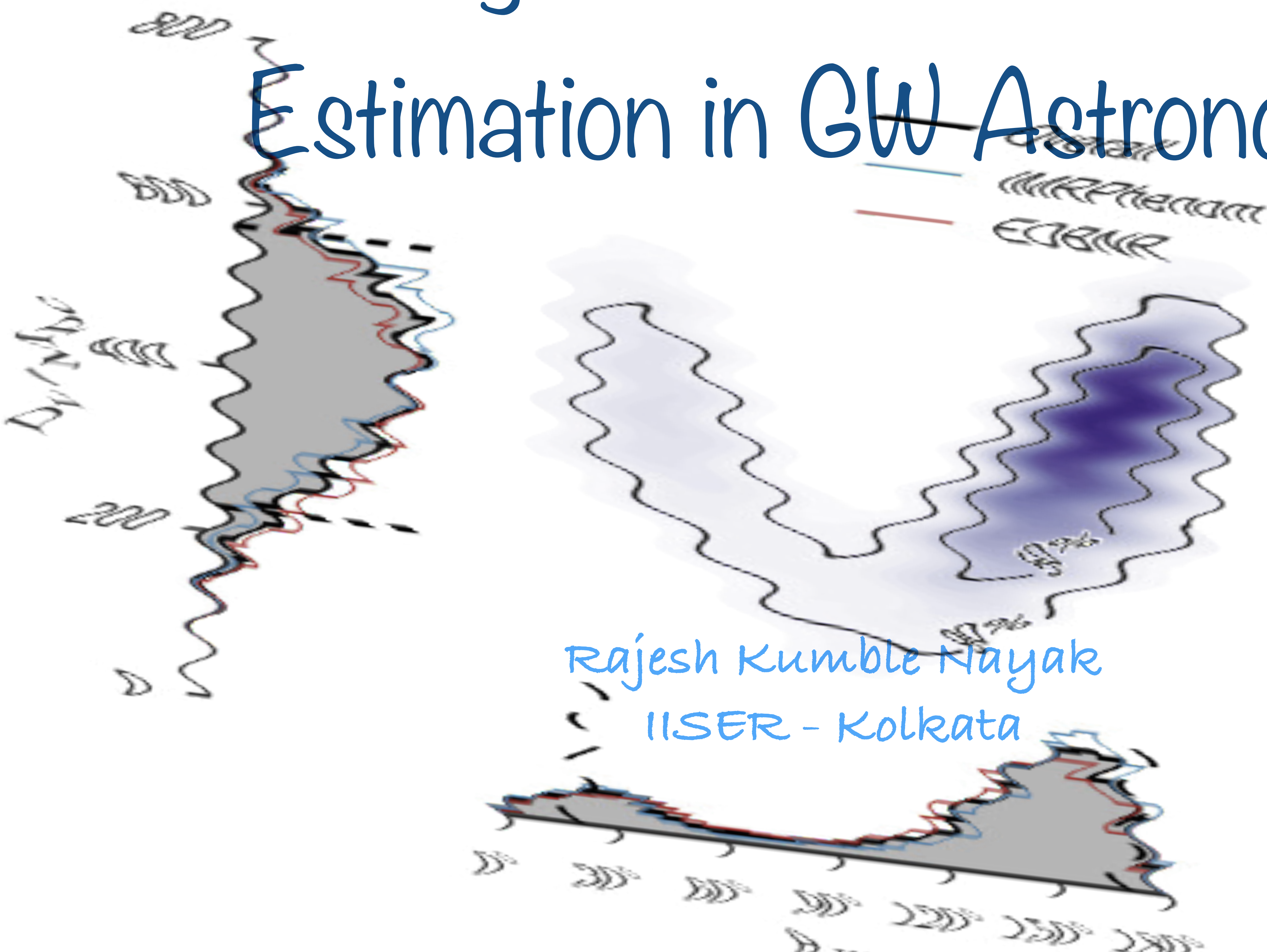
Challenges in Source Parameter Estimation in GW Astronomy



Rajesh Kumble Nayak
IISER - Kolkata

ICTS-SAMSI Workshop

Plan of the talk

- Introduction
- Detection Vs Parameter Estimation
- Parameter Estimation and Gravitational Wave Astronomy
- MCMC, etc
- Ergodic Theorem for MCMC
- Beyond Markov Chain

The Dawn of Gravitational Wave Astronomy

Post detection, Gravitational Wave Astronomy is entering into a new era

Because of weak signal buried in the noise, the Statistical Data Analysis plays an important role in GW Astronomy

Astronomy/Astrophysics part comes from the source modelling and parameter estimation

Detection Problem

Data recorded at the detector may be considered as Radom with certain statistical properties which changes in presence of signal.

And detection problem reduces to ability to distinguish between these two statistics!

