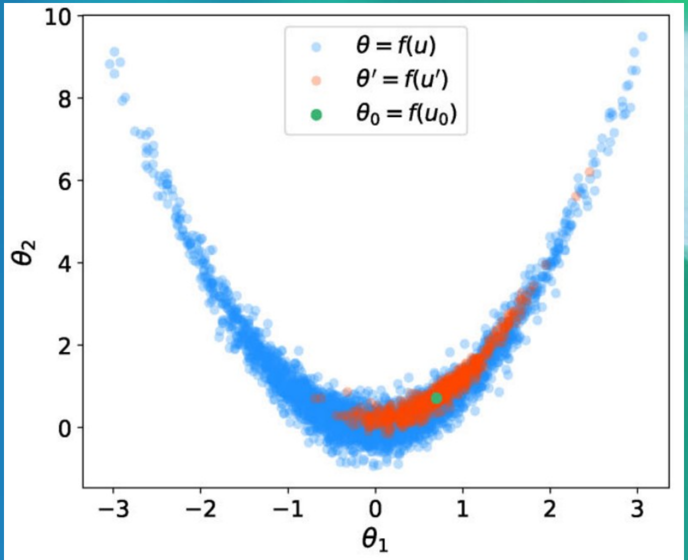
Accelerating astronomical and cosmological inference with preconditioned Monte Carlo
(2022, MNRAS, 516, 2, 1644-1653)

Minas Karamanis, Florian Beutler, John A. Peacock, David Nabergoj, Uros Seljak



Statistics Journal Club, OA UW, May 22nd, 2024

Probability density intergals from from a random sample

- Expected value of ***x*** is:

$$E(x)=\int x\, p(x)\, dx \quad \simeq \frac{1}{N}\sum_i x_i$$

If the points x_i are drawn according to the p(x) distribution

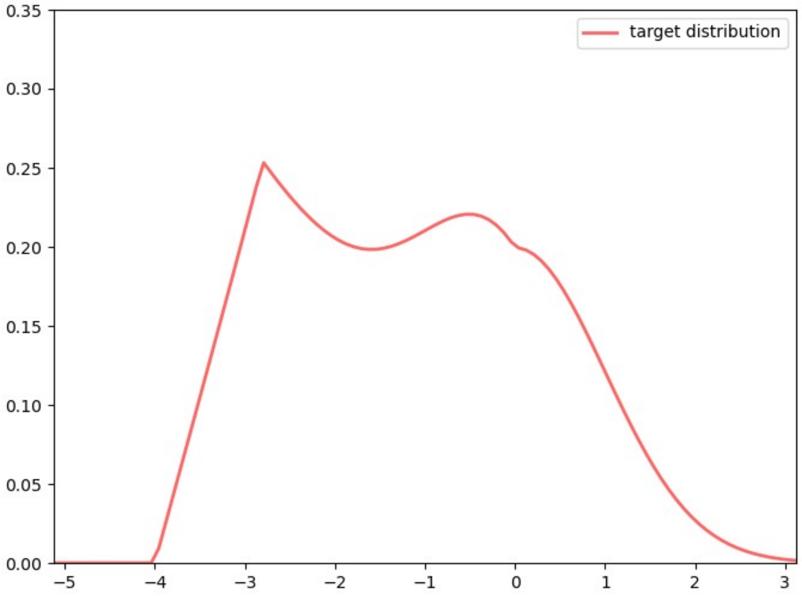
- Expected value of any function ***f*** of ***x*** is also easy to calculate

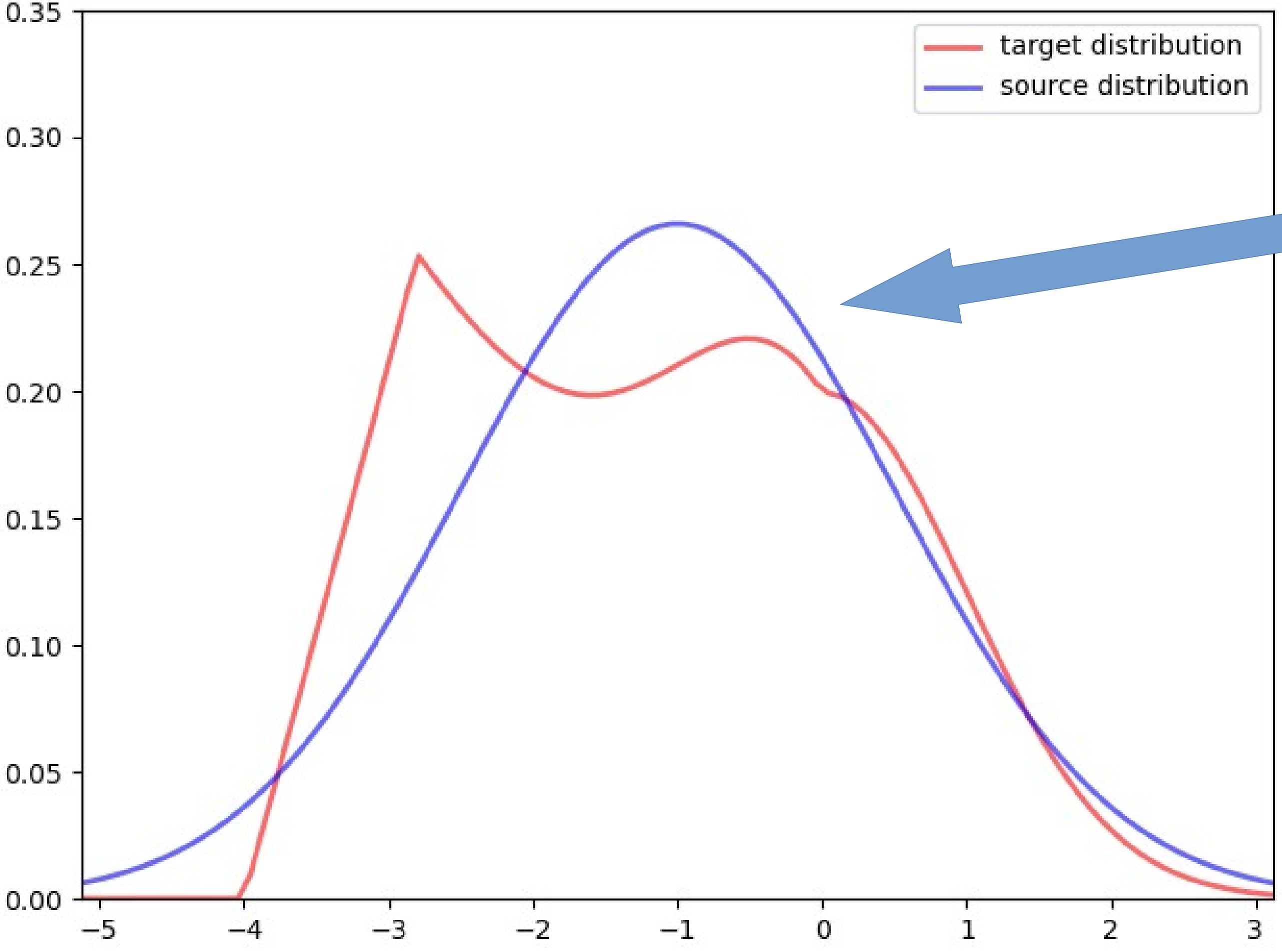
$$E(f)=\int f(x)\, p(x)\, dx \quad \simeq \frac{1}{N}\sum_i f(x_i)$$

