Bayesian Methods for Cosmology: Measuring Dark Energy

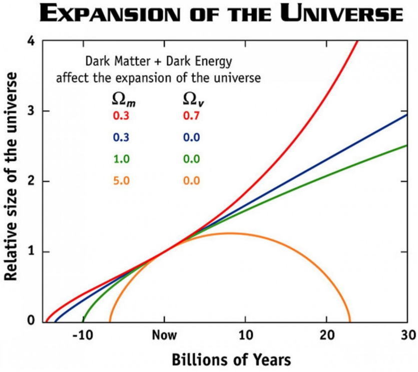
F. Sottovia and L. Di Filippo

Topics

1. Introduction to the physical model
2. MCMC (Monte Carlo Markov Chain)
3. ABC (Approximate Bayesian Computation)
4. Comparison and conclusions

Why dark energy?

- Universe expansion is accelerating
- Dark energy: scales as $\rho \propto a^0$
- What is it?
  - Cosmological constant
  - Quintessence
  - ...
- From experiments (PDG 2018): $\Omega_\Lambda \approx 0.69$, $\Omega_m \approx 0.31$

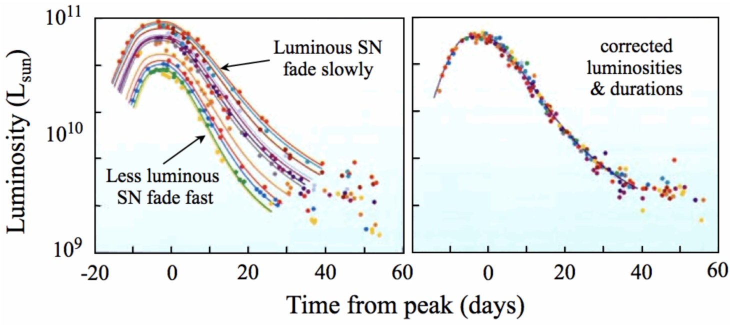


Type Ia Supernovae and Standard Candles

$$m_b = 5\log_{10} D_L + M \Rightarrow m_b = 5\log_{10} D_L - \alpha(s-1) + \beta c + M$$

where m_b: magnitude in the b-band, α: stretch parameter

β: color parameter, s: stretch, c: color



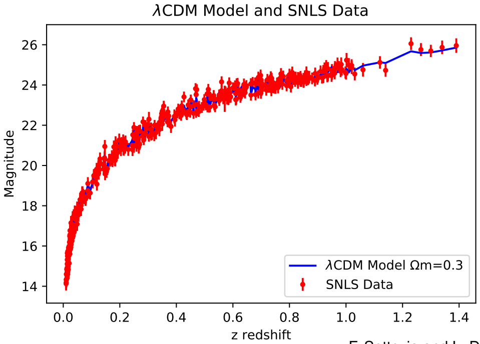
The Model

$$D_L = (1+z)\frac{c}{H_0}\int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda}}$$

Gaussian noise

Parameters: (Ω_m, α, β, M)

(for ABC (Ω_m, σ))



Why Bayesian methods in Cosmology?

- Incorporate previous knowledge about the Universe
- Methods useful for dealing with dimensionality curse:
  - MCMC: likelihood needed
  - ABC: no likelihood needed
- Tradition!

Monte Carlo Markov Chain (MCMC)

Bayes theorem: the posterior distribution of the parameters $\vec{\theta}$, given the prior $p(\vec{\theta})$ and the likelihood of the data $P(\text{data}|\vec{\theta})$, is:

$$P(\vec{\theta}|\text{data}) \propto P(\text{data}|\vec{\theta})p(\vec{\theta})$$

- So, given two parameter sets $\vec{\theta_1}$ and $\vec{\theta_2}$, we can use the ratio:

$$\frac{p(\vec{\theta_1}|\text{data})p(\vec{\theta_1})}{p(\vec{\theta_2}|\text{data})p(\vec{\theta_2})}$$

- If the prior is uniform, this simplifies to the likelihood ratio:

$$\frac{p(\vec{\theta_1}|\text{data})}{p(\vec{\theta_2}|\text{data})}$$

Algorithm

1. Choose an initial guess $\vec{\theta}_0$ and a proposal distribution

Algorithm

1. Choose an initial guess $\vec{\theta}_0$ and a proposal distribution
2. Using the proposal distribution guess new position $\vec{\theta}_1$
3. Should we accept the new position?

