Bayesian Inference in Astronomy & Astrophysics by Markov Chain Monte Carlo (MCMC)

Chaoli Zhang

Supervisors:

Li Chen & Zhengyi Shao

Shanghai Observatory

July 20, 2016

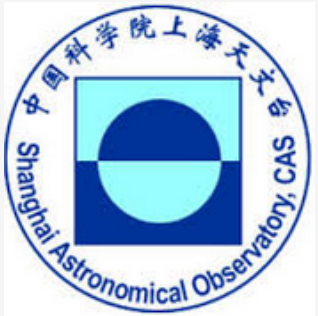


Table of contents

Bear in mind when you first time study MCMC.....

The things that I find hard to understand push me to my limits. One of the things that I have always found hard is **Markov Chain Monte Carlo Methods**. When I first encountered them, I read a lot about them but mostly it ended like this.

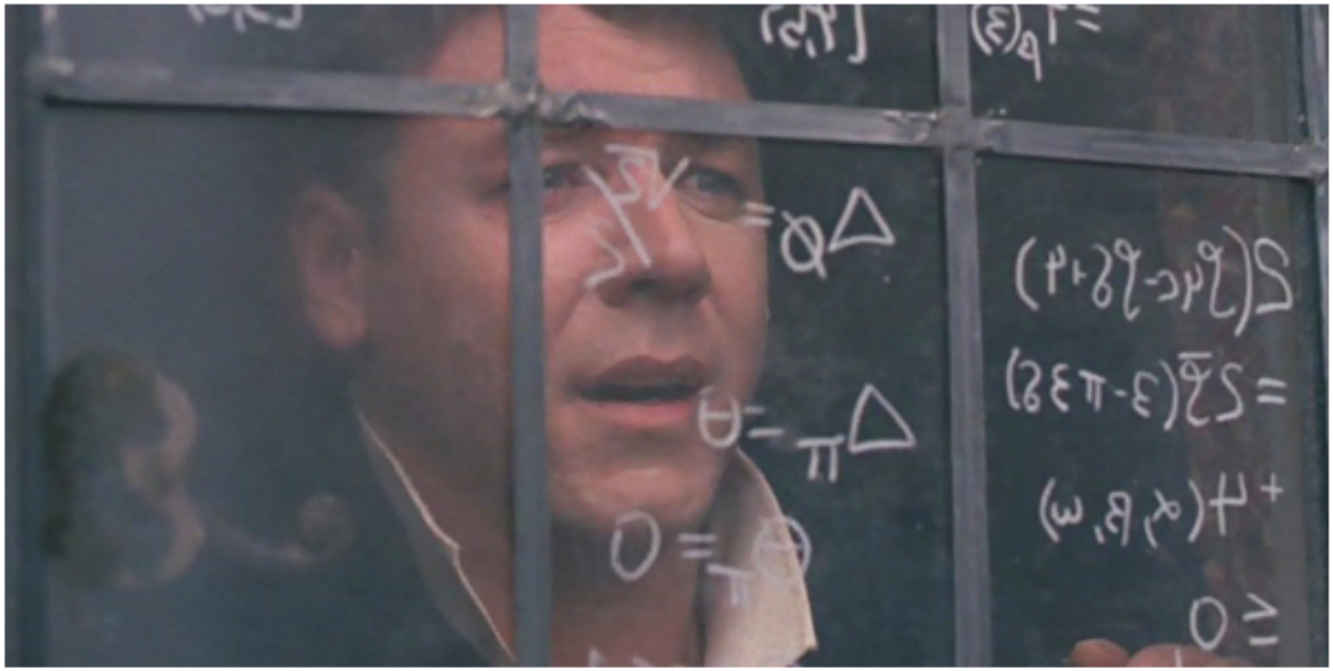
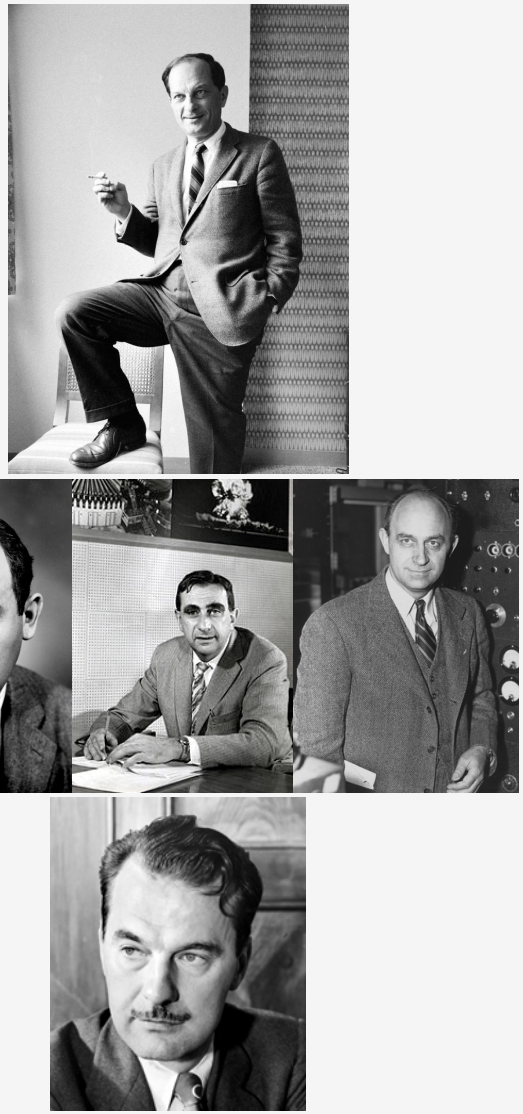