If the points x_i are drawn according to the p(x) distribution

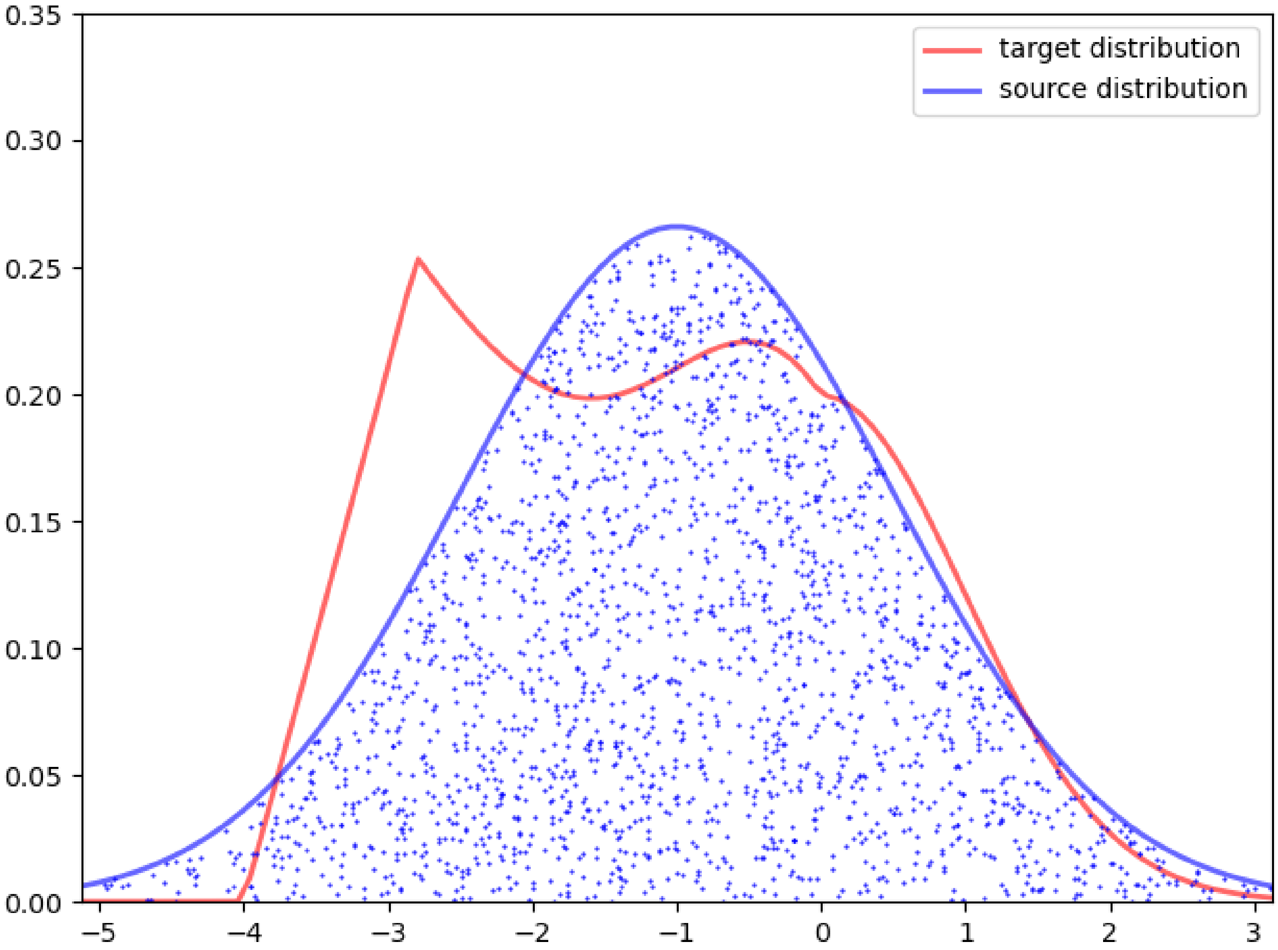
Importance sampling

Problem:
p(x) is a probability density, and we need a random sample from it,
For example, to calculate something for it (expected value, percentiles, or any integral)