Detection Probability	False Alaram Probability
$$P_D(R) = \int_R p_1(x)dx$$	$$P_F(R) = \int_R p_0(x)dx$$
PDF corresponding to signal being present	PDF corresponding to only noise

Detection Neyman-Pearson Approach

In this approach, fix a pre-determined false alarm prob. Say, α and maximise the detection prob.

And the solution is region R defined by,

$$R = [x : \Lambda(x) \geq k]$$

Where $\Lambda(x)$ is likelihood ratio defined as: $\Lambda(x) = \dfrac{p_1(x)}{p_0(x)}$

k is given by, $P_F\,[\Lambda(x) \geq k] = \alpha$

Given the nature of noise (Say Gaussian !) and the model of the signal one can construct the likelihood function which can be maximised for given α

Detection vs Parameter Estimation

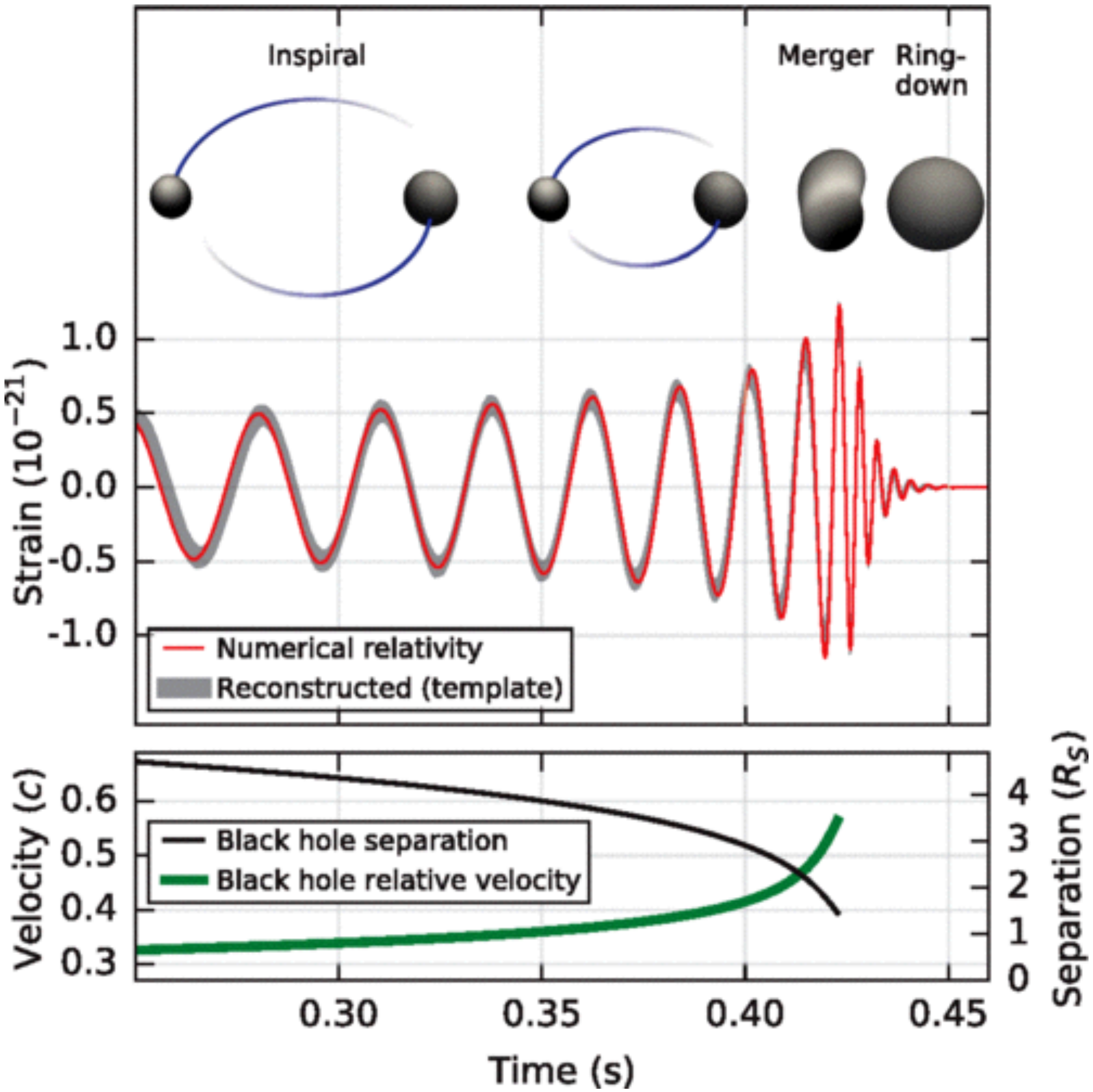
The likelihood function $\Lambda(x;\theta)$ depend on signal model and maximisation has to be performed over all source parameters θ

To reduce the computational cost and speed up the detection process, the parameter space is spanned efficiently, in other words as coarsely as possible. This done by template placement

This coarse sampling of likelihood function $\Lambda(x;\theta)$ not good enough for accurate parameter estimation

Signal Model

One of the promising source of GW signal is compact binary system



Parameter set $\theta = \left\{ m1, m2, \vec{S}_1, \vec{S}_2, t_a, \phi_0, \psi, \alpha, \delta, \theta_{J,N} \right\}$

Parameter Estimation Problem

A more precise estimation of source parameter is needed form Astrophysics point of view.

