MA8702 – Paper review

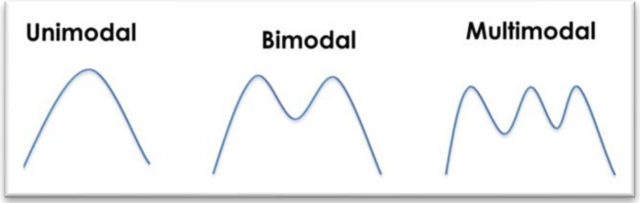
The Hastings algorithm at fifty

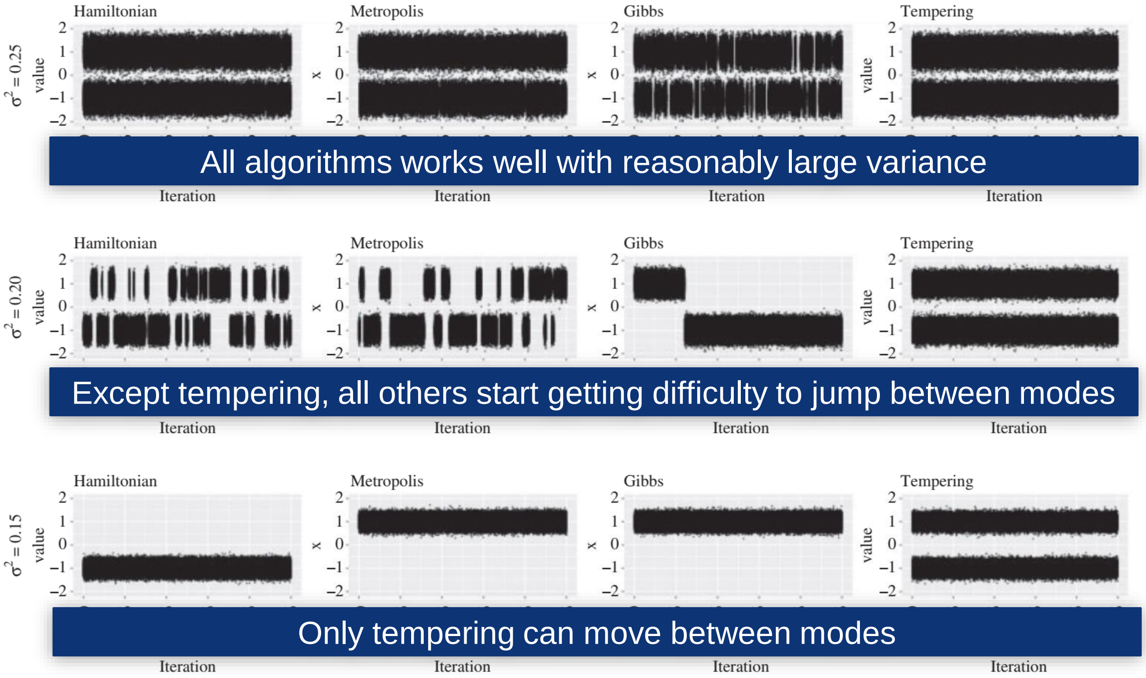
D. B. Dunson and J. E. Johndrow

01.2024

Norwegian University of Science and Technology

3. Challenging – Multimodal targets





All algorithms works well with reasonably large variance

Except tempering, all others start getting difficulty to jump between modes

Only tempering can move between modes

Solutions for multimodal:

- simulated
- tempering
- equi-energy sampler
- split–merge samplers
- birth–death algorithm

Hasting algorithm can fail to move among modes of the multimodal distribution.

3. Challenging – Intractable likelihoods

Example of intractable likelihoods:

- g-and-k distribution:

$$Q(u;A,B,g,k) = A + B\left[1 + c\frac{1-\exp\{-g\Phi(u)\}}{1+\exp\{-g\Phi(u)\}}\right]\{1+\Phi(u)^2\}^k\Phi(u)$$

Q(u;A,B,g,k)=A + B\left[1+c\dfrac{1-\exp\{-g\Phi(u)\}}{1+\exp\{-g\Phi(u)\}}\right]\{1+\Phi(u)^2\}^k\Phi(u)