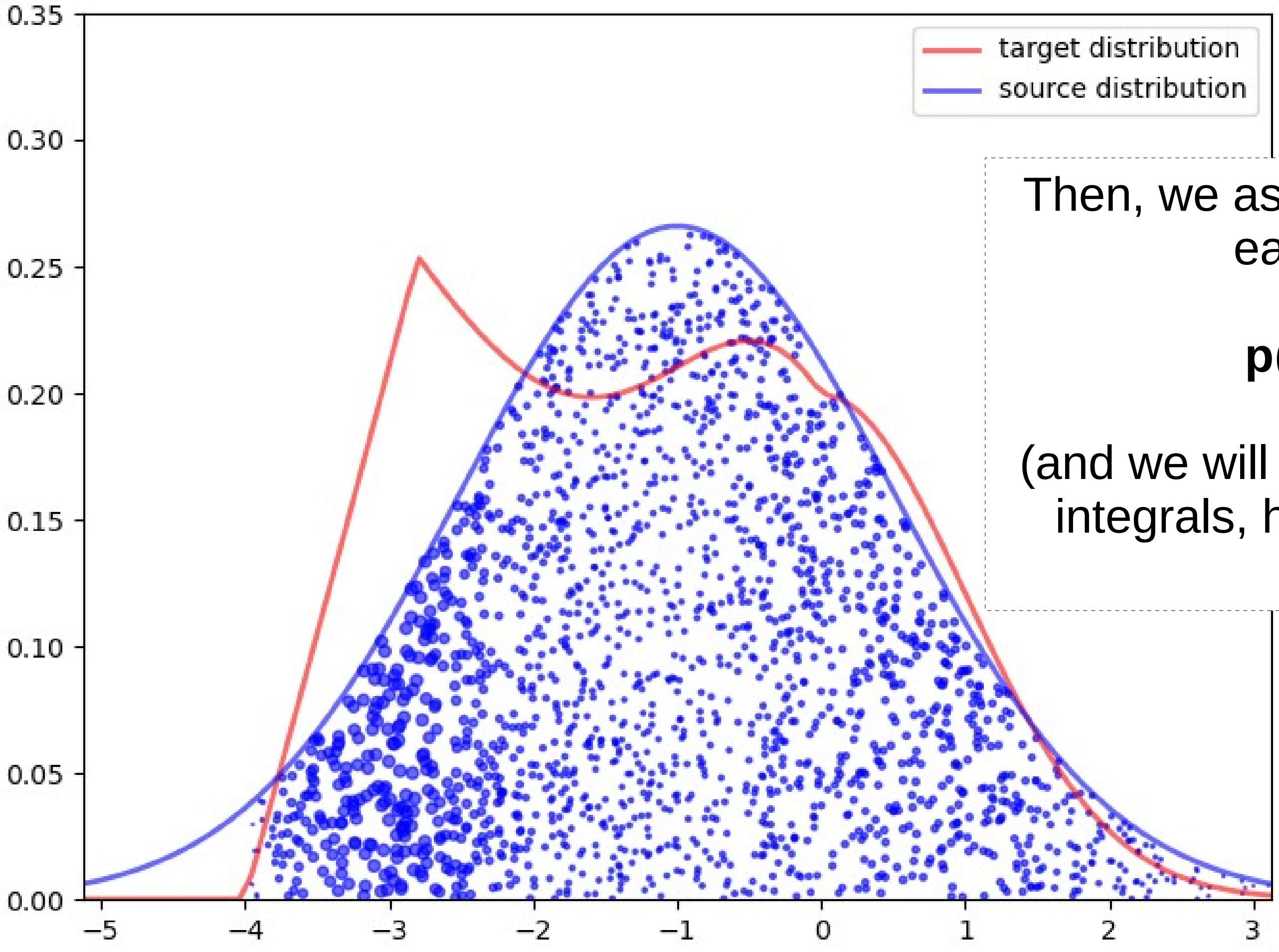




We choose any „similar" function, from which it is easy to get a random sample from:
ϱ(x)