In maximum likelihood estimator, the source parameters θ_0 which maximises the likelihood function gives the source parameters

$$\left.\frac{\partial \Lambda[x;\theta]}{\partial \theta_k}\right|_{\theta_0} = 0$$

The error/the accuracy is represented by Fisher information matrix

$$\Gamma_{ik} = E\left[\frac{\partial^2 \Lambda[x;\theta]}{\partial \theta_i \partial \theta_k}\right]_{\theta_0}$$

Parameter Estimation Bayesian Approach

Here, Bayes theorem is the basis for computing posterior prob. distribution. According to the Bayes theorem

$$p\left(\theta|d,H\right) = \frac{p(\theta|H)p(d|\theta,H)}{p(d|H)}$$

$p\left(\theta|d,H\right)$ Posterior prob. function -> The joint probability distribution on the multidimensional space describes the collective knowledge about all parameters as well as their relationships.

$p\left(\theta|H\right)$ Prior distribution, i.e. prior knowledge or ignorance about the problem

$p\left(d|\theta,H\right)$ Likelihood function, for the hypothesis

Results for a specific parameter are found by marginalising over the unwanted parameters

$$p(\theta_1|d, H) = \int \mathrm{d}\theta_2 \ldots \mathrm{d}\theta_N \, p(\boldsymbol{\theta}|d, H)$$

Fully marginalised likelihood, called evidence useful for comparing different Hypothesis

$$Z = p(d|H) = \int \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_N \, p(d|\boldsymbol{\theta}, H) \, p(\boldsymbol{\theta}|H).$$

Effectively sampling parameter space, especially in higher dimensional space.

J. Veitch, PHYSICAL REVIEW D **91,** 042003 (2015)

Monte Carlo Methods

Stochastic samplers are best bets for sampling multidimensional space. (Monte Carlo methods)

Markov Chain Monte Carlo(MCMC) samplers powerful and promising samplers.

One of the top 10 Algorithms of last century!

Success of the Bayesian methods are closely linked to success of samplers like MCMC

