stellar, brown dwarf and multiple star properties from a radiation hydrodynamical simulation of star cluster formation

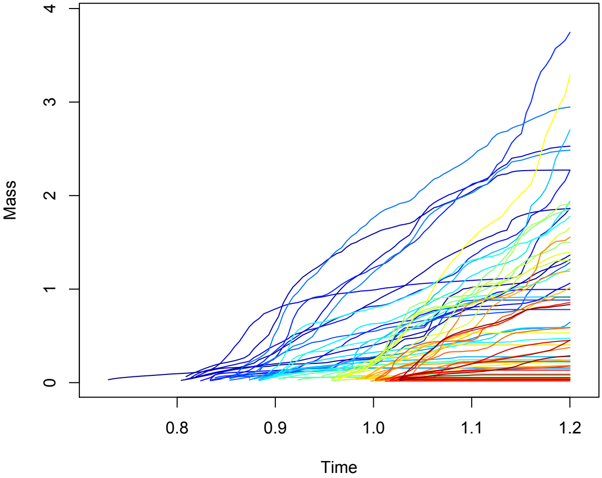
Matthew R. Bate[1,2]★
[1] *School of Physics, University of Exeter, Stocker Road, Exeter EX4 4QL*
[2] *Monash Centre for Astrophysics, School of Mathematical Sciences, Monash University, Clayton, Vic 3168, Australia*

- Summary of 183 stars
- Size and color of points based on final mass



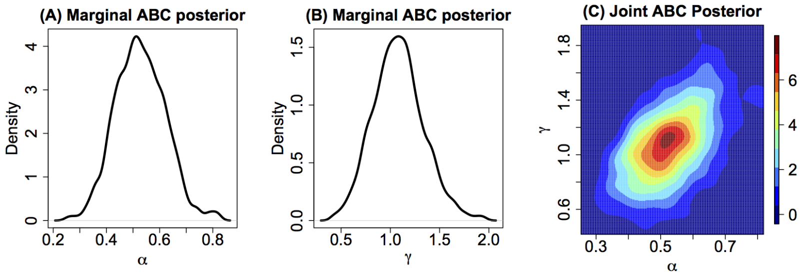
Bate (2012) simulation

- Growth of 183 stars
- Color of lines based on start time
  bluer = formed earlier, redder = formed later



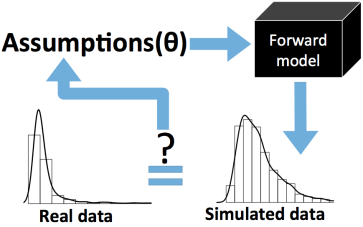
Bate (2012) simulation - ABC posteriors

- 2000 particles, 23 time steps, $M_{tot} = 88.68$



Summary

- ABC can be a useful tool when data are too complex to define a reasonable likelihood
- Selection of good summary statistics is crucial for ABC posterior to be meaningful
- A challenge in describing the Stellar Initial Mass Function is capturing the cluster formation mechanism
- We proposed a Preferential Attachment mechanism for the ABC forward model



THANK YOU!!!

Bibliography I

Akeret, J., Refregier, A., Amara, A., Seehars, S., and Hasner, C. (2015), "Approximate Bayesian Computation for Forward Modeling in Cosmology," *arXiv preprint arXiv:1504.07245*.

Bate, M. R. (2012), "Stellar, brown dwarf and multiple star properties from a radiation hydrodynamical simulation of star cluster formation," *Monthly Notices of the Royal Astronomical Society*, 419, 3115–3146.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009), "Adaptive approximate Bayesian computation," *Biometrika*, 96, 983 – 990.

Bonassi, F. V. and West, M. (2004), "Sequential Monte Carlo with Adaptive Weights for Approximate Bayesian Computation," *Bayesian Analysis*, 1, 1–19.

Cameron, E. and Pettitt, A. N. (2012), "Approximate Bayesian Computation for Astronomical Model Analysis: A Case Study in Galaxy Demographics and Morphological Transformation at High Redshift," *Monthly Notices of the Royal Astronomical Society*, 425, 44–65.

Chabrier, G. (2003), "The galactic disk mass function: Reconciliation of the hubble space telescope and nearby determinations," *The Astrophysical Journal Letters*, 586, L133.

D'Souza, R. M., Borgs, C., Chayes, J. T., Berger, N., and Kleinberg, R. D. (2007), "Emergence of tempered preferential attachment from optimization," *Proceedings of the National Academy of Sciences*, 104, 6112–6117.

Ishida, E., Vitenti, S., Penna-Lima, M., Cisewski, J., de Souza, R., Trindade, A., Cameron, E., et al. (2015), "cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation," *arXiv preprint arXiv:1504.06129*.

Kroupa, P. (2001), "On the variation of the initial mass function," *Monthly Notices of the Royal Astronomical Society*, 322, 231–246.

Moral, P. D., Doucet, A., and Jasra, A. (2011), "An adaptive sequential Monte Carlo method for approximate Bayesian computation," *Statistics and Computing*, 22, 1009–1020.

Pritchard, J. K., Seielstad, M. T., and Perez-Lezaun, A. (1999), "Population Growth of Human Y Chromosomes: A study of Y Chromosome Microsatellites," *Molecular Biology and Evolution*, 16, 1791 – 1798.

Salpeter, E. E. (1955), "The luminosity function and stellar evolution." *The Astrophysical Journal*, 121, 161.

Bibliography II

Schafer, C. M. and Freeman, P. E. (2012), *Statistical Challenges in Modern Astronomy V*, Springer, chap. 1, Lecture Notes in Statistics, pp. 3 – 19.

Simon, H. A. (1955), "On a class of skew distribution functions," *Biometrika*, 425–440.

Weyant, A., Schafer, C., and Wood-Vasey, W. M. (2013), "Likeihood-free cosmological inference with type Ia supernovae: approximate Bayesian computation for a complete treatment of uncertainty," *The Astrophysical Journal*, 764, 116.

Yule, G. U. (1925), "A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS," *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 21–87.