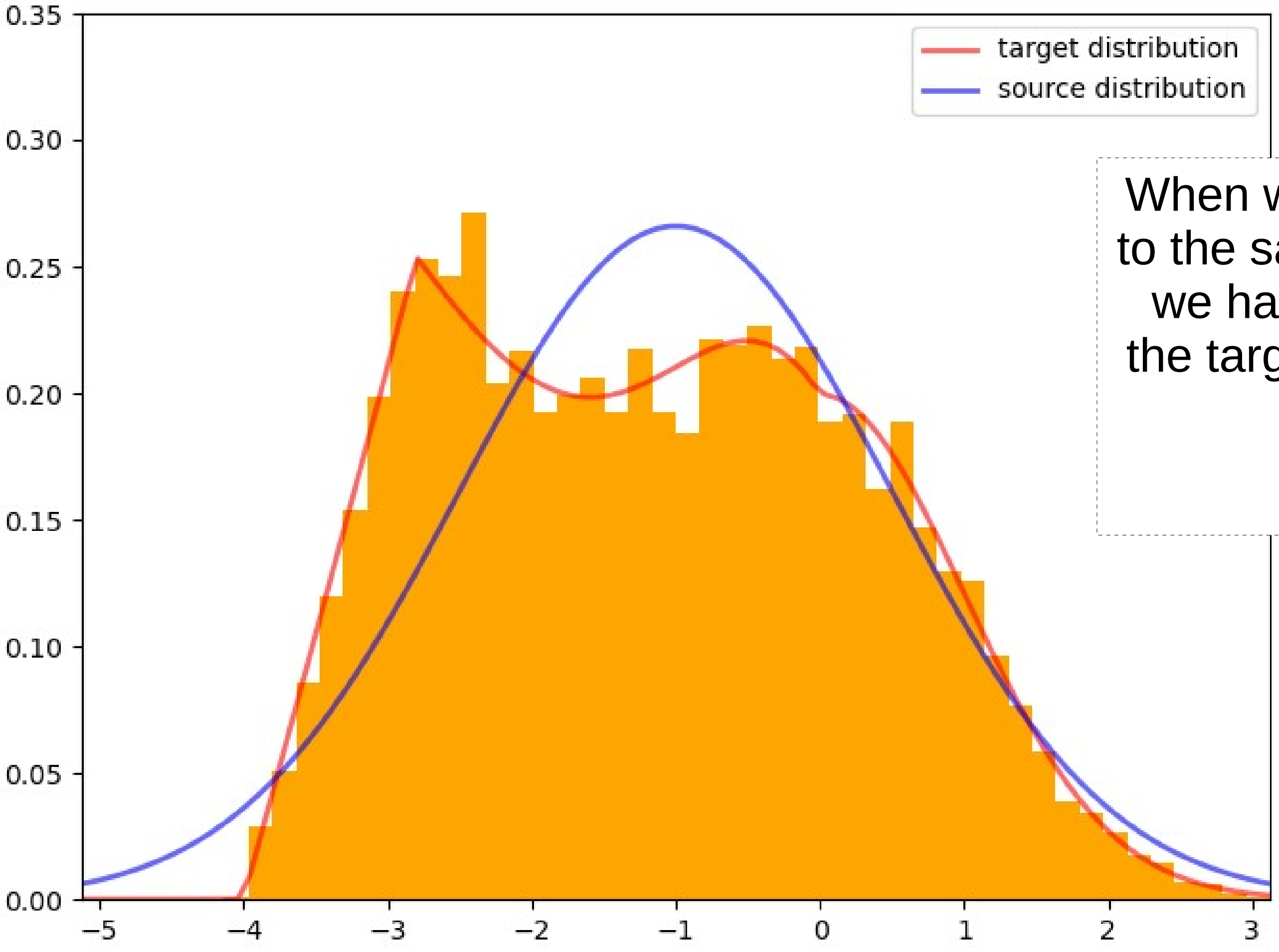
We draw a sample from the distribution **ρ**



Then, we assign „**weights**" for each point

p(x) / **ϱ**(x)

(and we will use this weight for integrals, historgrams, etc.)



When we add the weights to the sample, it looks as if we have a sample from the target **p**(x) distribution

Probability density integrals in importance sampling

- Expected value of ***x*** is:

$$E(x)=\int x\,p(x)\,dx=\int x\frac{p(x)}{\rho(x)}\rho(x)\,dx=\int [x\,w(x)]\rho(x)\,dx$$

$$\simeq\frac{1}{N}\sum_i [x_i w(x_i)]=\frac{1}{N}\sum_i [x_i w_i],\ \text{where}\ w(x)=\frac{p(x)}{\rho(x)}$$

- Expected value of any function ***f*** of ***θ*** is also easy to calculate

normalization

$$\mathrm{E}_p[f(\theta)] = \int f(\theta)w(\theta)\rho(\theta)\mathrm{d}\theta \Big/ \int w(\theta)\rho(\theta)\mathrm{d}\theta$$

$$\rightarrow=\frac{1}{N}\sum_i [f(\theta_i)w(\theta_i)]$$

If the points θ_i are drawn according to the p(θ) distribution

Notation

Bayes formula:

$$P(\theta \mid D, M) = \frac{P(D \mid \theta, M)\, P(\theta \mid M)}{P(D \mid M)}$$

likelihood prior

posterior $$\mathcal{P}(\theta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{\mathcal{Z}}$$

evidence

We could use importance sampling to sample form the posterior

01 Draw sample from: $\pi(\theta)$

02 We need sample from: $\mathcal{L}(\theta)\pi(\theta)$

03 So, we weight sample with:

$$w(\theta) = \mathcal{L}(\theta)\pi(\theta) / \pi(\theta) = \mathcal{L}(\theta)$$

04 This is very inefficient, if likelihood is significantly different from the prior

Sequential Monte Carlo (SMC)

It is similar to the importance sampling, but in couple of stages

$$p_1(\theta) = \pi(\theta)$$

$$p_T(\theta) = \pi(\theta)\mathcal{L}(\theta)$$

$$p_t(\theta) \propto \pi(\theta)\mathcal{L}(\theta)^{\beta_t}, \quad t = 1, \ldots, T$$

$$\beta_1 = 0 < \beta_2 < \ldots < \beta_T = 1$$

Sequential Monte Carlo (SMC)

Weight at every step:

$$w_t(\theta_t) = p_t(\theta_{t-1}) / p_{t-1}(\theta_{t-1})$$

So, the total, combined weight is:

$$w = w_1 w_2 \ldots w_T = \frac{p_2}{p_1} \frac{p_3}{p_2} \ldots \frac{p_T}{p_{T-1}} = \frac{p_T}{p_1}$$

We need some reshuffling / resampling of the sample

Step 1

(i) *Mutation*. The population of particles is moved from $\{\theta_{t-1}^k\}_{k=1}^N$ to $\{\theta_t^k\}_{k=1}^N$ using a *Markov transition kernel* $K_t(\theta^{'}|\theta)$ that defines the next important sampling density

$$p_t(\theta') = \int p_{t-1}(\theta)K_t(\theta'|\theta)\mathrm{d}\theta. \tag{4}$$

In practice, this step consists of running multiple short MCMC chains (i.e. one for each particle) to get the new states $\theta^{'}$ starting from the old ones θ.