Figure: John Nash, "A Beautiful Mind (2001)"

A bit of history



- Stan Ulam: Solitaire (1946)
- John Von Neumann (Manhattan Project, Nuclear Bomb)
- Nick Metropolis (Klari Von Neumann $\rightarrow$ Metropolis & Ulam, 1949)
- Fermi, Teller, Feynman, & Gamow go to Metropolis for solving their problems
- Metropolis and Hastings ($\Rightarrow$Metropolis-Hastings algorithms 1970s)
- Top 10 ranked algorithm in 20th century.

Markov Chain Monte Carlo method

Markov Chain Monte Carlo (MCMC) : The goal of MCMC is to quickly draw samples from some probability distribution (presumably in very high dimension) without having to know its exact height at any point.

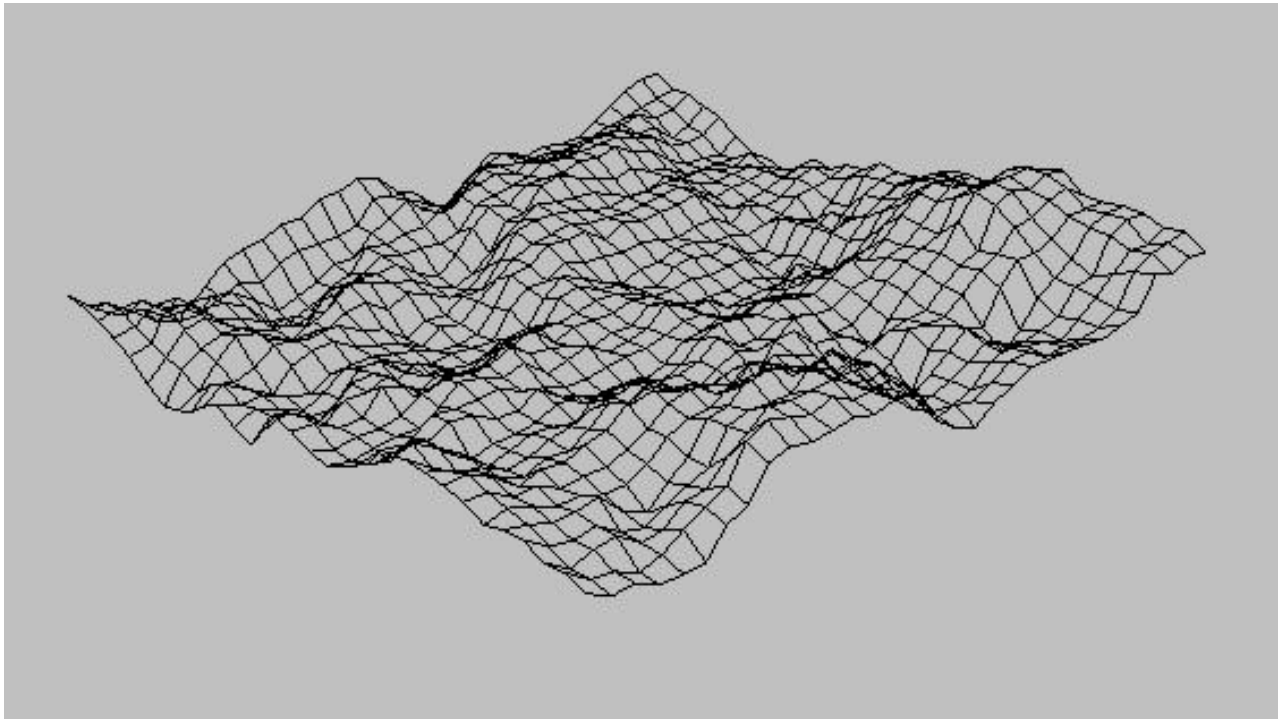
Curse of dimensionality (Richard E. Bellman) :
when the dimensionality increases, the volume of the space increases so fast that the available data become sparse. e.g. if 10 evaluations provide adequate accuracy in one dimension, then 10^{100} points are needed for 100 dimension – way too many to be computed.

Markov Chain Monte Carlo method

Markov Cha quickly
draw sampl very high
dimension) nt.

Curse of d
when the di eases so
fast that th s provide
adequate ac ed for 100
dimension -

probability density surface (2D)