- Two-dimensional summary statistics:

$$x_1, \dots, x_n \overset{\text{iid}}{\sim} \mathrm{N}(\theta, \sigma^2),$$
$$S(x_1, \dots, x_n) = (\mathrm{med}(x_1, \dots, x_n), \mathrm{mad}(x_1, \dots, x_n)),$$

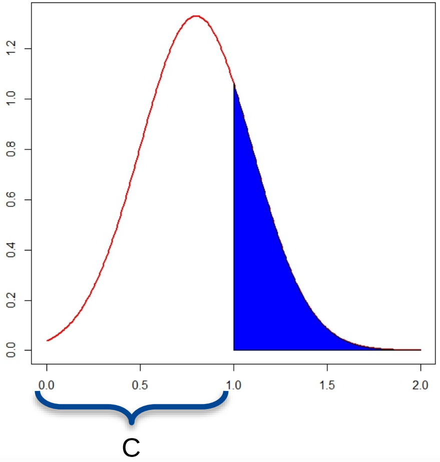
Solutions for incomputable likelihoods:
- auxiliary variable scheme
- rejection sampling
- Pseudo-marginal Metropolis Hastings

=> We need unbiased estimate of the likelihood in the acceptance probability

Likelihood functions can be intractable, meaning it is not computable even up to a normalizing constant.

NTNU

3. Challenging – Distributions with constrained support



Solution for constrained support distribution:

- ignore the constraint and simply reject proposals falling outside of C;
- reparameterize to an unconstrained space before running the sampler
- Gibbs sampling with the conditional posterior distributions truncated to reflect the constraint

Hard to implement appropriate proposal distribution with the same support

NTNU

Bonus: The effect of proposal selection in MCMC

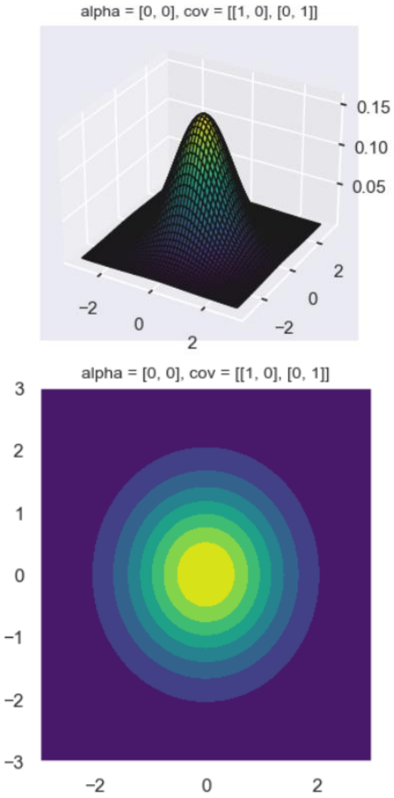
Given a target distribution,

- What is the difference we select one proposal over another?
- Is there a “better” proposal distribution over some others?

