Statistics in Cosmology

Day 4 – Likelihood-free inference

11th TRR33 Winter School in Cosmology
10-16 December 2017, Passo del Tonale - Italy

Emille E. O. Ishida

Laboratoire de Physique de Clermont - Université Clermont-Auvergne
Clermont Ferrand, France

A little cosmology ...

Type Ia supernovae

$$d_L(z;\Omega_m,w) \propto (1+z)\int_0^z \frac{dx}{H(x;\Omega_m,w)}$$

$$\mu = 5\log_{10}\left(d_L(z;\Omega_m,w)\right) + 25$$

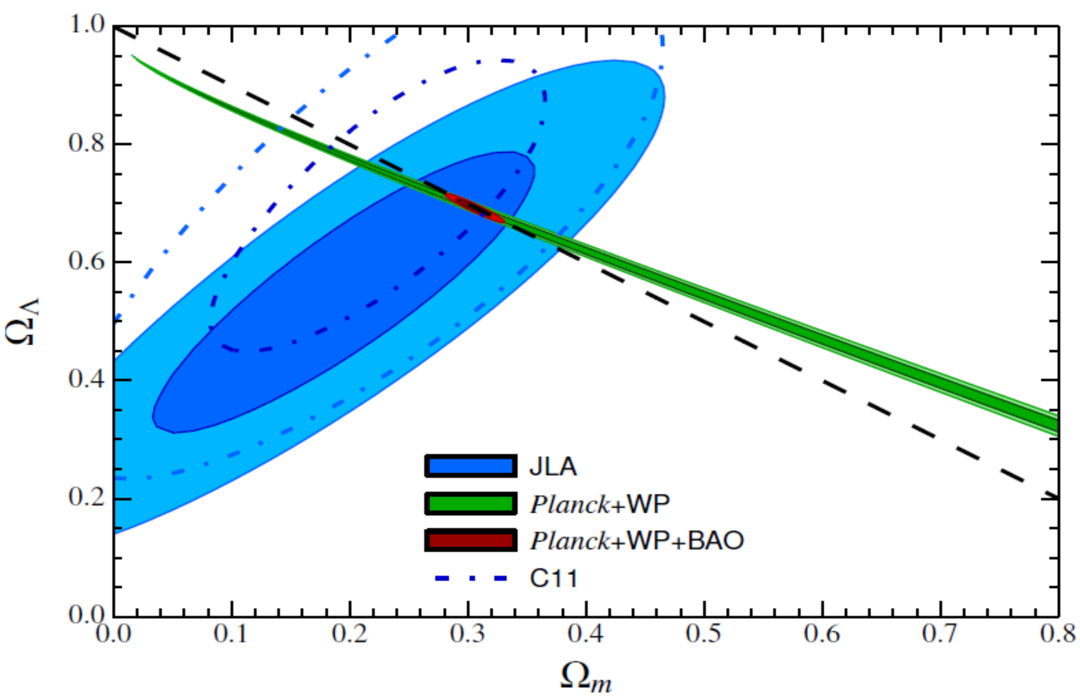
$$\mu = m_B - M + \alpha \times \mathrm{width} - \beta \times \mathrm{colour}$$

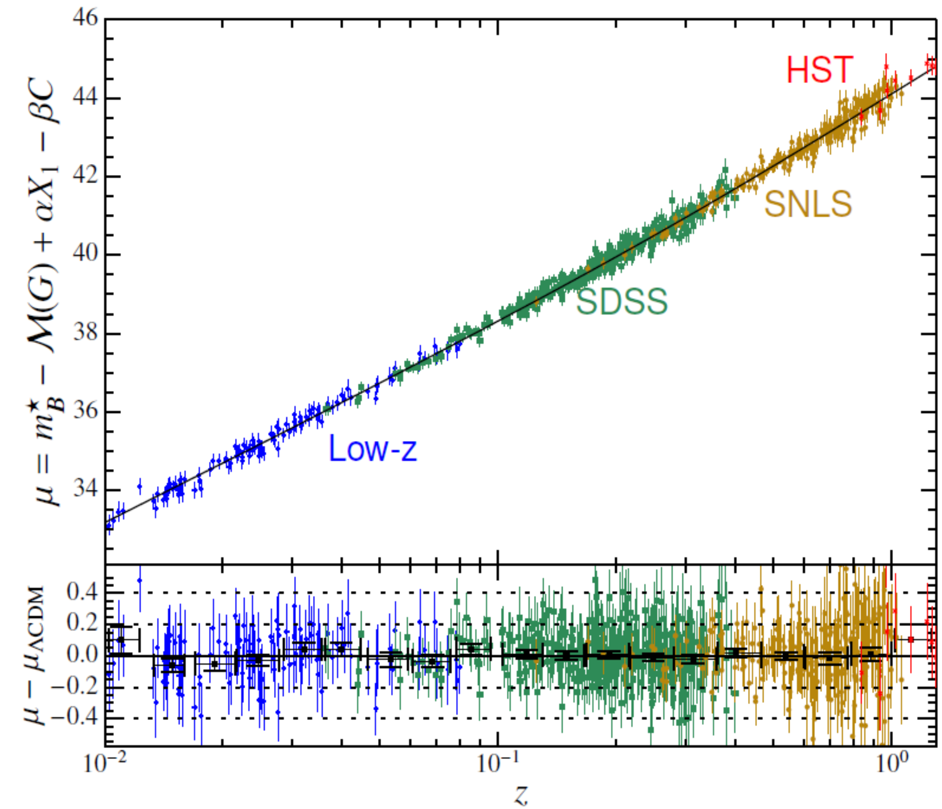
Standard analysis

$$-2\log\mathcal{L} = \chi^2 = \sum_i \frac{\left(\mu(z_i,\mathcal{C}) - [\hat{m}_{B,i} - M + \alpha\hat{x}_{1,i} - \beta\hat{c}_i]\right)^2}{\sigma_{\rm int}^2 + \sigma_{\rm fit}^2}$$

$$\sigma_{\rm fit}^2 = \sigma_{m_B}^2 + \alpha^2\sigma_{x_1}^2 + \beta^2\sigma_c^2 + \text{correlations}$$

	Ω_m	α	β	M_B^1
JLA (stat+sys)	0.295 ± 0.034	0.141 ± 0.006	3.101 ± 0.075	-19.05 ± 0.02