Monte Carlo

Pinball Machine

- is named by Metropolis referring to the Monte Carlo Casino in Monaco

Monte Carlo

Casino in Monaco

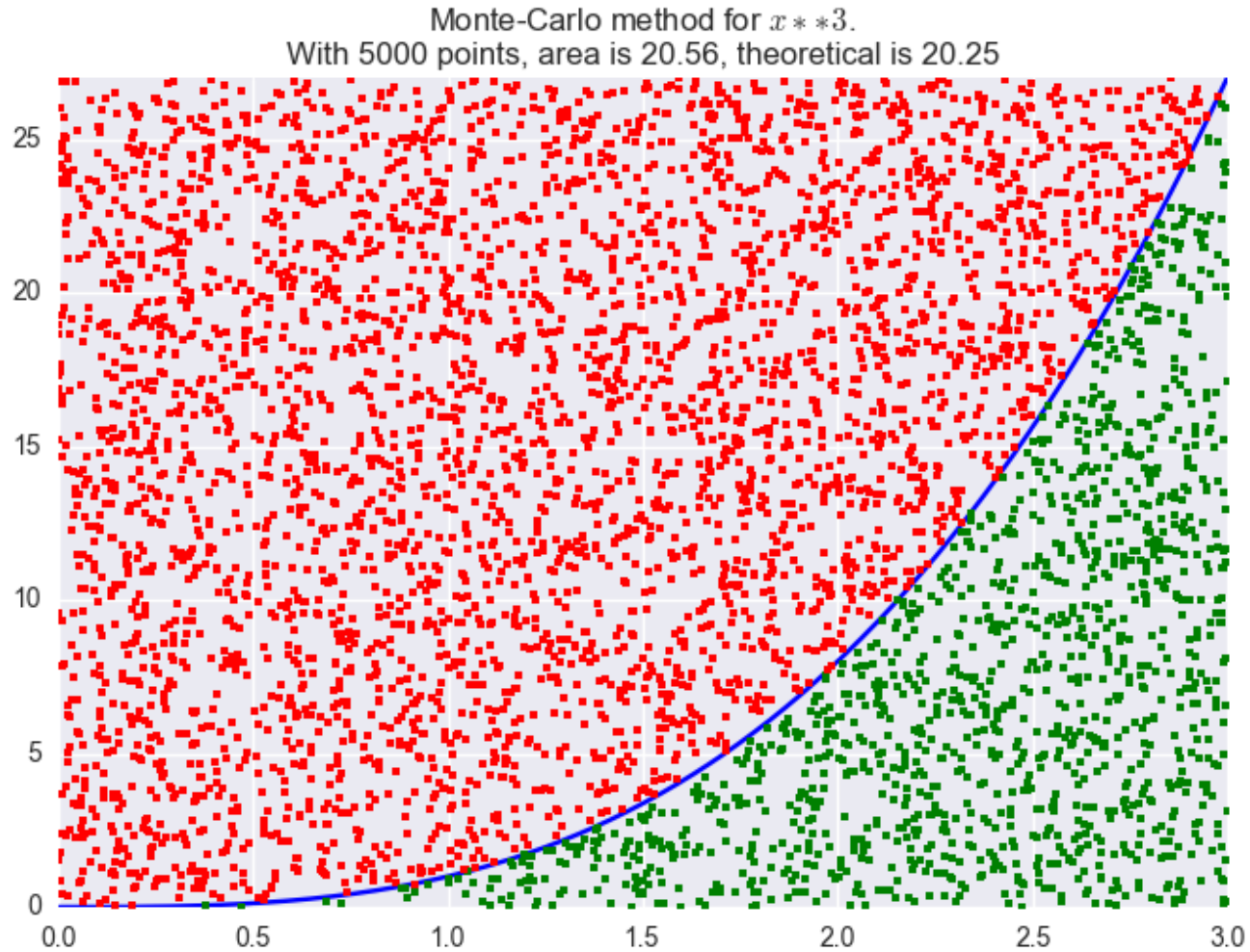
Monte Carlo

Pinball Machine

- is named by Metropolis referring to the Monte Carlo Casino in Monaco
- analytically complex problems
- use simulation to find solutions to the following
  1. optimization
  2. numerical integration
  3. sampling probability distribution.