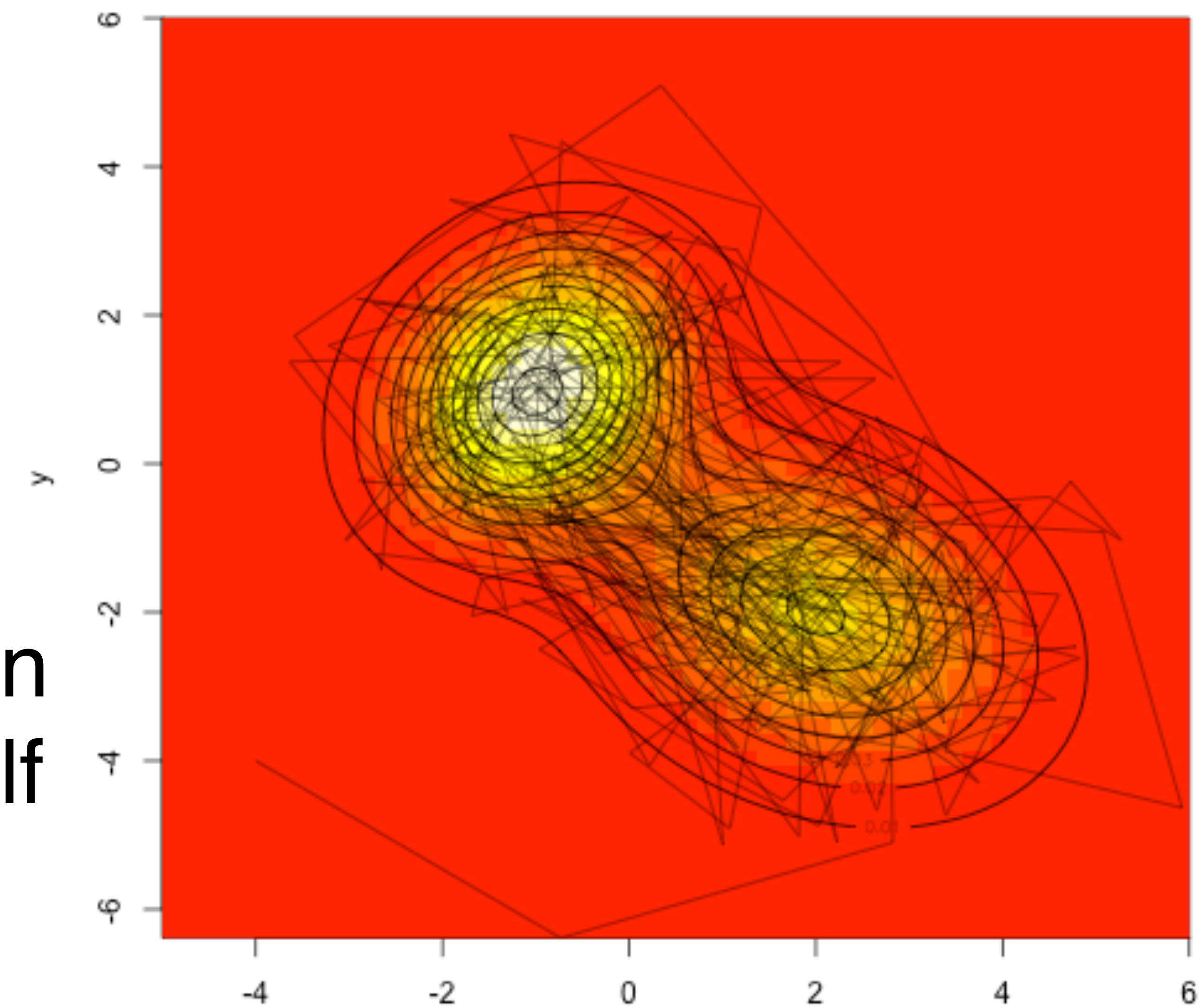
PARAMOUNT PICTURES PRESENTS
KEN ANNAKIN'S PRODUCTION OF
MONTE CARLO OR BUST!
U
A RACE FOR GLORY, FOR LOVE AND FOR THE FUN OF IT!
STARRING (IN ALPHABETICAL ORDER)
BOURVIL • LANDO BUZZANCA • WALTER CHIARI • PETER COOK • TONY CURTIS • MIREILLE DARC • MARIE DUBOIS • GERT FROBE
SUSAN HAMPSHIRE • JACK HAWKINS • NICOLETTA MACHIAVELLI • DUDLEY MOORE • PEER SCHMIDT • ERIC SYKES • TERRY-THOMAS
PRODUCED AND DIRECTED BY KEN ANNAKIN • ORIGINAL STORY AND SCREENPLAY BY JACK DAVIES AND KEN ANNAKIN • ASSOCIATE PRODUCER BASIL KEYS
TITLE SONG, MUSIC AND LYRIC BY RON GOODWIN—SUNG BY JIMMY DURANTE, AND OTHER MUSIC FROM THE MOTION PICTURE SCORE AVAILABLE ON PARAMOUNT RECORDS
TECHNICOLOR® • PANAVISION • A PARAMOUNT PICTURE

MCMC 101

A Markov chain is a series of states of a system such that future state is conditionally independent of every prior state

Markov chain consists of n states and a transition probability matrix describe change

Final goal is to sample the function with distribution function similar/close to the function itself