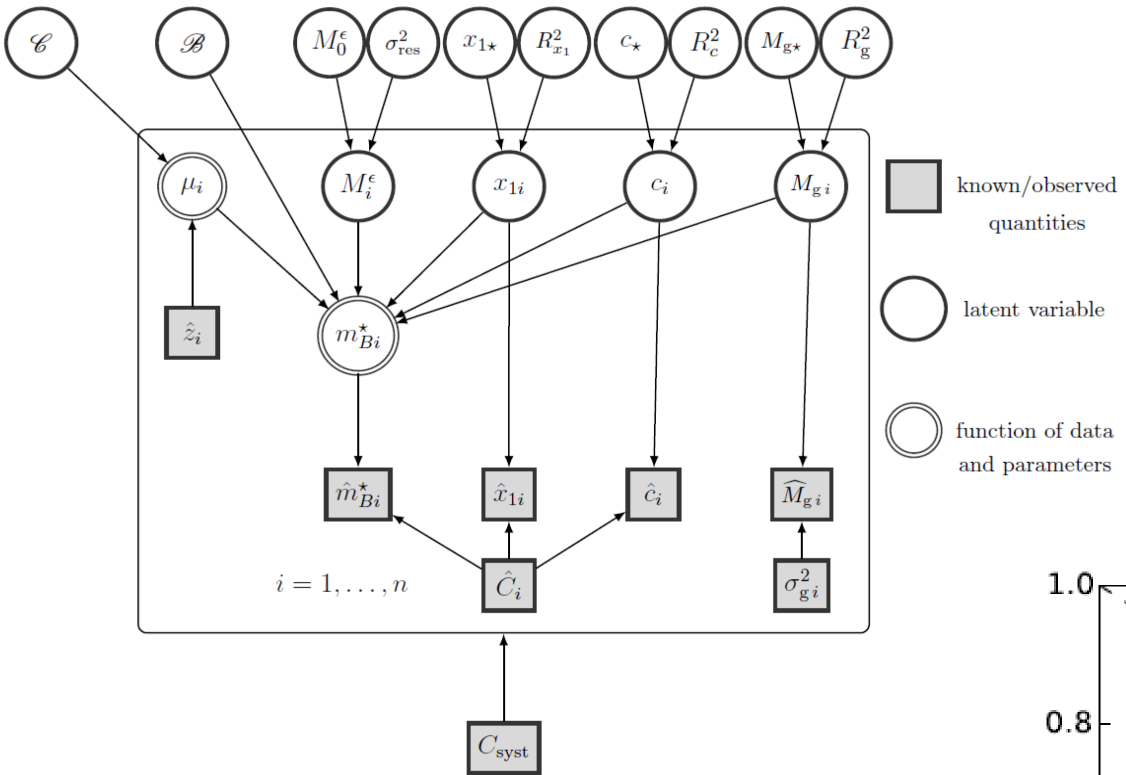


Betoule et al.,2014

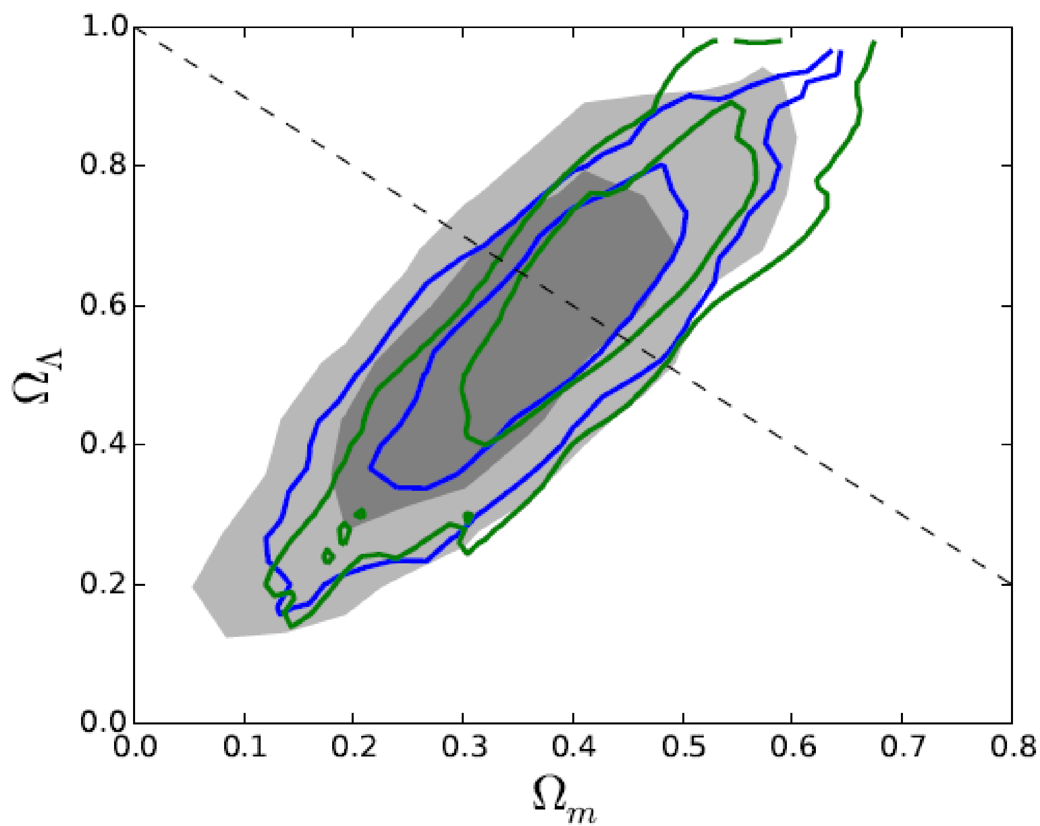
Type Ia supernovae

BAHAMAS

BAyesian HierArchical Modeling for the Analysis of Supernova



$\Omega_{\rm m}$	0.343 ± 0.096
Ω_Λ	0.523 ± 0.144
Ω_κ	0.134 ± 0.232
α	0.141 ± 0.006
β	3.058 ± 0.095



Shariff et al., 2016

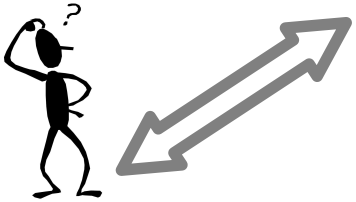
Approximate Bayesian Computation

This is not a replacement for MCMC!

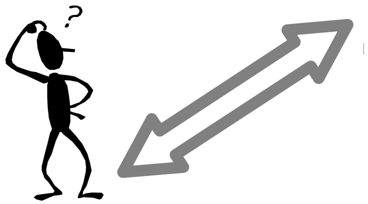
Forward Modeling

$$-2\log\mathcal{L} = \chi^2 = \sum_i \frac{\left(\overbrace{\mu(z_i,\mathcal{C})}^{\text{theory}} - \overbrace{[\hat{m}_{B,i} - M + \alpha\hat{x}_{1,i} - \beta\hat{c}_i]}^{\text{measurement}}\right)^2}{\sigma_{\text{int}}^2 + \sigma_{\text{fit}}^2}$$

Synthetic catalog



Can simulations be used for inference?



Comparing catalogs (or basic ABC)

Model + $\{\Omega_{m1}, w_1\}$

z	ξ
2.0	1.8
11.5	10.5
13.1	12.9
25.4	26.7
..	...

Observed

z	ξ
2.5	0.9
5.6	2.1
7.4	2.9
10.5	4.2
..	...