Case 1: Target and proposal are both Gaussian

Proposals

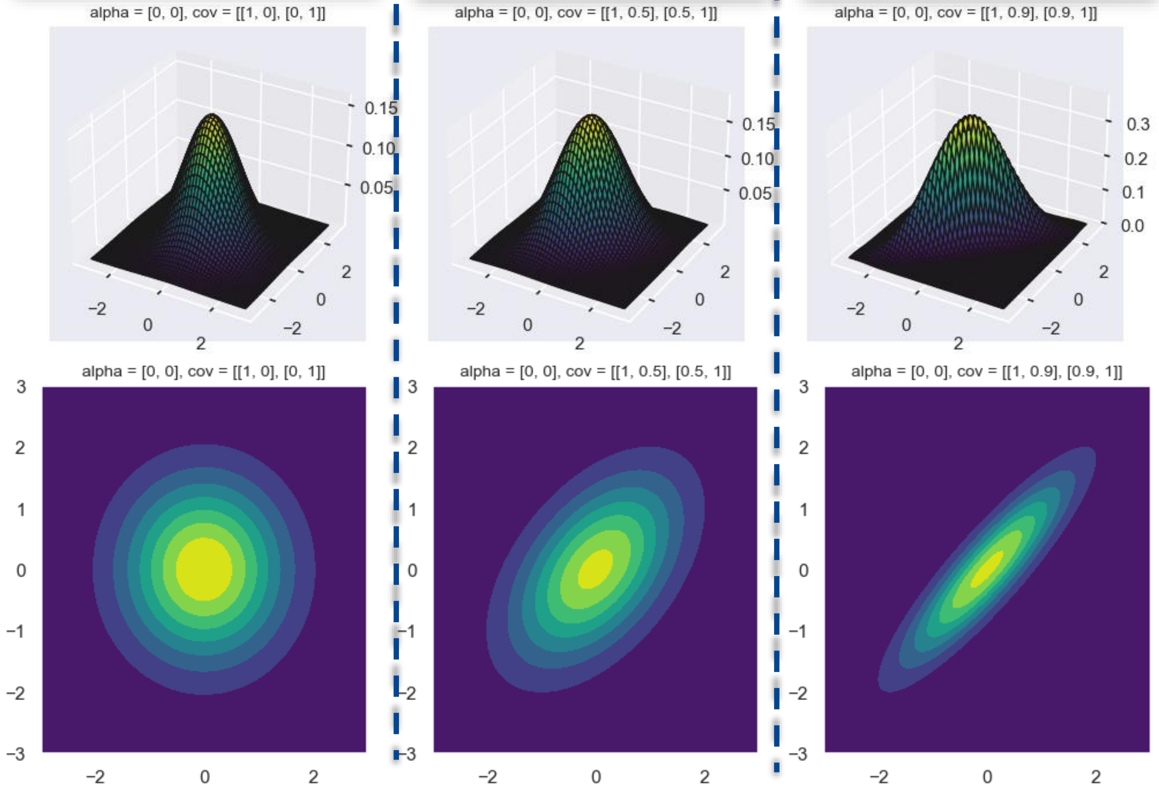
Target



Cov = [[1, 0], [0, 1]]

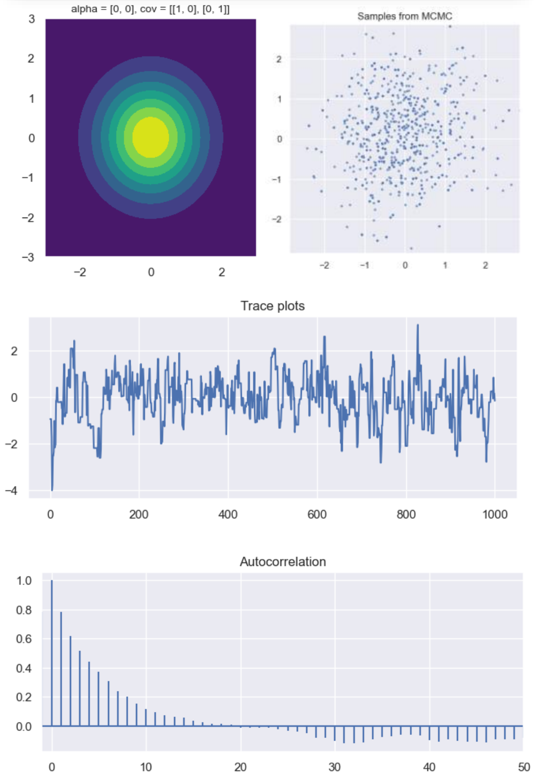
Cov = [[1, 0.5], [0.5, 1]]

Cov = [[1, 0.9], [0.9, 1]]



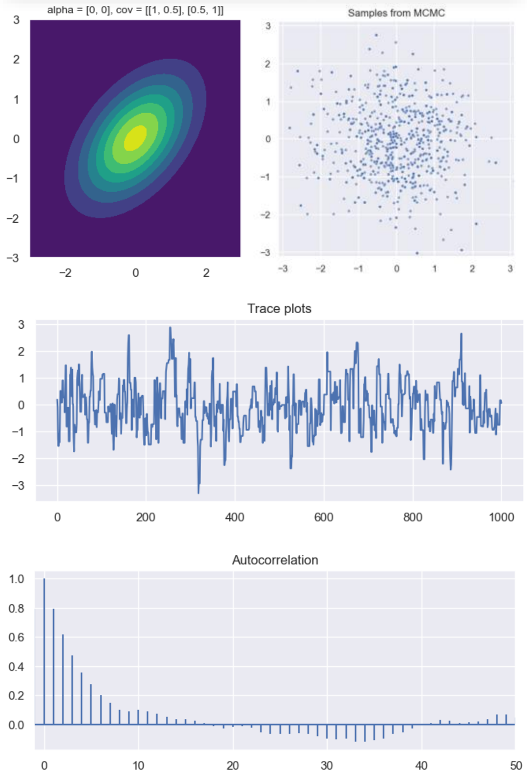
Case 1: Target and proposal are both Gaussian