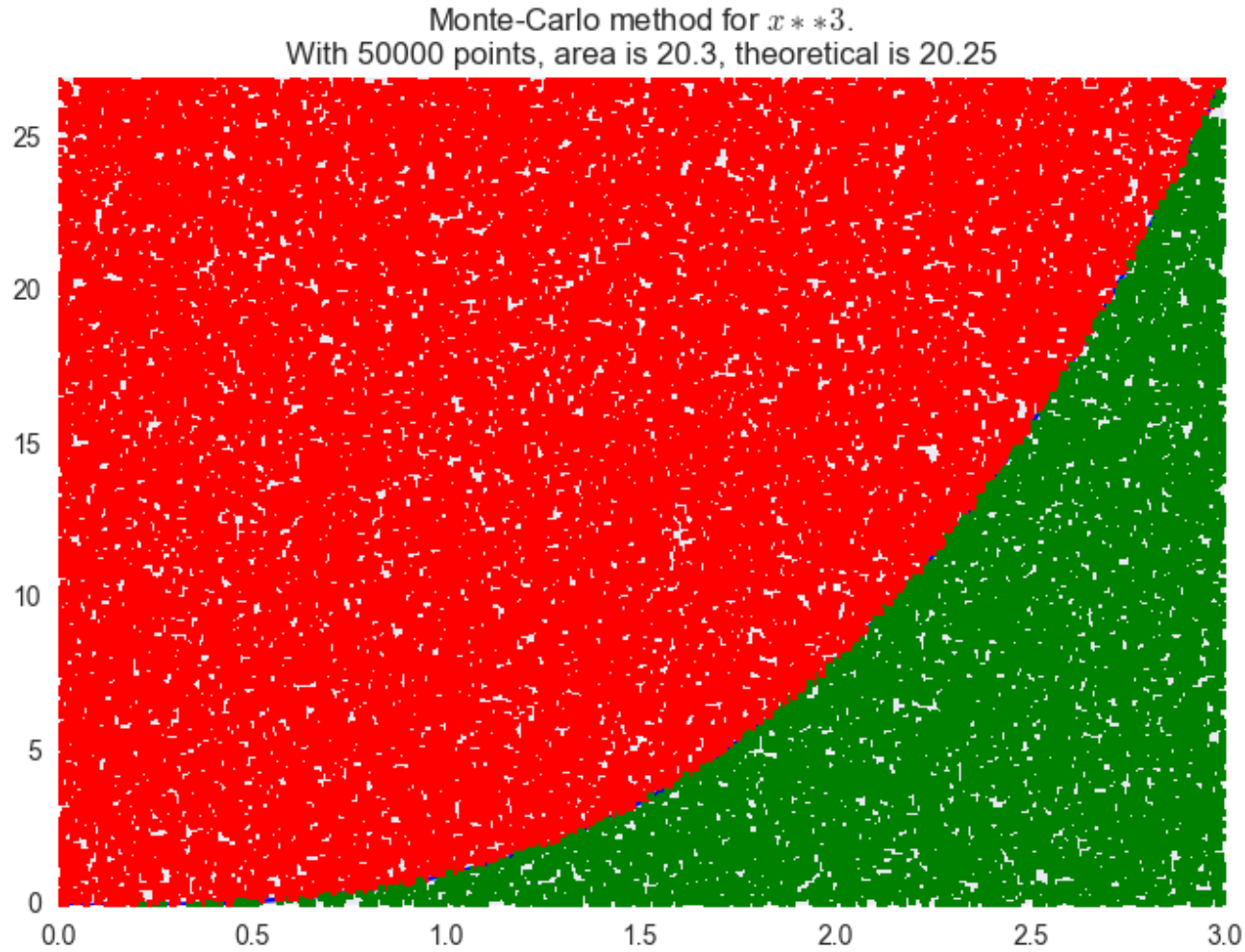
Monte Carlo

Sampling method



Monte Carlo

Sampling method



Monte Carlo Method

Now let's solve the problem by numerical integration i.e. we want to evaluate the integral

$$I = \int_{\text{lower}}^{\text{upper}} x^3 \, dx \tag{1}$$

Then the Monte Carlo method suggested we firstly to sample uniformly at interval [upper , lower] which gives