Step 2

(ii) *Correction*. The particles are reweighted according to the next density in the sequence. This step consists of multiplying the current weight W_t^k of each particle by the appropriate importance weight

$$w_t(\theta_t) = p_t(\theta_{t-1})/p_{t-1}(\theta_{t-1}). \quad (5)$$

Step 3

(iii) *Selection*. The particles are resampled according to their weights, which are then set to $1/N$. This can be done using *multinomial resampling* or more advanced schemes. The purpose of this step is to eliminate particles with low weight and multiply the ones with high weights.

Evidence

- We start from a sample from the prior and set the evidence to $\mathcal{Z}=1$
- Evidence at every step can be updated, according to the sum of the weights (at every stage)

$$\mathcal{Z}_t / \mathcal{Z}_{t-1} = \sum_{k=1}^{N} W_{t-1}^{k} w_t(\theta_{t-1}^{k})$$

Metropolis-Hastings

“standard algorithm”

- proposal step
- acceptance step

How we sample effectively?

- By using proposal distribution “close” to the target distribution

Alternative way

Use simple proposal distribution, but in changed parametrization

$$\mathcal{N}(u, 1) \quad \leftarrow \text{proposal}$$

where

$$u = f(\theta)$$

“Preconditioning”

The authors approach

- Density estimation algorithm (from the current sample)
- Neural networks
  - Encoders (sample → Gaussian)
  - Special network that is a *bijective* mapping, so:
    - sample ↔ Gaussian
- Sample in multidimensional Gaussian
  - easy and control-able and short!

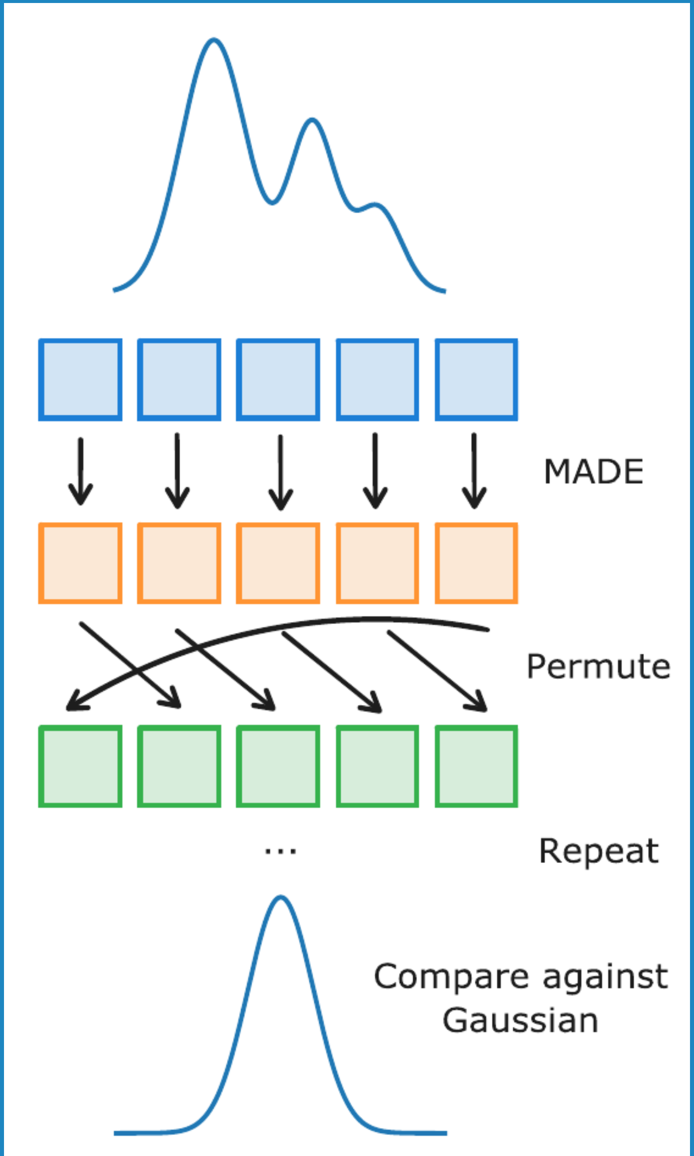
Metropolis-Hastings

Optimal proposal distribution for Gaussian target distribution

$$\mathcal{N}\left(\theta, 2.38^2 \Sigma / ndim\right)$$

where Σ is covariance matrix of the Gaussian

And then the acceptance rate is 23.4%



"Normalizing flows"