Simulate

Draw $\{\Omega_{m1}, w_1\}$ from priors

Compare

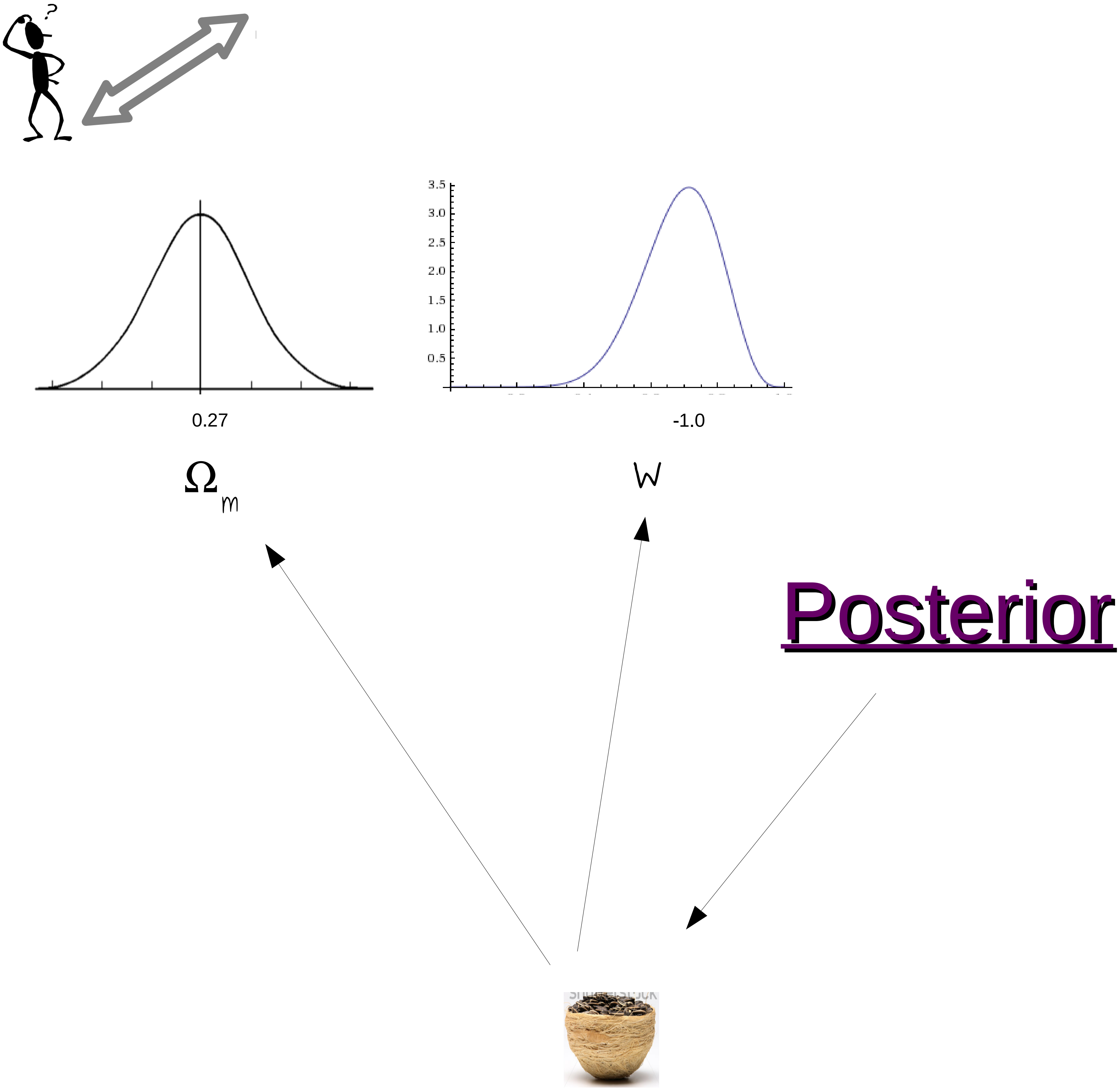
If close enough

If too different

Throw away

Start again

Store $\{\Omega_{m1}, w_1\}$


3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.27
-1.0
Ωm
w
Posterior

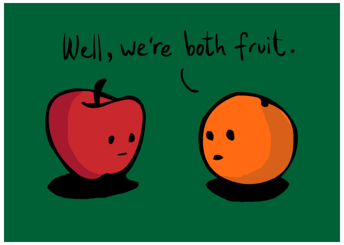
Approximate Bayesian Computation (ABC)

Ingredients

1. Priors

2. Simulator (cheap)

3. Distance function
(or summary statistics)



Sufficient summary statistics

Statistics → measured of an attribute from a sample

Summary statistics → summarizes the information contained in the sample

Sufficient summary statistics → contains all possible information available in the sample

Example of (not sufficient) summary statistics for supernovae:

$$\Delta_{\text{data}} =$$

$$\frac{1}{N_{\text{data}}} \sum_{i}^{N_{\text{data}}} \frac{[\mu(z_i^{data}, \theta^*) - (m_{b,i}^{data} + \alpha^* x_{1,i}^{data} - \beta^* c_i^{data} - M_0 - \delta M_0^*)]^2}{\sigma_{m_{b,i}}^2 + (\alpha^* \sigma_{x_{1,i}})^2 + (\beta^* \sigma_{c_i})^2 + \sigma_{\text{int}}^2}$$

$$\Delta_{\text{sim}} =$$

$$\frac{1}{N_{\text{sim}}} \sum_{j}^{N_{\text{sim}}} \frac{[\mu(z_j^{sim}, \theta^*) - (m_{b,j}^{sim} + \alpha^* x_{1,j}^{sim} - \beta^* c_j^{sim} - M_0 - \delta M_0^*)]^2}{\sigma_{m_{b,j}}^2 + (\alpha^* \sigma_{x_{1,j}})^2 + (\beta^* \sigma_{c_j})^2 + \sigma_{\text{int}}^2},$$

$$\boxed{\rho = \left| \Delta_{\text{data}} - \Delta_{\text{sim}} \right|}$$

Weyant et al., , 2013

Further developments in ABC

Particle System

(parameter values
+ distances)

Simulate

Model + $\{\Omega_{m1}, w_1\}$

z	ξ
2.0	1.8
11.5	10.5
13.1	12.9
25.4	26.7
..	...

Draw $\{\Omega_{m1}, w_1\}$
from priors

Calculate
distance

Store $\{\Omega_{m1}, w_1, d_1\}$

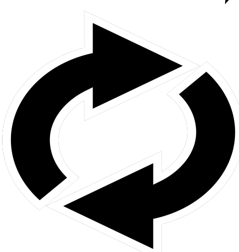
Further developments in ABC

Particle System
(parameter values + distances)

Take the N parameter vectors with smallest distances

Determine distance threshold, ε

Repeat basic ABC algorithm, accept if d < ε



Use the new particle system as a guide to subsequent drawings

Further developments in ABC

Importance sampling: guiding draws

Draw $\{a_1, b_1\}$ from previous weighted particle system

Repeat until there are N particles stored

if d < ε, store

$\{\Omega_{m1}, w_1\}$

Associate a weight with each particle

$$W_t^j = \frac{p(\boldsymbol{\theta}_t^j)}{\sum_{i=1}^{N} W_{t-1}^i \mathcal{N}(\boldsymbol{\theta}_t^j; \boldsymbol{\theta}_{t-1}^i, C_{t-1})},$$

Determine ε

Further developments in ABC

Importance sampling: guiding draws

Draw $\{a_1, b_1\}$ from previous weighted particle system

Repeat until there are N particles stored

if $d < \varepsilon$, store $\{\Omega_{m1}, w_1\}$

Associate a weight with each particle

$$W_t^j = \frac{p(\boldsymbol{\theta}_t^j)}{\sum_{i=1}^N W_{t-1}^i \mathcal{N}(\boldsymbol{\theta}_t^j; \boldsymbol{\theta}_{t-1}^i, C_{t-1})},$$