N uniform samples $\{s_1, s_2, s_3 ... s_n\} \in [lower, upper]$, then the integral can be approximated by

$$\tilde{I} = V * \frac{1}{N} \sum_{i=1}^{N} s_i^3 \tag{2}$$

$$= (\text{upper-lower}) \frac{1}{N} \sum_{i=1}^{N} s_i^3 \tag{3}$$

Markov Chain

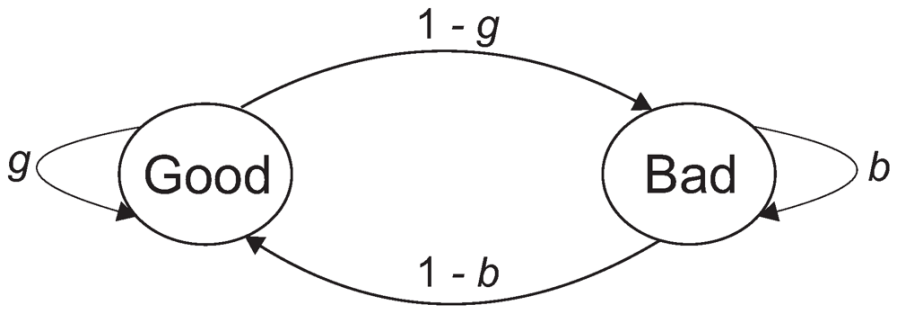
Markov Chain: a stochastic process in which future states depend solely on current state. i.e.