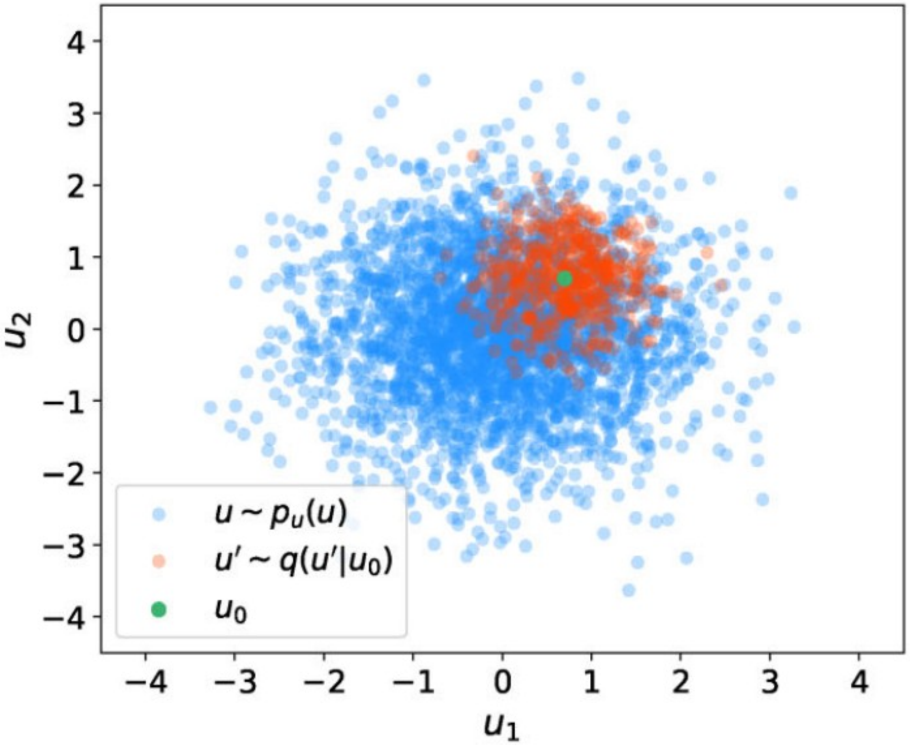
- uses: Masked Autoregressive Density Estimator (MADE)
- maps any distribution into multidimensional Gaussian

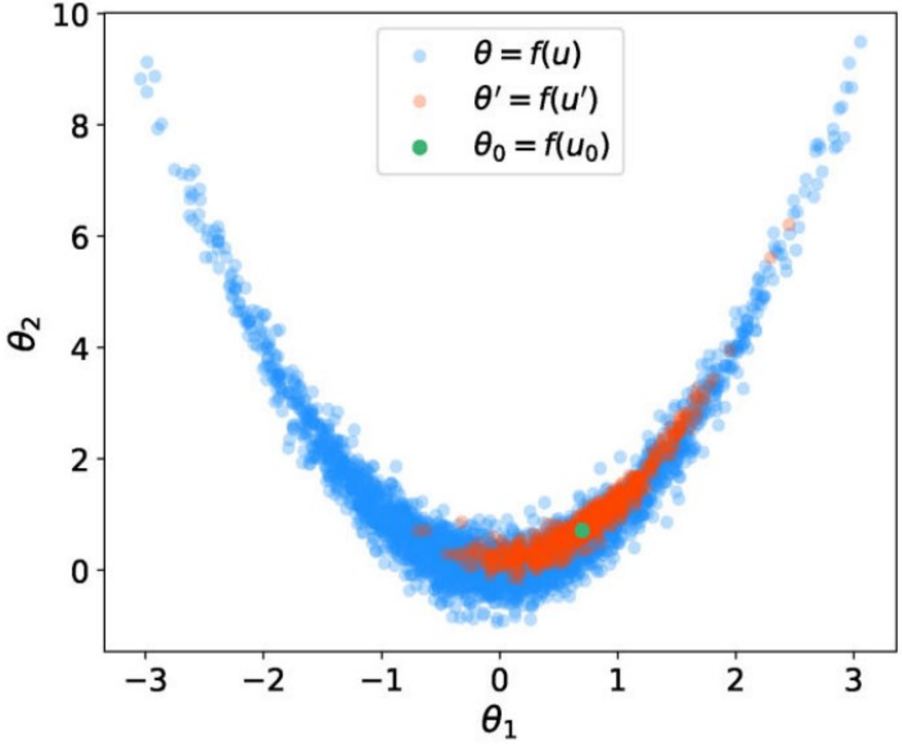
MCMC in the changed (preconditioned) coordinates

$$\alpha = \min\left(1, \frac{p_\theta(\theta')q(\theta|\theta')}{p_\theta(\theta)q(\theta'|\theta)}\right)$$

$$\alpha = \min\left(1, \frac{p_\theta(f^{-1}(u'))q(u|u')\left|\det \frac{\partial f^{-1}(u')}{\partial u'}\right|}{p_\theta(f^{-1}(u))q(u'|u)\left|\det \frac{\partial f^{-1}(u)}{\partial u}\right|}\right)$$

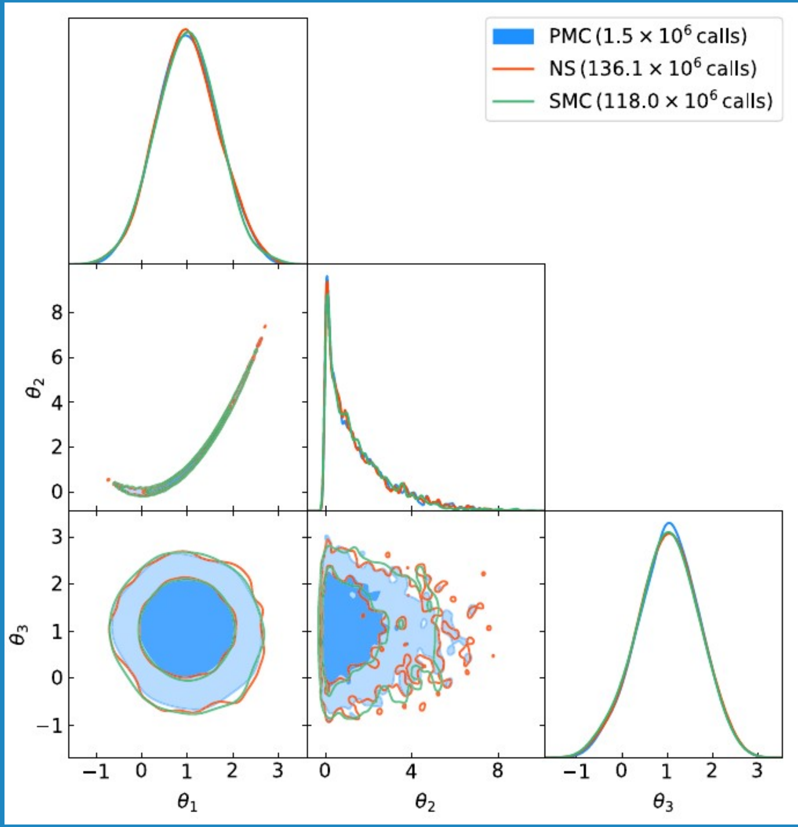
Proposal distribution in preconditioned *vs* regular coordinates