Determine ε

Example

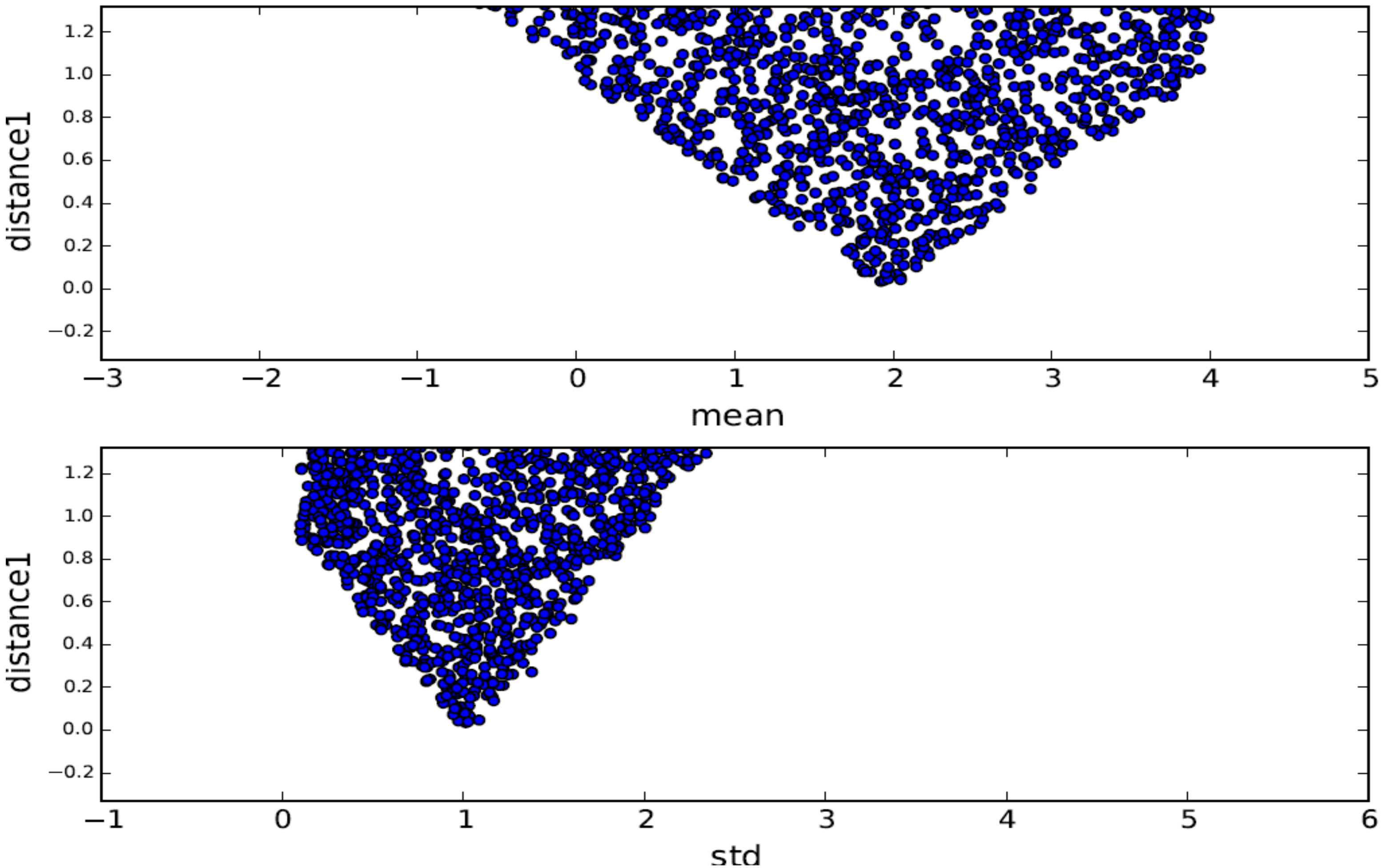
Fiducial model: *Y ~ Gaussian(μ, σ)*

Priors: $-2.0 < \mu < 4.0$ and $0.1 < \sigma < 5.0$

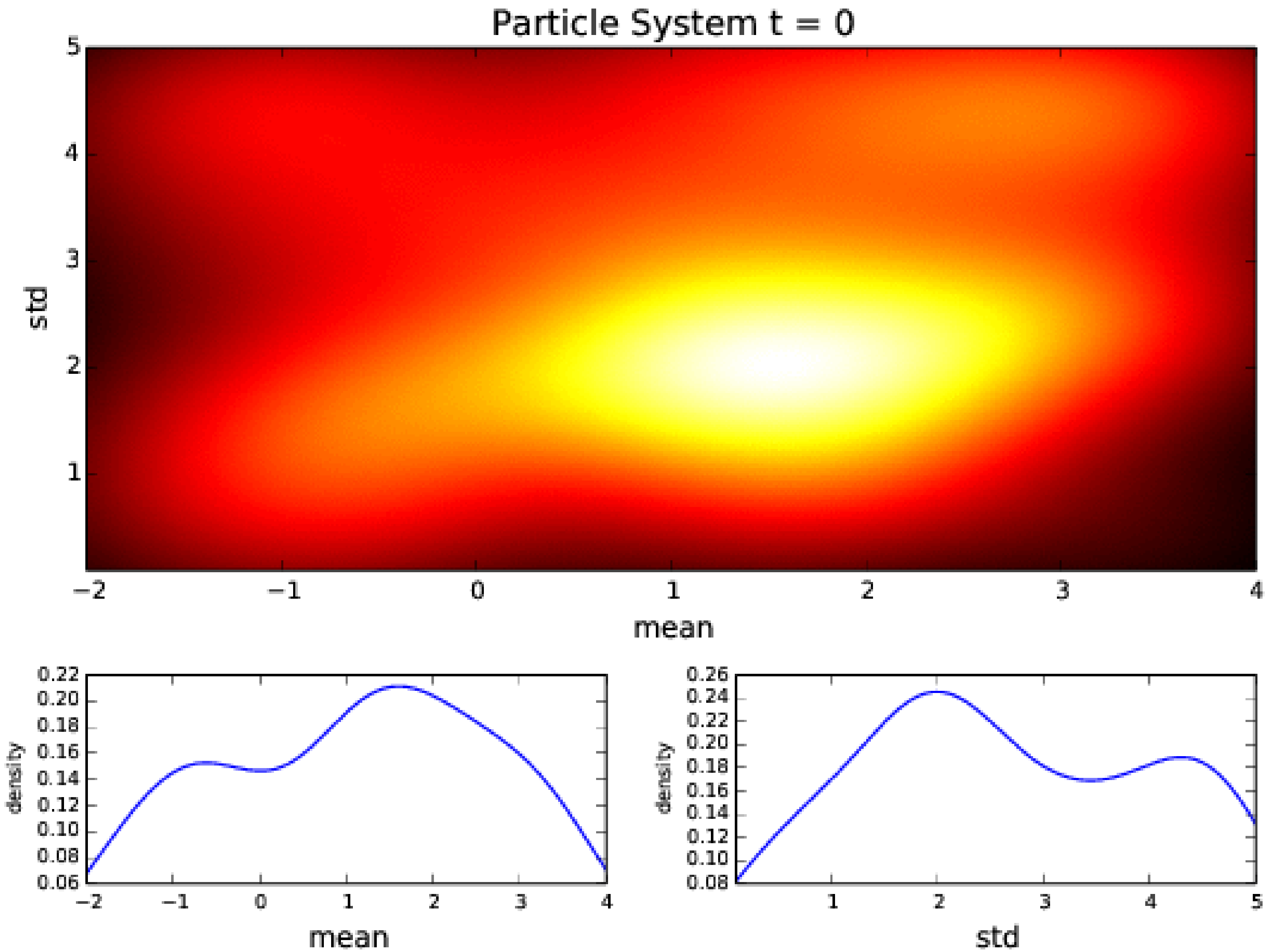
Distance: $$\rho = \mathrm{abs}\left(\frac{\bar{\mathcal{D}} - \bar{\mathcal{D}}_S}{\bar{\mathcal{D}}}\right) + \mathrm{abs}\left(\frac{\sigma_{\mathcal{D}} - \sigma_{\mathcal{D}_S}}{\sigma_{\mathcal{D}}}\right).$$

Y
11.2
15.3
200.0
0.5
14.3

Example: distance preview



Example: particle system evolution



in Astronomy

2012

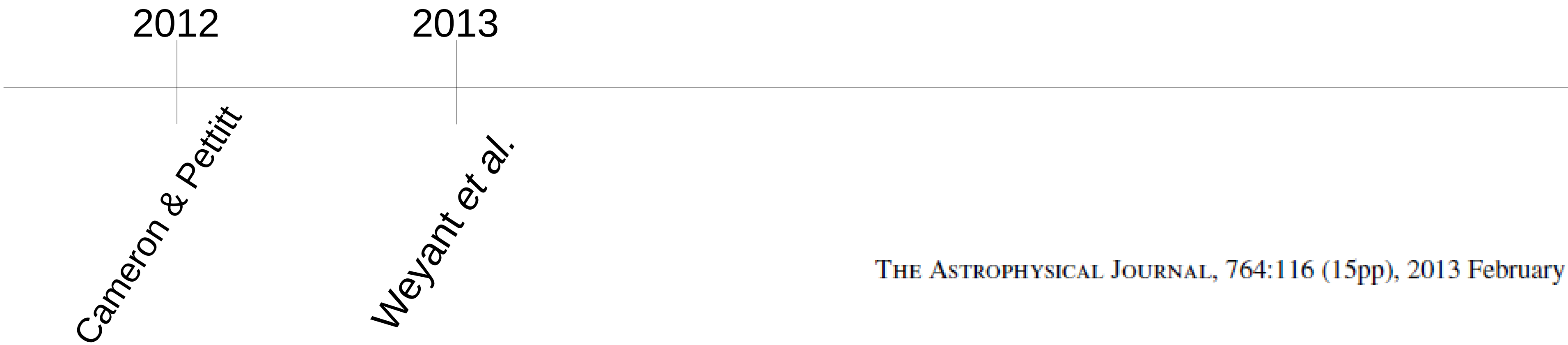
Cameron & Petit

Mon. Not. R. Astron. Soc. **425,** 44–65 (2012)