Algorithm

1. Choose an initial guess $\vec{\theta}_0$ and a proposal distribution
2. Using the proposal distribution guess new position $\vec{\theta}_1$
3. Should we accept the new position? $\Rightarrow$ look at the likelihood ratio:
   If $\frac{p(\vec{\theta}_1|\text{data})}{p(\vec{\theta}_0|\text{data})} > 1$: accept $\vec{\theta}_1$

Algorithm

1. Choose an initial guess $\vec{\theta}_0$ and a proposal distribution
2. Using the proposal distribution guess new position $\vec{\theta}_1$
3. Should we accept the new position? $\Rightarrow$ look at the likelihood ratio:
   If $\frac{p(\vec{\theta}_1|\text{data})}{p(\vec{\theta}_0|\text{data})} > 1$: accept $\vec{\theta}_1$
   If $\frac{p(\vec{\theta}_1|\text{data})}{p(\vec{\theta}_0|\text{data})} < 1$: accept $\vec{\theta}_1$ with probability $\frac{p(\vec{\theta}_1|\text{data})}{p(\vec{\theta}_0|\text{data})}$.
4. If $\vec{\theta}_1$ is not accepted, store $\vec{\theta}_0$ a second time.

Algorithm

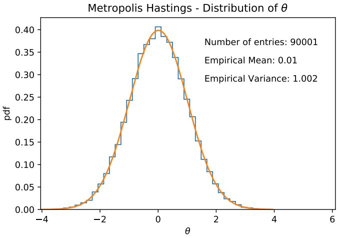
1. Choose an initial guess $\vec{\theta}_0$ and a proposal distribution
2. Using the proposal distribution guess new position $\vec{\theta}_1$
3. Should we accept the new position? $\Rightarrow$ look at the likelihood ratio:
   If $\frac{p(\vec{\theta}_1|\text{data})}{p(\vec{\theta}_0|\text{data})} > 1$: accept $\vec{\theta}_1$
   If $\frac{p(\vec{\theta}_1|\text{data})}{p(\vec{\theta}_0|\text{data})} < 1$: accept $\vec{\theta}_1$ with probability $\frac{p(\vec{\theta}_1|\text{data})}{p(\vec{\theta}_0|\text{data})}$.
4. If $\vec{\theta}_1$ is not accepted, store $\vec{\theta}_0$ a second time.
5. Repeat the iteration N times

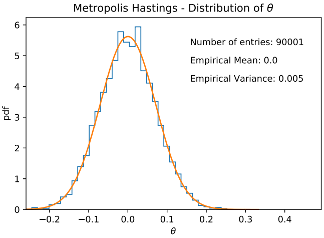
Toy model: One-dimensional Gaussian distribution

- Simulate a sample of N independent events according to $\mathcal{N}(\mu = 0, \sigma^2 = 1)$.
  The estimator of their mean will be distributed according to $\mathcal{N}(\mu = 0, \sigma^2 = 1/N)$ (CLT). (Here $\theta_{j+1} \sim \mathcal{N}(\theta_j, 1)$)
  Results: Mean and variance as expected!
  $\sigma^2_{N=200} = 0.005$ and $\sigma^2_{N=1} = 1.0$

Toy model: One-dimensional Gaussian distribution

- Simulate a sample of N independent events according to $\mathcal{N}(\mu = 0, \sigma^2 = 1)$.
  The estimator of their mean will be distributed according to $\mathcal{N}(\mu = 0, \sigma^2 = 1/N)$ (CLT). (Here $\theta_{j+1} \sim \mathcal{N}(\theta_j, 1)$)
  Results: Mean and variance as expected!
  $\sigma^2_{N=200} = 0.005$ and $\sigma^2_{N=1} = 1.0$





Back to our model (MCMC)

Back to our model (MCMC)

Parameter space is 4-dimensional: $\vec{\theta} = (\Omega_m, \alpha, \beta, M)$

Proposal distribution: try 4-dimensional Gaussian distro $\Rightarrow$ what covariance matrix?

Back to our model (MCMC)

Parameter space is 4-dimensional: $\vec{\theta} = (\Omega_m, \alpha, \beta, M)$
Proposal distribution: try 4-dimensional Gaussian distro $\Rightarrow$ what covariance matrix?