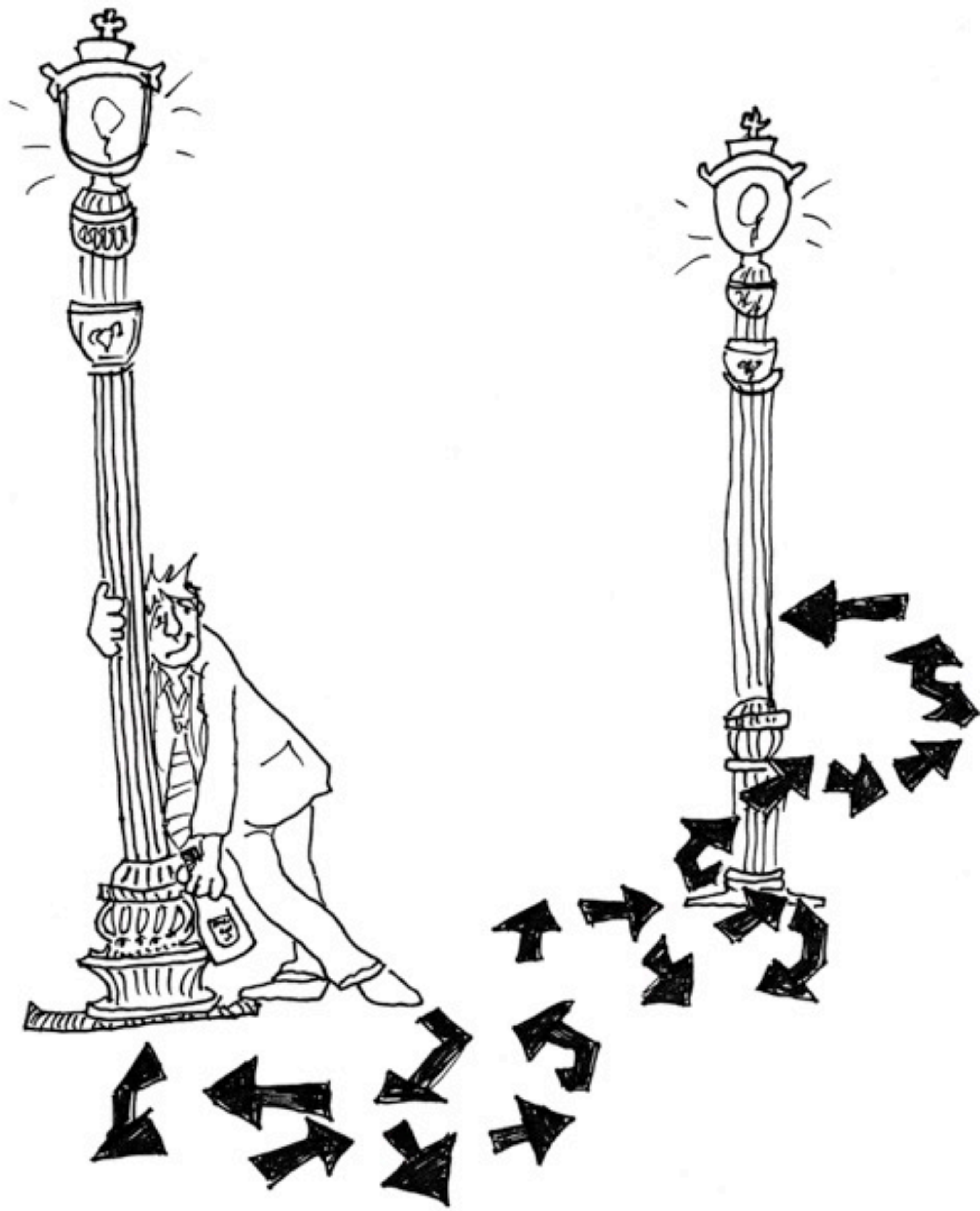


Convergence of MCMC

Convergence is important property.

Convergence means, as number of samples $N \to \infty$ distribution of samples should converge to desired distribution.

This ensures, can do better as we take more and more sample



Convergence of MCMC

Tierney, Annals of statistics, 22, 1701(1994)

It has been shown that under three conditions Markov Chain converge: irreducibility, aperiodicity and invariance

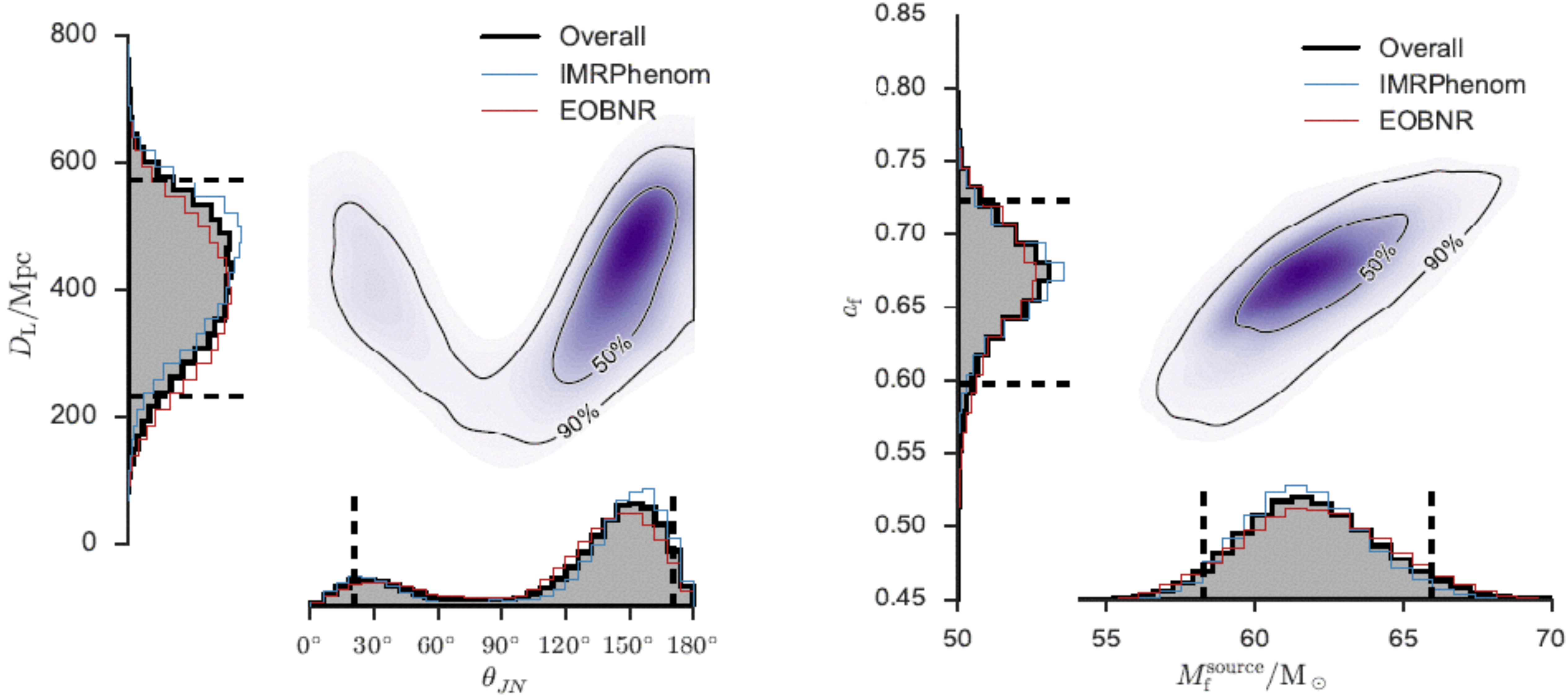
irreducibility: A Markov chain is said to be irreducible if it is possible to get to any state from any state.