Approximate Bayesian Computation for astronomical model analysis: a case study in galaxy demographics and morphological transformation at high redshift

E. Cameron* and A. N. Pettitt

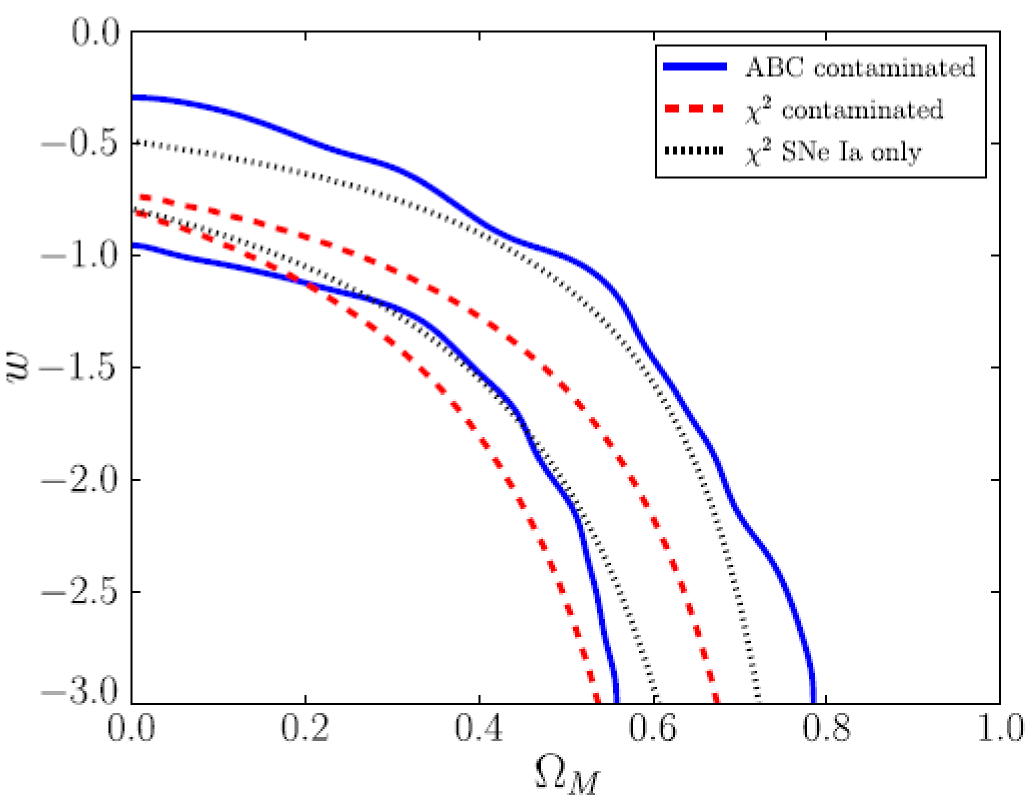
School of Mathematical Sciences (Statistical Science), Queensland University of Technology (QUT), GPO Box 2434, Brisbane 4001, QLD, Australia

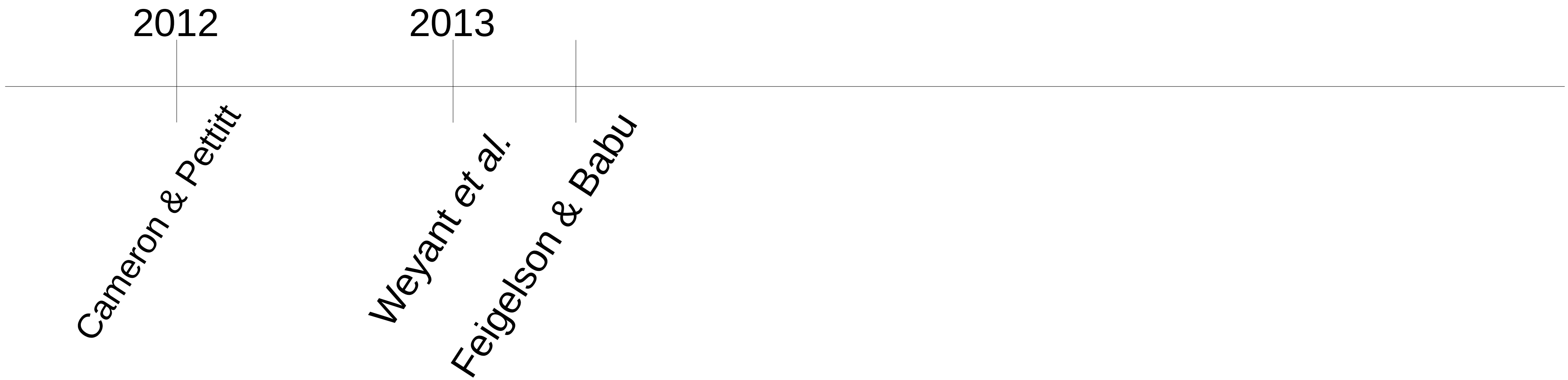


The Astrophysical Journal, 764:116 (15pp), 2013 February 20

LIKELIHOOD-FREE COSMOLOGICAL INFERENCE WITH TYPE Ia SUPERNOVAE: APPROXIMATE BAYESIAN COMPUTATION FOR A COMPLETE TREATMENT OF UNCERTAINTY

Anja Weyant[1], Chad Schafer[2], and W. Michael Wood-Vasey[1]

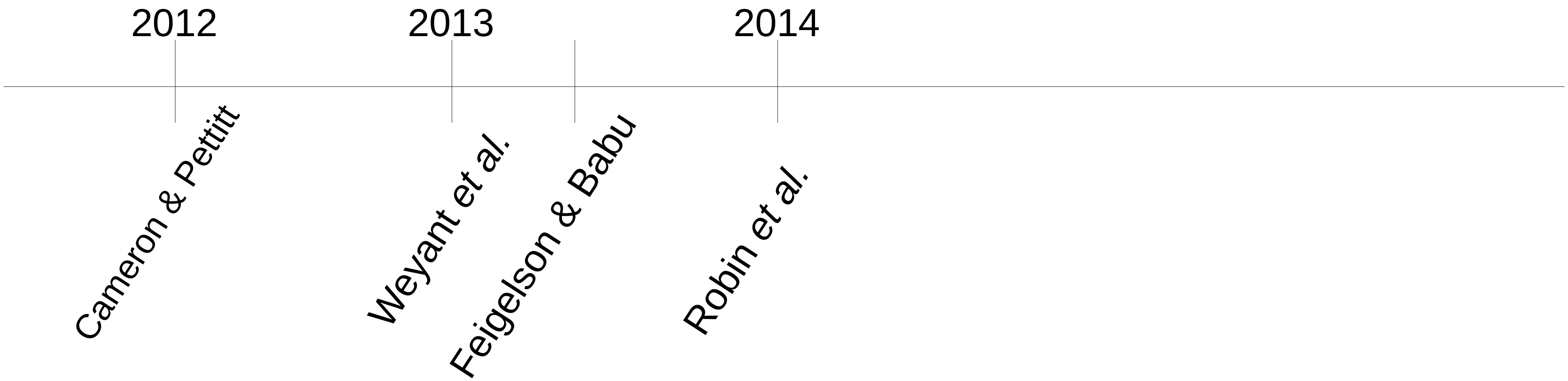




E.D. Feigelson and G.J. Babu (eds.), *Statistical Challenges in Modern Astronomy V*,
Lecture Notes in Statistics 209, DOI 10.1007/978-1-4614-3520-4_1,

Chapter 1
Likelihood-Free Inference in Cosmology: Potential for the Estimation of Luminosity Functions