$$p(x^{(i+1)}|x^{(1)}, x^{(2)}, x^{(3)}..., x^{(i)}) = p(x^{(i+1)}|x^{(i)}) \tag{4}$$

Example:
two states markov chain

$$T = \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix} \tag{5}$$

where $g = 0.9$ and $b = 0.6$



Two states Markov Chain

Markov Chain

We see that regardless of any initial distribution (a, b) we use, the chain will stabilise at $p(x) = (0.8, 0.2)$. This stability result plays a fundamental role in MCMC simulation.
In term of mathematics, the Markov chain will convergence to the invariant distribution p(x), as long as Markov Chain obeys following properties [1]:

- *Irreducibility* For any state of the Markov chain,there is a positive probability of visiting all other states.
- *Aperiodicity* The chain should not get trapped in cycles.

Markov Chain played such important role in our daily life e.g. every time we google (or baidu), the public bike in the city...

[1] (Ergodic theorem: the behavior of a dynamical system when $t \to \infty$)

Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm is the most popular MCMC method. And in fact, most MCMC algorithms can be interpreted as special cases or extensions of this algorithm.

Basic idea:

- pick a new "proposed" location (similar to transitional kernel)
- figure out how much higher or lower that location is compared to your current location
- probabilistically stay put or move to that location

Metropolis-Hastings algorithm

The Me IC
method
special

Basic id
- pic
- fig to
you
- pro

Metropolis Hastings Algorithm

1. Initialise $x^{(0)}$.
2. For $i = 0$ to $N - 1$
   - Sample $u \sim \mathcal{U}_{[0,1]}$.
   - Sample $x^\star \sim q(x^\star | x^{(i)})$.
   - If $u < \mathcal{A}(x^{(i)}, x^\star) = \min\left\{1, \frac{p(x^\star)q(x^{(i)}|x^\star)}{p(x^{(i)})q(x^\star|x^{(i)})}\right\}$

     $$x^{(i+1)} = x^\star$$

     else

     $$x^{(i+1)} = x^{(i)}$$

Bayesian Statistics

$$\mathrm{P}(\Theta|data) = \frac{\mathrm{P}(data|\Theta)\mathrm{P}(\Theta)}{\mathrm{P}(data)} \tag{6}$$

where

- $\Theta \in \{Z, \text{age}, \text{distance}, \text{E(B-V)}\}$
- $\mathrm{P}(\Theta|data)$ is posterior
- $\mathrm{P}(data|\Theta)$ is likelihood
- $\mathrm{P}(\Theta)$ is prior
- $\mathrm{P}(data)$ is bayesian evidence (challenging for model selection!!!)

Example on flipping the coins

You flip the coin say 100 times, and get 41 heads. Questions: is this coin fair (what is the fairness of the coin) ?

Example on normal distribution with fixed variance

- data: generate from normal distribution $Normal(\mu = 0, \sigma = 1)$ (500)
- prior is $P(\mu) = \frac{1}{\sqrt{2\pi\sigma^2_{prior}}} \exp\left\{-\frac{(\mu_{prior}-\mu)^2}{2\sigma^2_{prior}}\right\}$ i.e. $Normal(\mu_{prior} = 0, \sigma_{prior} = 1)$
- likelihood is $P(x_i|\mu_{likelihd}) = \prod_{i=1}^{500} \frac{1}{\sqrt{2\pi\sigma^2_{likelihd}}} \exp\left\{-\frac{(x_i-\mu_{likelihd})^2}{2\sigma^2_{likelihd}}\right\}$ i.e. $Normal(\mu_{likelihd}, \sigma_{likelihd} = 1)$
- the analytic posterior is then $Normal(\mu = \frac{\sigma^2_{\text{prior}}}{\sigma^2_{data}+\sigma^2_{\text{prior}}}\mu_{\text{data}}, \sigma = \left(\frac{1}{\sigma^2_{\text{prior}}} + \frac{1}{\sigma^2_{data}}\right)^{-1})$