periodicity: A state in a Markov chain is periodic if the chain can return to the state only at multiples of some integer larger than 1.

periodicity:Invariance refers to the property that if we start with a state vector generated from the desired posterior distribution, then a further transition in the chain leave the distribution invariant.

Lal Inference in LIGO

LIGO tool, Lal-Inference extensively uses MCMC sampling with very good results



Beyond Markovian Sampler

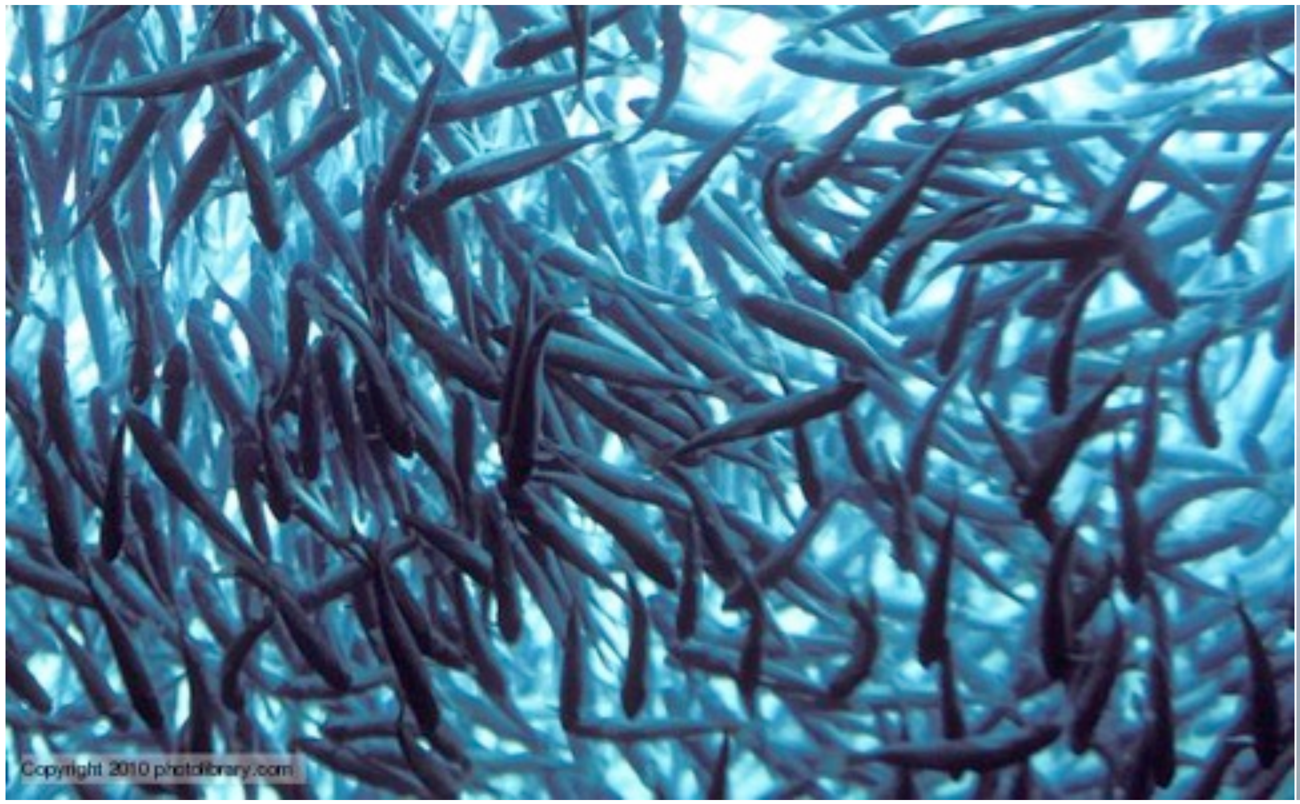
MCMC sampler is computationally intense. At least in gravitation wave astronomy, if more and more sources are detected, computation can be very challenging

As there is no theoretical studies such as Ergodic theorem (convergence), it is non-trivial to move away from MCMC or similar algorithms

It may be worth exploring!