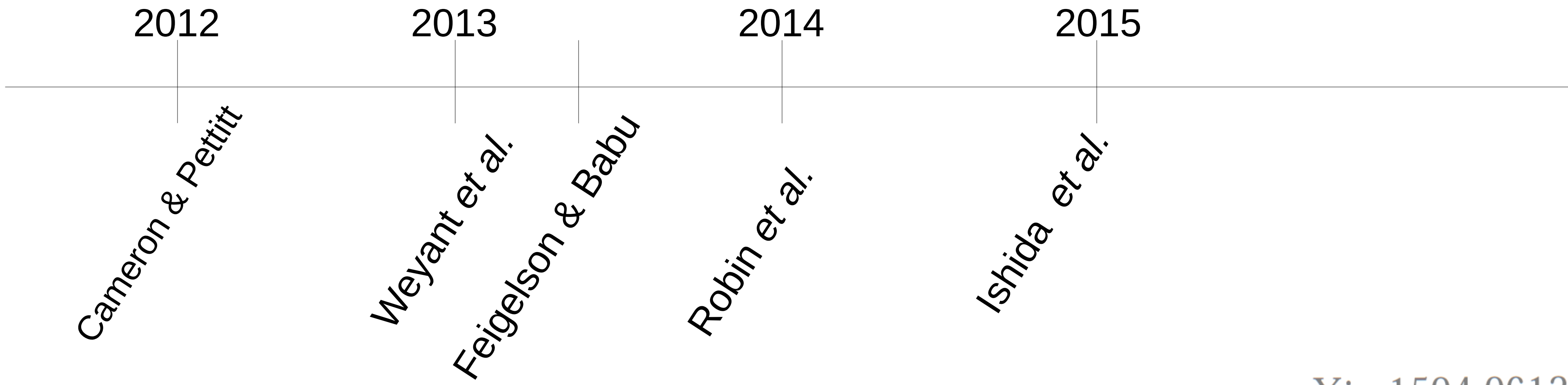
Chad M. Schafer and Peter E. Freeman



A&A 569, A13 (2014)

Constraining the thick disc formation scenario of the Milky Way★

A. C. Robin[1], C. Reylé[1], J. Fliri[2,3], M. Czekaj[4], C. P. Robert[5], and A. M. M. Martins[1]



arXiv:1504.06129v2

cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation

E. E. O. Ishida[1], S. D. P. Vitenti[2], M. Penna-Lima[3,4], J. Cisewski[5], R. S. de Souza[6], A. M. M. Trindade[7,8]
E. Cameron[9] and V. C. Busti[10], for the COIN collaboration