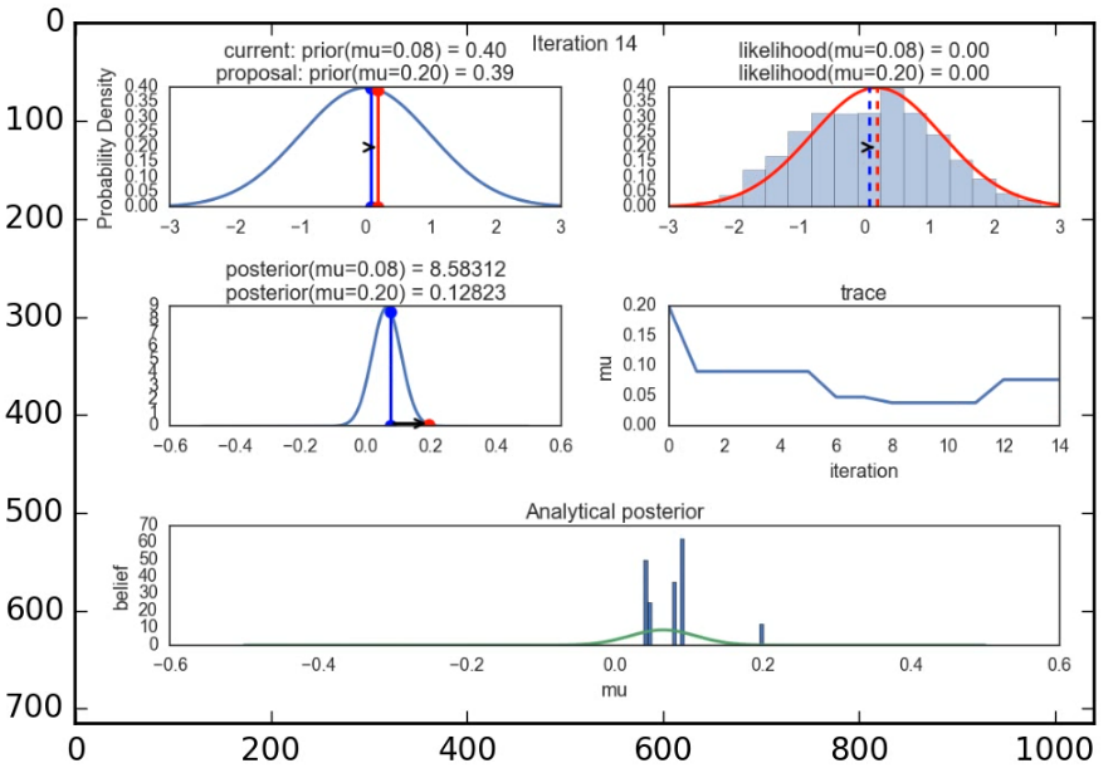
Example on normal distribution with fixed variance

- data: g
- prior is
  Norma
- likeliho
  Norma
- the an
  Norma

= 1) (500)

$)^2$ } i.e.