Particle swarm optimisation(PSO)

Stochastic method Inspired by the Nature,



What is not achieved by one, is done in coordination with many. In the case of Biological system life depends on it!

Nature might have it’s own rules. PSO as simple rule by which N particles, communicate with each others, to take successive step towards the global maximum.

PSO Algorithm

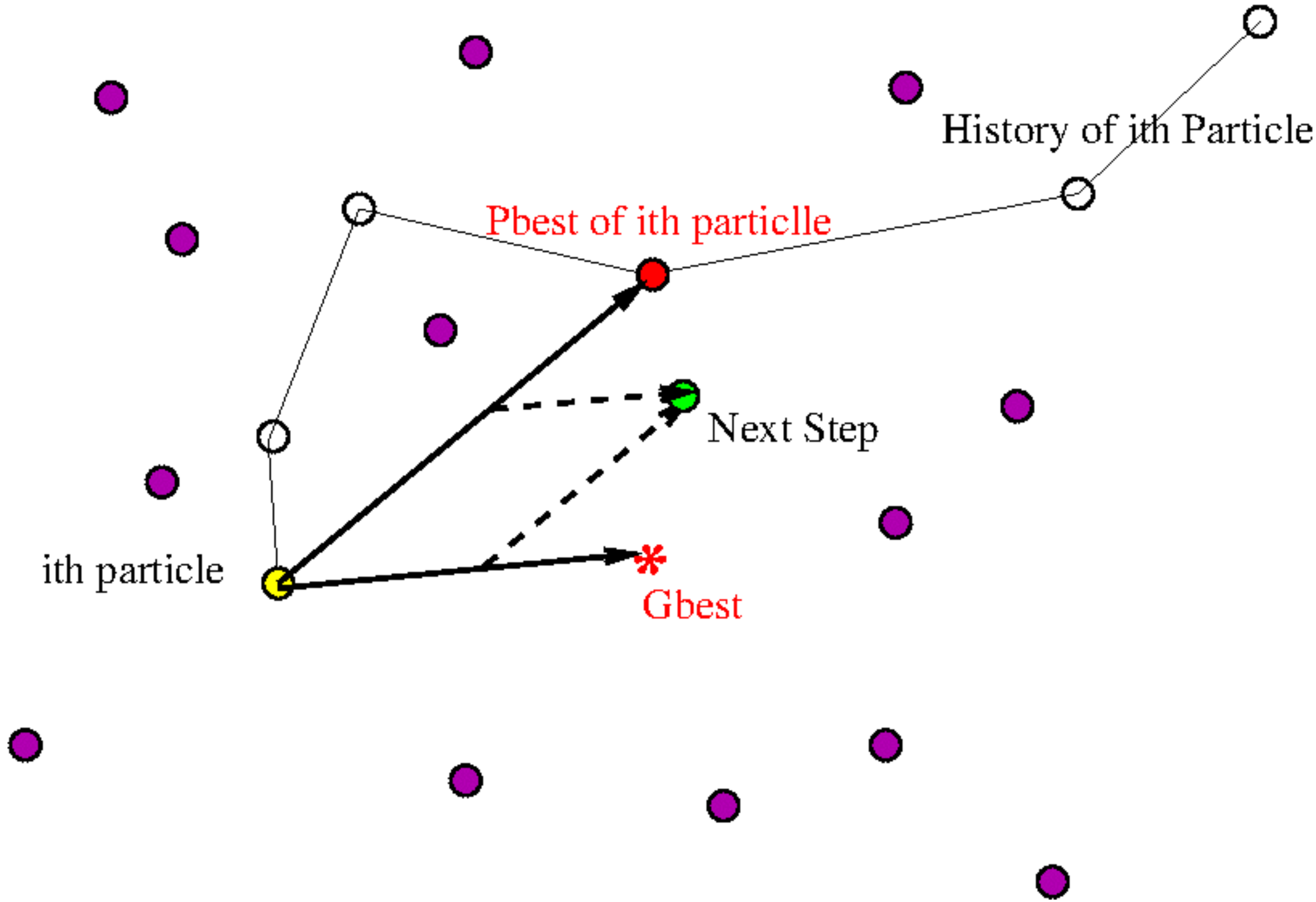
1. Let $f(\vec{X})$ function we want maximize, in M dimensional domain.
2. We start with N particles, Uniformly distributed in the domain.
3. Each particle start with a random velocity $\vec{V}_0$, which is also uniformly distributed.
4. Each particle keep track of it's own best location, namely $\vec{X}_{pbest}$.
5. $\vec{X}_{gbest}$ is the best location found by all particles.
6. Velocity for the next step is computed based on the rule

$$\vec{V}_{n+1} = w\vec{V}_n + c_1\chi_1\left(\vec{X}_{pbest} - \vec{X}\right) + c_2\chi_2\left(\vec{X}_{gbest} - \vec{X}\right)$$

   w, c_1 and c_2 are free parameters, used for tuning the walk. χ_1 and χ_2 are uniformly distributed random number.