- Id $\Rightarrow$ acceptance ratio $\approx 4 \cdot 10^{-5}$
- Estimated covariance with first iteration with Id $\Rightarrow$ acceptance ratio $\gtrapprox 0.6$: sweet spot!
- Estimated covariance after a couple iterations $\Rightarrow$ acceptance ratio $\gtrapprox 0.3$: still good

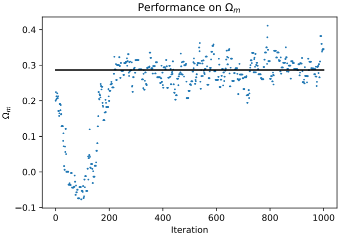
Back to our model (MCMC)

Parameter space is 4-dimensional: $\vec{\theta} = (\Omega_m, \alpha, \beta, M)$

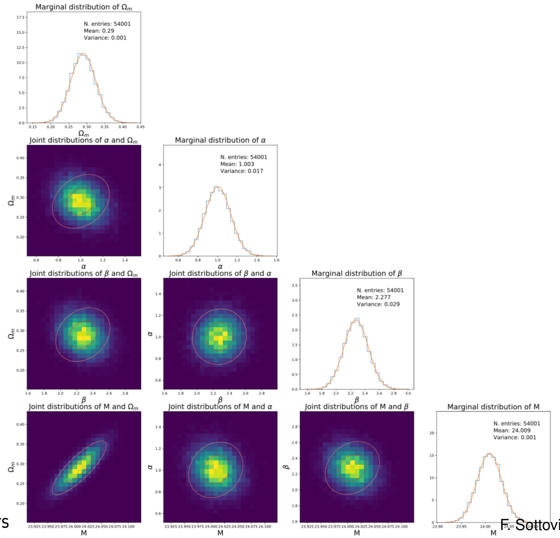
Proposal distribution: try 4-dimensional Gaussian distro $\Rightarrow$ what covariance matrix?

- Id $\Rightarrow$ acceptance ratio $\approx 4 \cdot 10^{-5}$
- Estimated covariance with first iteration with Id $\Rightarrow$ acceptance ratio $\gtrapprox 0.6$: sweet spot!
- Estimated covariance after a couple iterations $\Rightarrow$ acceptance ratio $\gtrapprox 0.3$: still good

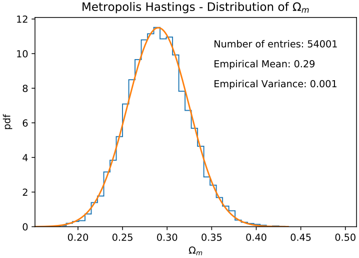
- How fast is the convergence?

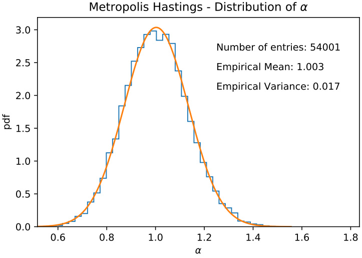


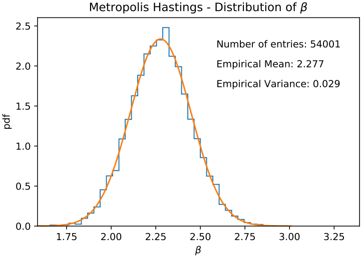
Estimated distr., N = 60000, discarding first 10%

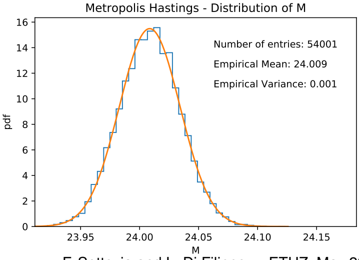


MCMC: Results

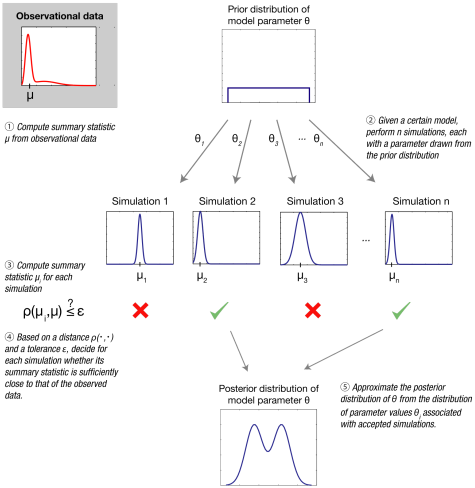








ABC (Approximate Bayesian Computation)