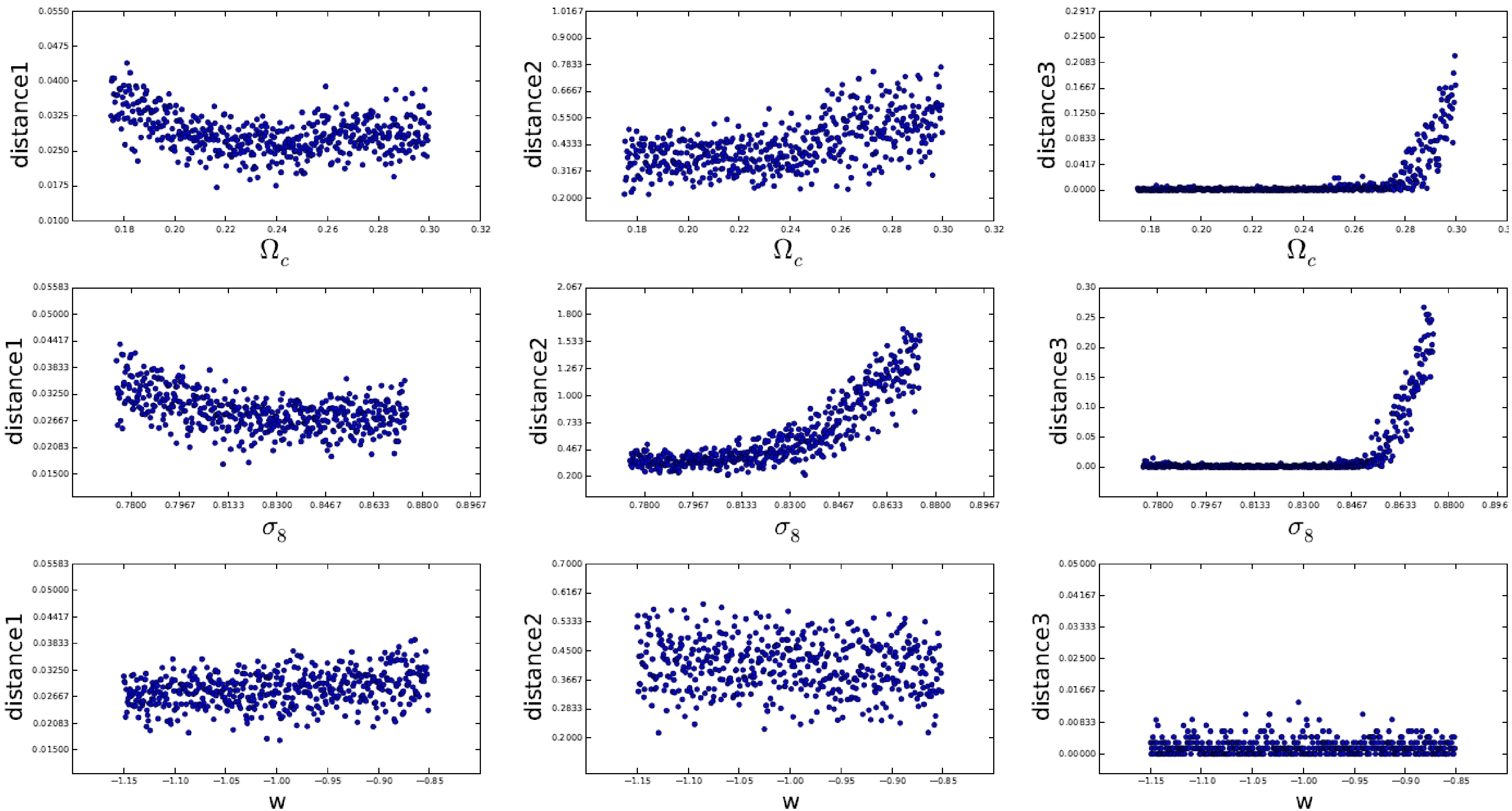
Cosmostatistics Initiative

COIN

cosmoabc

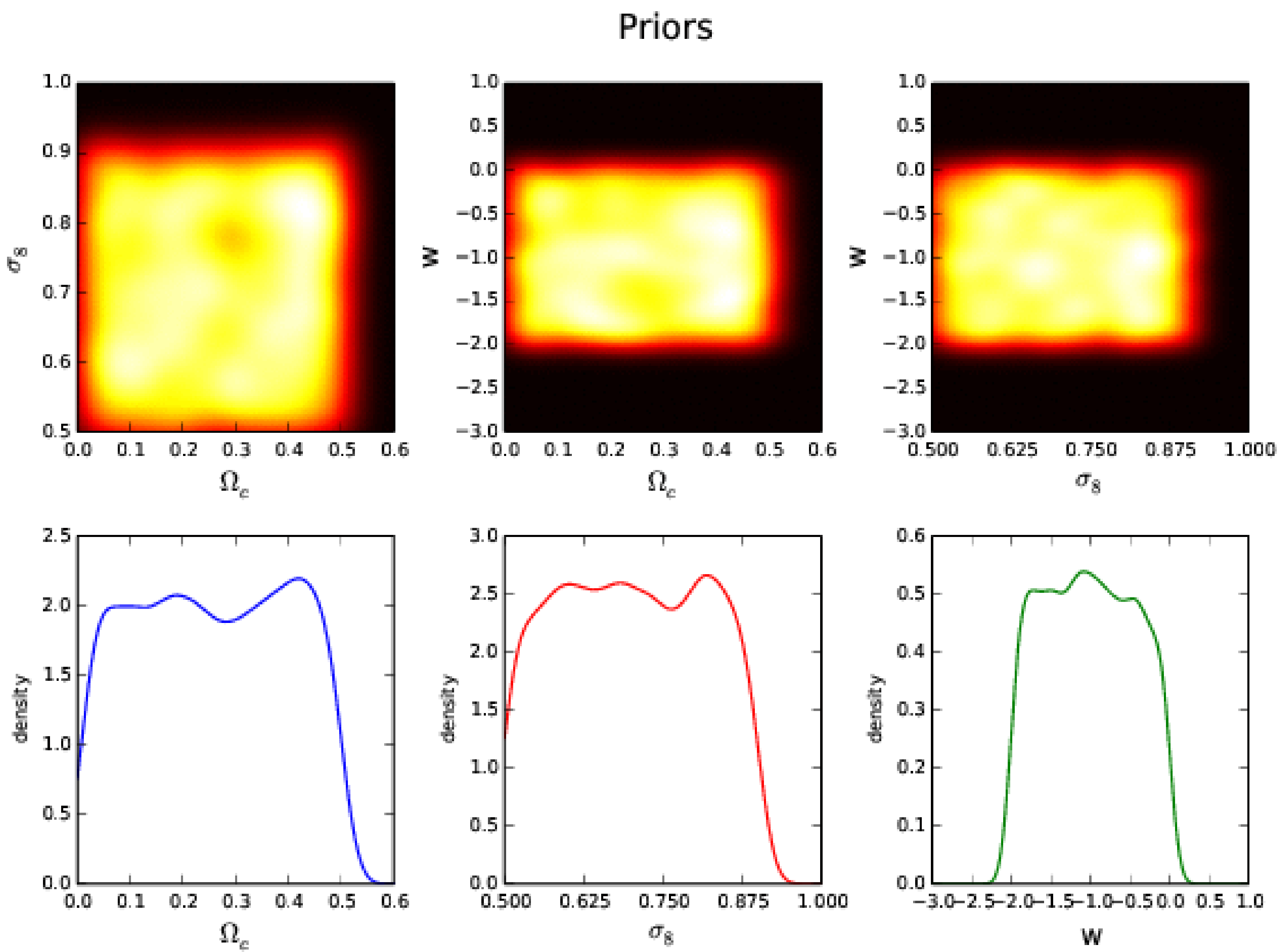
arXiv:1504.06129v2

Distance

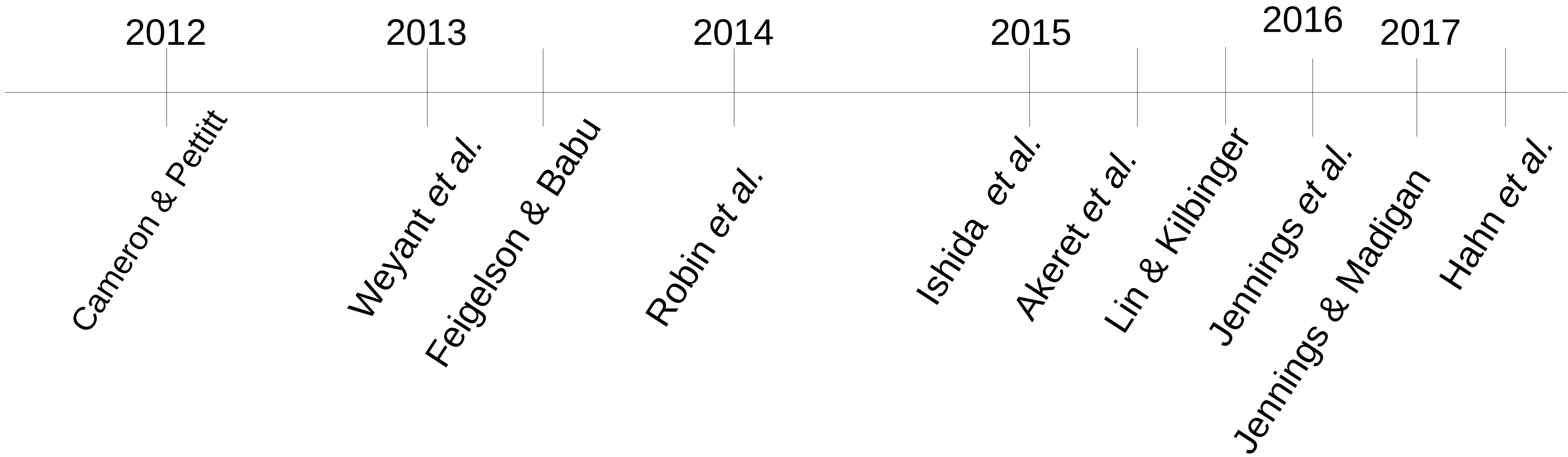


http://cosmoabc.readthedocs.org/en/latest/

cosmoabc



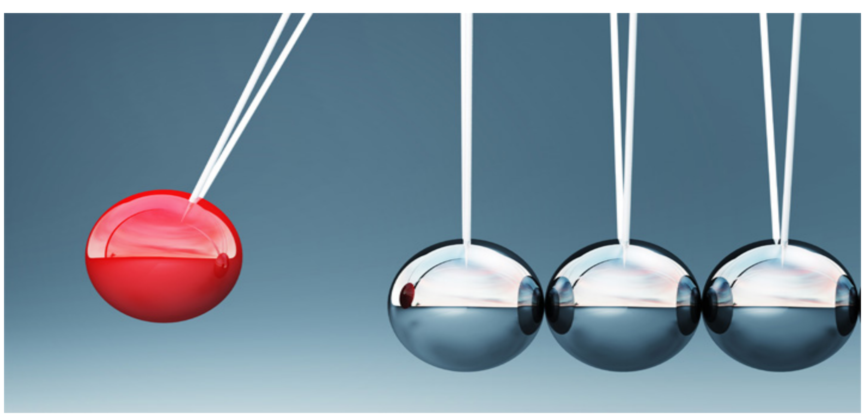
See frames in the pdf arxiv (open with Adobe Reader)


2012
2013
2014
2015
2016
2017
Cameron & Pettitt
Weyant et al.
Feigelson & Babu
Robin et al.
Ishida et al.
Akeret et al.
Lin & Kilbinger
Jennings et al.
Jennings & Madigan
Hahn et al.

Summary

- ✓ Good alternative when likelihood is not available.
- ✓ Becomes more attractive with faster simulations.