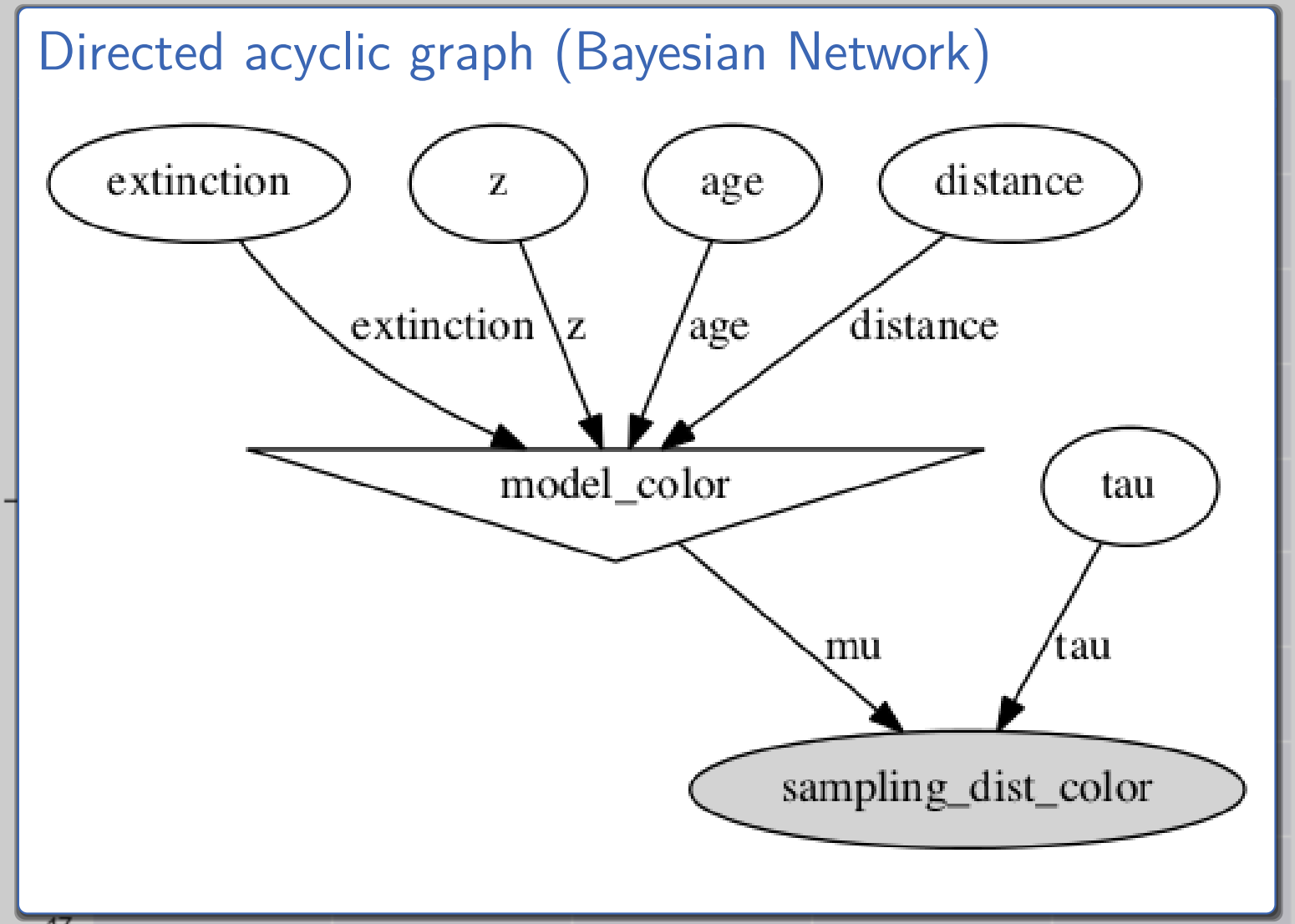
Confused by the ugly math? right, Movie time!



Star Cluster Isochrone Fitting: open cluster NGC2168

Star Cluster Isochrone Fitting: open cluster NGC2168

Directed acyclic graph (Bayesian Network)



Conclusion

- MCMC is a powerful sampling technique in Bayesian inference, it can sample high dimension posterior distribution that would otherwise be impossible to sample from
- 1> construct a Markov Chain and allowed it to reach its stationary distribution, 2> generated samples from that stationary distribution (the samples draw from stationary distribution would appear as if we draw from our posterior distribution as $t \to \infty$), 3> making any statistics as you wish.
- Metropolis-Hastings algorithm is the most popular MCMC method, and it has good theoretical guarantee for the convergence.
- MCMC is only one of the method for bayesian inference, there are plenty of alternatives e.g. MultiNest method (Cambridge) – LuLee