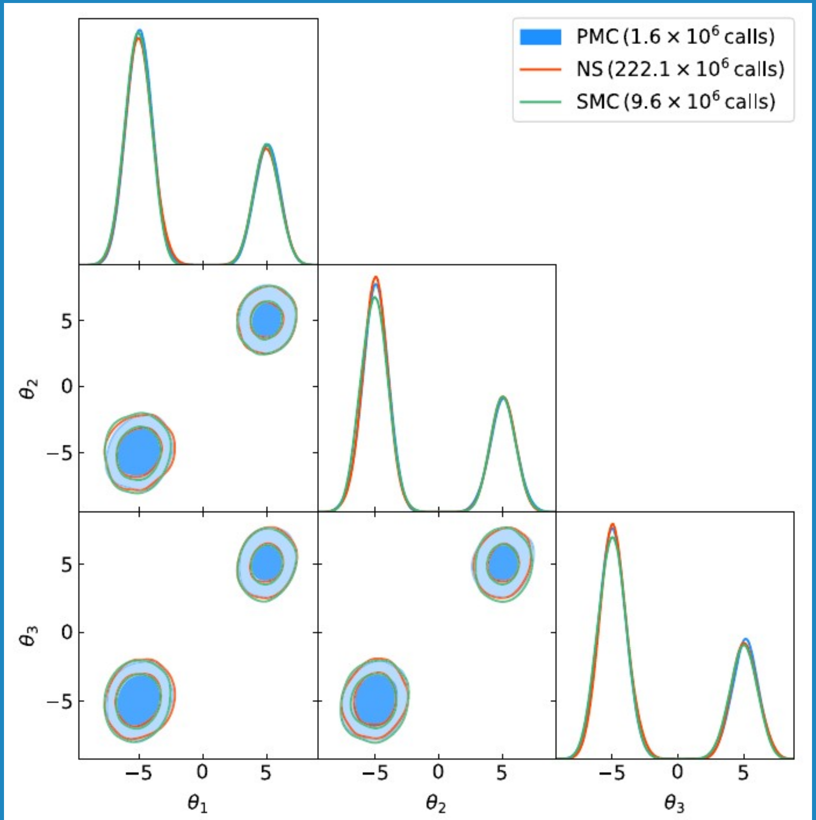


Examples

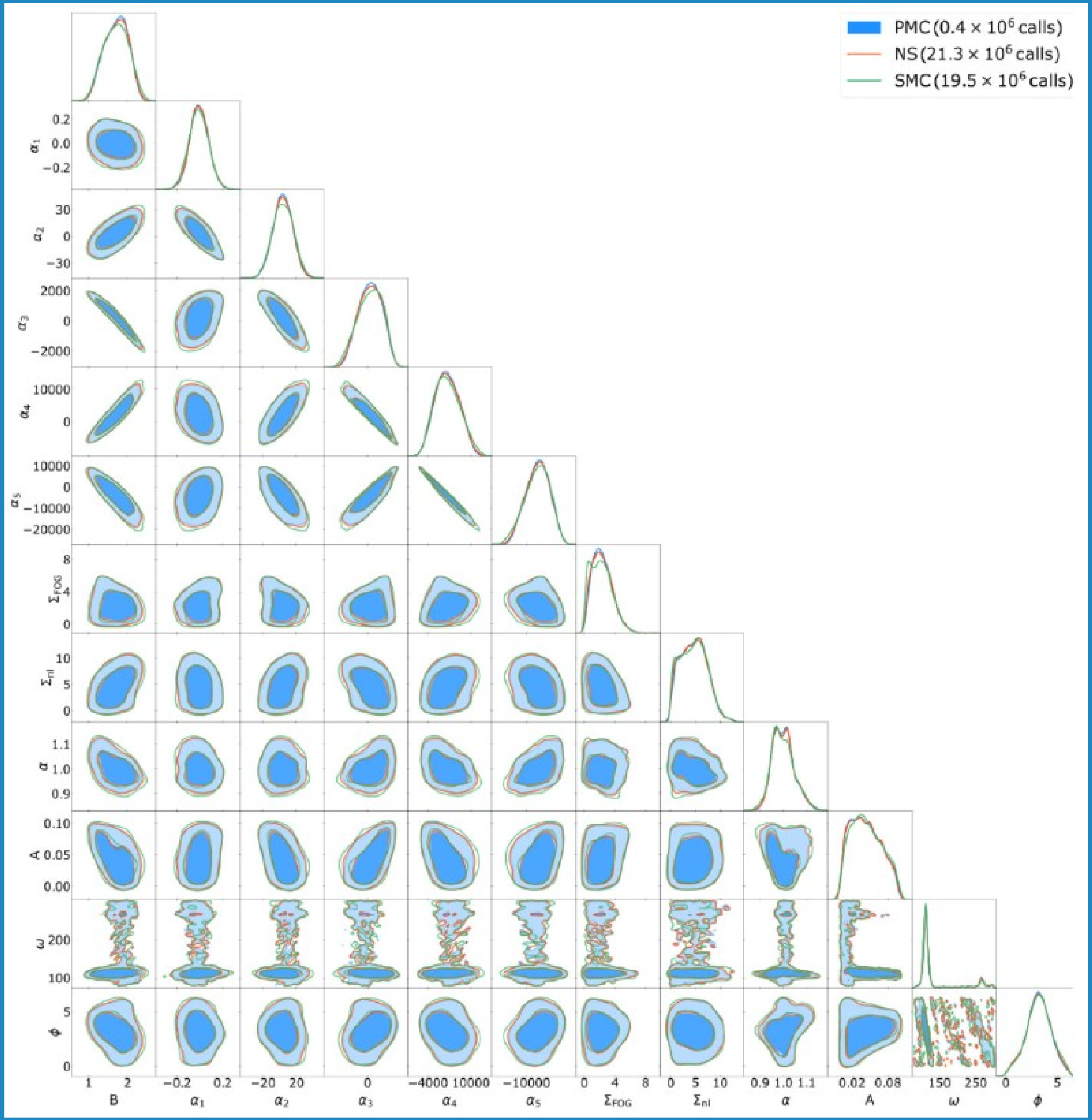
20-dim generalization of the Rosenbrock distribution