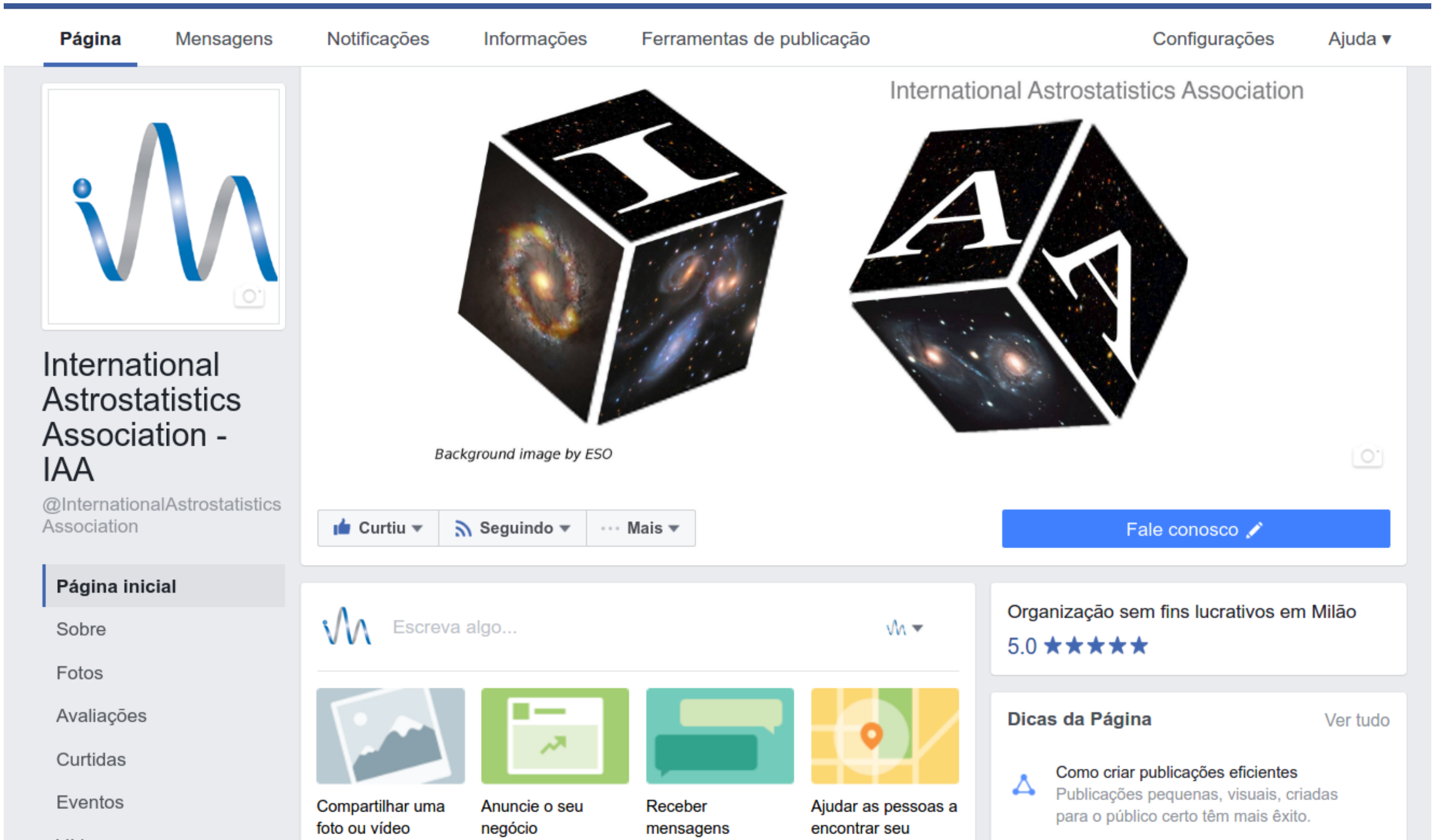
- ✗ Definition of distance function/summary statistics



Promising perspective as it gains momentum inside the astronomical community!

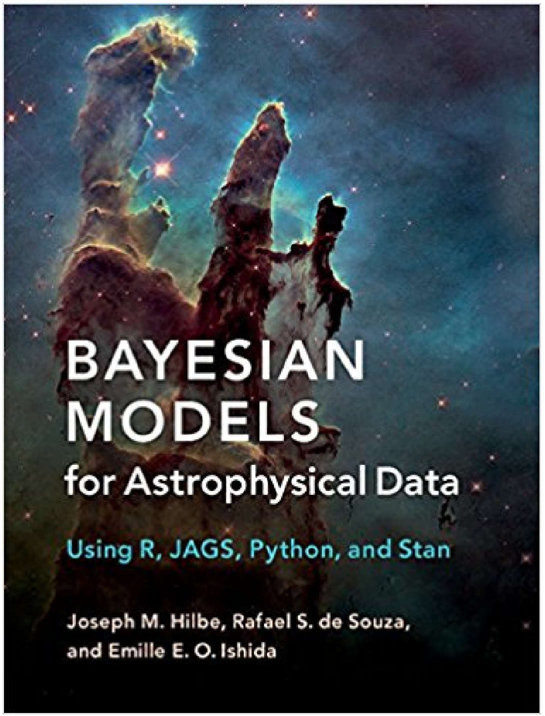
Thank you!

IAA facebook page



COIN on twitter

A unique endeavor requires unique tools



Designed to foster a new generation of interdisciplinary – and multi-lingual scientists

Cover statements by:

Benjamin Wandelt – *Lagrange Institute, Paris*
Eric Feigelson – *Penn State, USA*
Roberto Trotta – *Imperial College, UK*

Published by

CAMBRIDGE
UNIVERSITY PRESS

in April/2017

First book *for astronomers* to show working examples in Stan

Examples of ABC application

...The obvious class of problems is population studies that use population synthesis (i.e., simulating an astrophysical population). The earliest work I know of using ABC-like ideas is a thread of papers from the late 1990s/early 00s by Zaven Arzoumanian, David Chernoff, and Jim Cordes, using population synthesis (including Galactic orbital dynamics) to simulate the pulsar population in order to constrain the initial kick distribution. If you do an ADS search on Arzoumanian and Chernoff you'll find the string of papers. It culminated in this paper, which has had a big impact:

The Velocity Distribution of Isolated Radio Pulsars
http://adsabs.harvard.edu/abs/2002ApJ...568..289A

Tom Loredo, ASAIP Bayesian forum

Examples of ABC application

The hidden Potts model (used, e.g., in image segmentation Algorithms).

The model is easy to write down statistically in terms of a simple relationship between adjacent pixels, and it's easy to Gibbs sample from pixel-by-pixel at a fixed temperature (for which the normalization constant is unnecessary), BUT to perform inference of the temperature parameter too we need the proper likelihood which requires computation of the normalization constant (a sum over a huge combinatorial space) at each given temperature. When the image size is above ~1000x1000 pixels we simply don't have the computational power to make this calculation multiple times (as for an MCMC algorithm) so we consider the likelihood Intractable.

See section 4 of http://arxiv.org/pdf/1403.4359.pdf

Ewan Cameron, private communication

Algorithm

cosmoabc

Data: $\mathcal{D} \longrightarrow$ observed catalogue.
Result: ABC-posteriors distributions over the model parameters.

$t \longleftarrow 0$
$K \longleftarrow M$
for $J = 1, \ldots, M$ **do**
  Draw θ, from the prior, $p(\theta)$.
  Use θ to generate $\mathcal{D}_S$.
  Calculate distance, $\rho = \rho(\mathcal{D}_S, \mathcal{D})$.
  Store parameter and distance values,
  $\mathcal{S}_{ini} \leftarrow \{\theta, \rho\}$
end
Sort elements in $\mathcal{S}_{ini}$ by $|\rho|$.
Keep only the N parameter values with lower distance in $\mathcal{S}_{t=0}$.
$C_{t=0} \longleftarrow$ covariance matrix from $\mathcal{S}_{t=0}$
for $L = 1, \ldots, N$ **do**
  $W_1^L \longleftarrow 1/N$
end
while $N/K > \Delta$ **do**
  $K \longleftarrow 0$.
  $t \longleftarrow t + 1$.
  $\mathcal{S}_t \longleftarrow []$
  $\epsilon_t \longleftarrow 75^{th}$-quantile of distances in $\mathcal{S}_{t-1}$.
  while len($\mathcal{S}_t$) $<$ N **do**
    $K \longleftarrow K + 1$
    Draw θ_0 from $\mathcal{S}_{t-1}$ with weights W_{t-1}.
    Draw θ, from $\mathcal{N}(\theta_0, C_{t-1})$.
    Use θ to generate $\mathcal{D}_S$.
    Calculate distance, $\rho = \rho(\mathcal{D}_S, \mathcal{D})$
    if $\rho \leq \epsilon_t$ **then**
      $\mathcal{S}_t \longleftarrow \{\theta, \rho, K\}$
      $K \longleftarrow 0$
    end
  end
  for $J = 1, \ldots, N$ **do**
    $W_t^J \longleftarrow$ equation (3).
  end
  $W_t \longleftarrow$ normalized weights.
  $C_t \longleftarrow$ weighted covariance matrix from $\{\mathcal{S}_t, W_t\}$.
end

$$W_t^j = \frac{p(\boldsymbol{\theta}_t^j)}{\sum_{i=1}^N W_{t-1}^i \mathcal{N}(\boldsymbol{\theta}_t^j; \boldsymbol{\theta}_{t-1}^i, C_{t-1})}, \qquad (3)$$