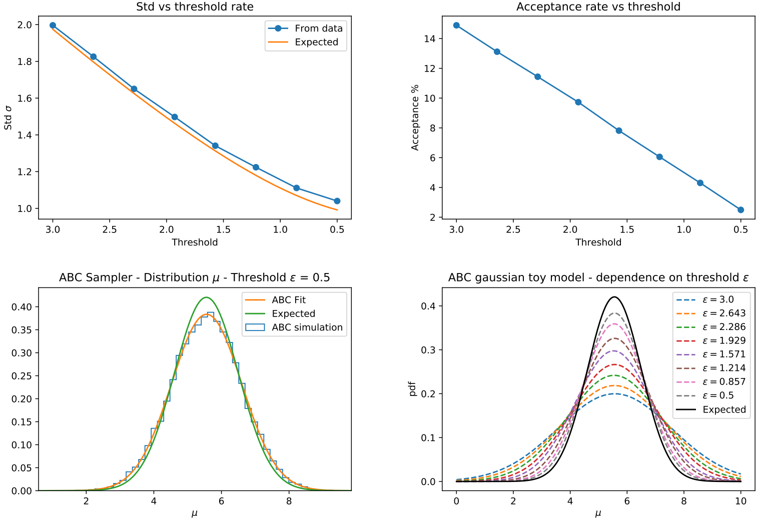
ABC Toy Model

- Gaussian model: 100 samples, $\sim \mathcal{N}(\mu, \sigma = 10)$
  $\Rightarrow$ estimate μ
- Expected variance on μ: $\sigma_\mu^2 = \frac{10^2}{100} = 1$ (CLT)
- From theory [2]: $\sigma_\mu^2(\epsilon) = \frac{\sigma_{\text{data}}^2}{N} + \frac{\epsilon^2}{3}$ (for flat prior)

ABC Toy Model

- Gaussian model: 100 samples, $\sim \mathcal{N}(\mu, \sigma = 10)$
  $\Rightarrow$ estimate μ
- Expected variance on μ: $\sigma_\mu^2 = \frac{10^2}{100} = 1$ (CLT)
- From theory [2]: $\sigma_\mu^2(\epsilon) = \frac{\sigma_{\text{data}}^2}{N} + \frac{\epsilon^2}{3}$ (for flat prior)
- **Results**: $\mu \approx 5.55$ (uncertainty depends on threshold)

ABC Toy Model - Results



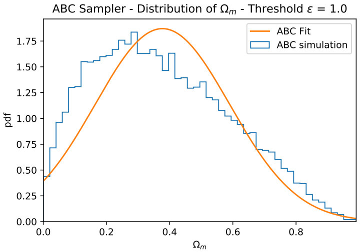
Back to our model (ABC)

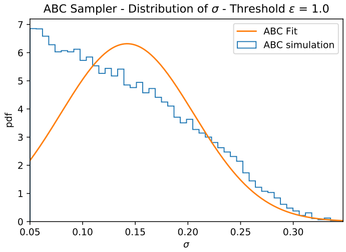
- Our model: gaussian noise $\sim \mathcal{N}(0, \sigma)$ on each (z, D_L)
- 2-dimensional parameter space, $\vec{\theta} = (\Omega_m, \sigma)$
- Distance measure
  a) $d = \frac{1}{N} \sum_z \left| D_{L,\text{obs.}}(z) - D_{L,\text{sim.}}(z) \right|$
  b) $d = \frac{1}{N} \sum_z \left(D_{L,\text{obs.}}(z) - D_{L,\text{sim.}}(z) \right)^2$
  c) $d = \max_z \left| D_{L,\text{obs.}}(z) - D_{L,\text{sim.}}(z) \right|$
- Flat prior on both parameters:

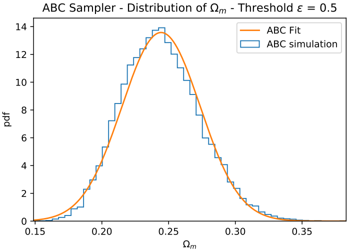
  $\Omega_m \in [0, 1]$

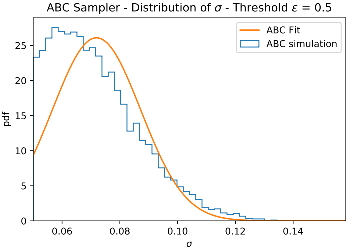
  $\sigma \in [0.05, 0.35]$: est. noise from data with $\sigma = 0.15$

ABC: Results

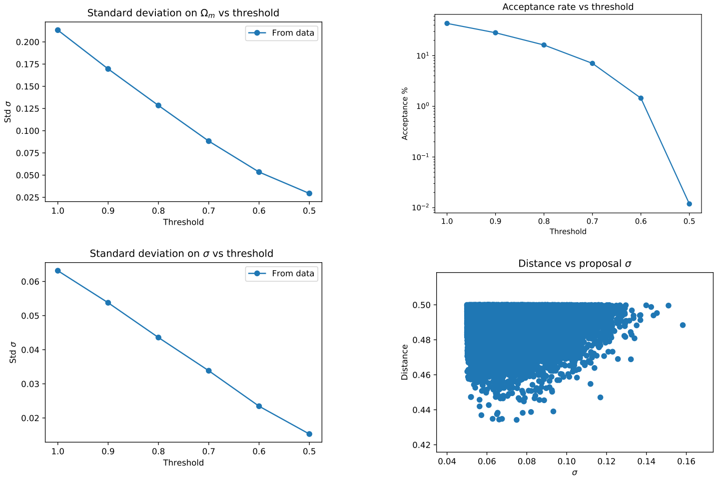

ABC Sampler - Distribution of Ω_m - Threshold ϵ = 1.0
ABC Fit
ABC simulation
pdf
Ω_m


ABC Sampler - Distribution of σ - Threshold ϵ = 1.0
ABC Fit
ABC simulation
pdf
σ


ABC Sampler - Distribution of Ω_m - Threshold ϵ = 0.5
ABC Fit
ABC simulation
pdf
Ω_m


ABC Sampler - Distribution of σ - Threshold ϵ = 0.5
ABC Fit
ABC simulation
pdf
σ

ABC: Performance



Hints of convergence for $\epsilon \lesssim 0.5$...

ABC: Results and Observations

- Results: $\Omega_m = 0.24 \pm 0.03$, $\sigma = 0.07 \pm 0.02$
- Improper fitting on σ!!

ABC: Results and Observations

- Results: $\Omega_m = 0.24 \pm 0.03$, $\sigma = 0.07 \pm 0.02$
- Improper fitting on σ!!
- Why?
  1. Choice of the distance
  2. Threshold - computation time
  3. Noise model
  4. Should we really get a gaussian distribution?

Conclusions

- MCMC more efficient (smaller variance), more “robust”

Conclusions

- MCMC more efficient (smaller variance), more “robust”
- ABC: no likelihood needed

Conclusions

- MCMC more efficient (smaller variance), more “robust”
- ABC: no likelihood needed
- Convergence: better in MCMC, but depends on various factors!
- Important: in MCMC high acceptance ratio not at expense of distribution
  ABC: high acceptance ratio for high threshold $\Rightarrow$ “bad” posterior

Conclusions

- MCMC more efficient (smaller variance), more “robust”
- ABC: no likelihood needed
- Convergence: better in MCMC, but depends on various factors!
- Important: in MCMC high acceptance ratio not at expense of distribution
  ABC: high acceptance ratio for high threshold $\Rightarrow$ “bad” posterior
- Results:
  - MCMC: $\Omega_m = 0.29 \pm 0.01$
  - ABC: $\Omega_m = 0.24 \pm 0.03$
  - Literature [3]: $\Omega_m = 0.308 \pm 0.012$

References

- ABC page on Wikipedia: https://en.wikipedia.org/wiki/Approximate_Bayesian_computation
- Joel Akeret, Alexandre Refregier, Adam Amara, Sebastian Seehars, Caspar Hasner
  Approximate Bayesian Computation for Forward Modeling in Cosmology
  https://en.wikipedia.org/wiki/Approximate_Bayesian_computation
- Particle Data Group 2018
  http://pdg.lbl.gov/2018/html/computer_read.html

Thank you for your attention!