50-dim Gaussian mixture

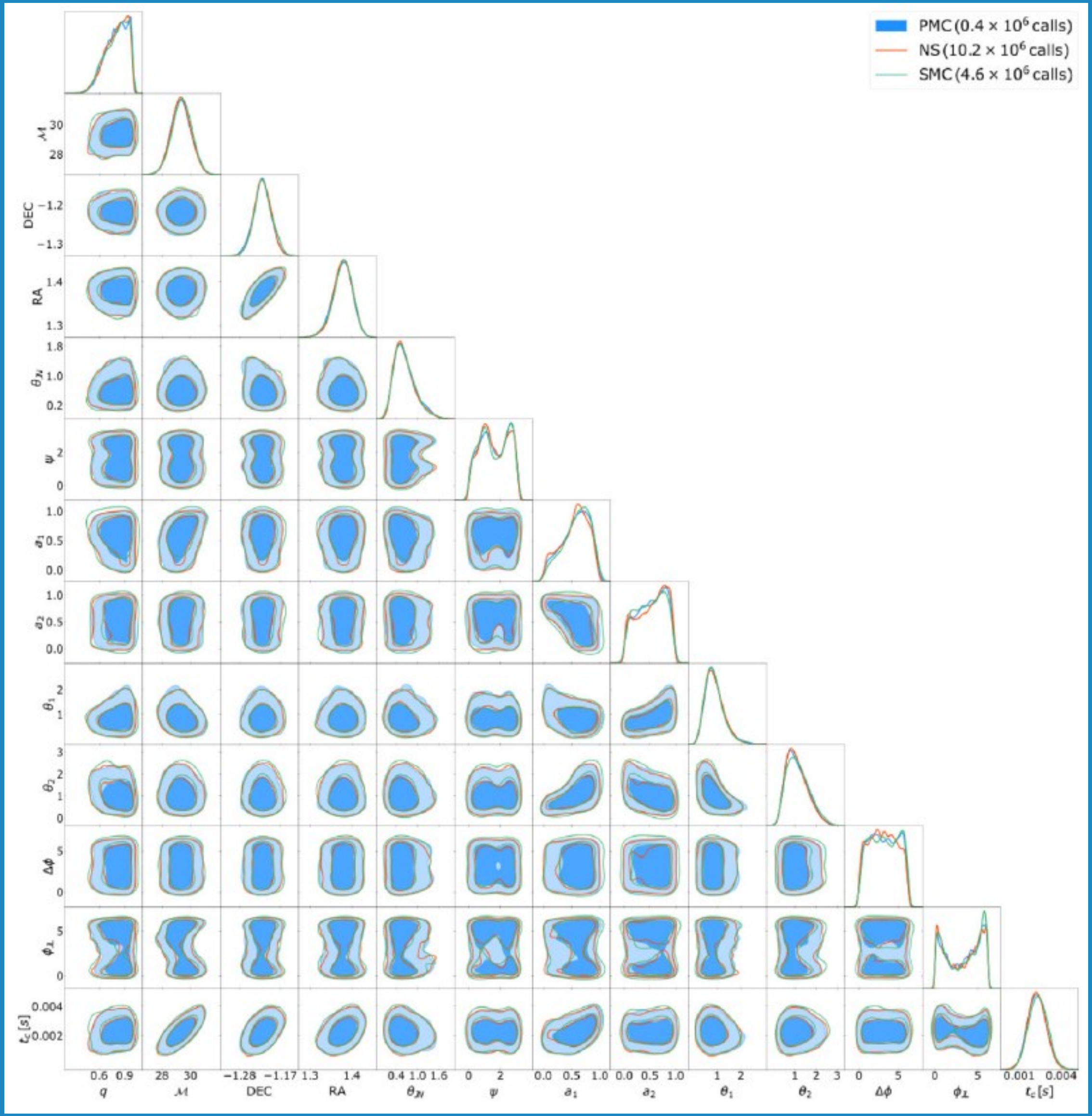


Examples



12-dim Baryon Accoustic Oscillation features in SDSS galaxies

Examples



13-dim gravitational wave signal from the compact binary coalescence

Speed

Distribution	Model evaluations ($\times 10^6$)		
	PMC	NS	SMC
Rosenbrock	**1.5**	136.1	118.0
Gaussian mixture	**1.6**	222.1	9.6
Primordial features	**0.4**	21.3	19.5
Gravitational waves	**0.4**	10.2	4.6

Thank you!