7. One can stop when sufficient convergence is obtained.

PSO Algorithm

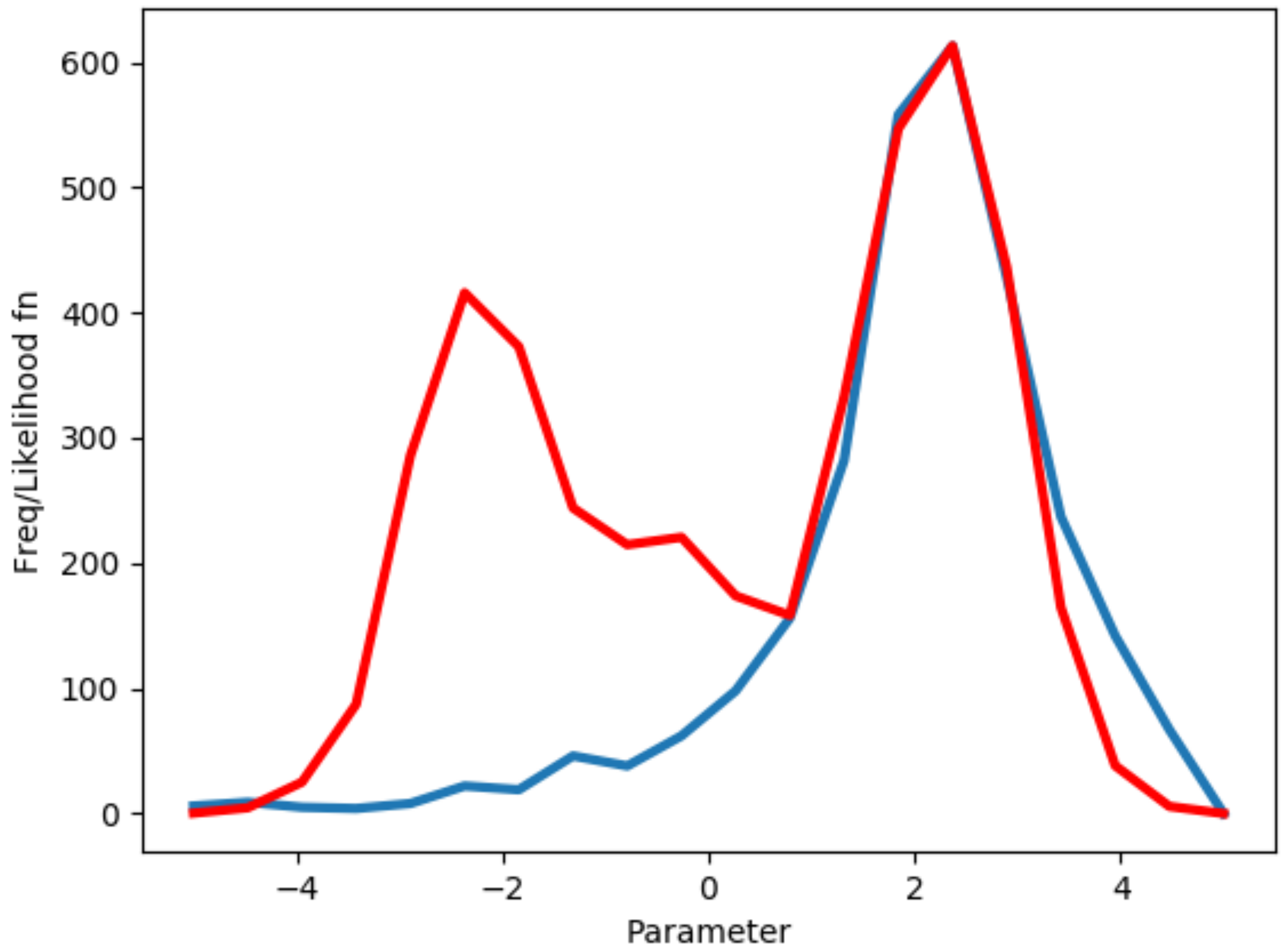


Simple results, 1-D Case

Problem is to sample the multimodal function using PSO sampler

Distribution of samples and function value

Run PSO with 100 Particles



Simple results

This can be extended to higher dimensions. PSO is effective in larger dimension too.

PSO ignores the local peaks, larger samples are taken at the global peak

Multidimensional integral using this samples.

However, distribution of samples does not match with the value of function except around global maxima