Cov = [[1, 0], [0, 1]]



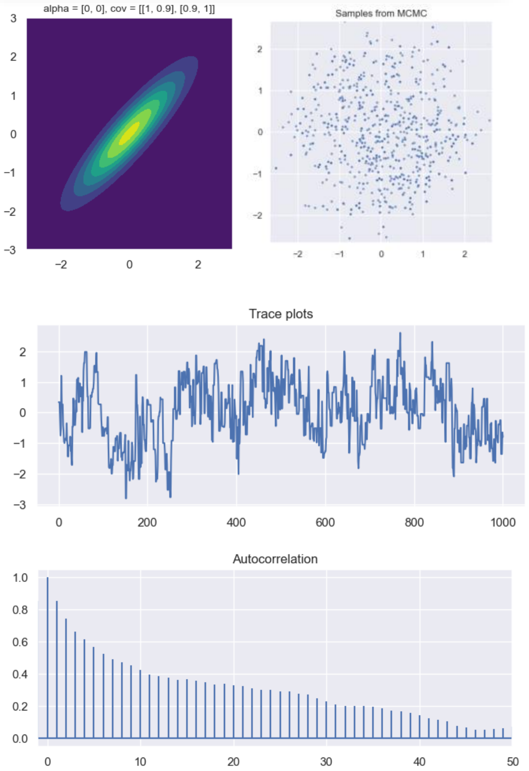
Acceptance rate: 57%

Cov = [[1, 0.5], [0.5, 1]]



Acceptance rate: 55%

Cov = [[1, 0.9], [0.9, 1]]

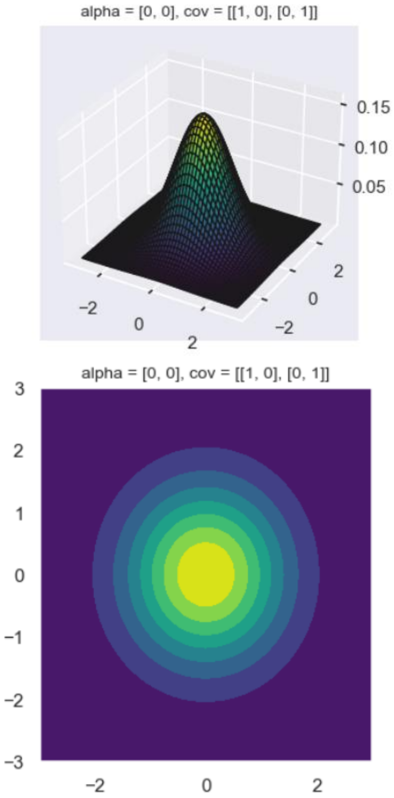


Acceptance rate: 58%

NTNU

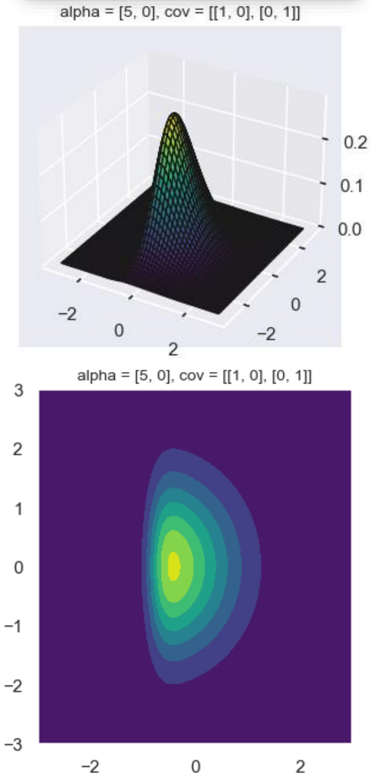
Case 2: Target is Gaussian, proposals are skew Gaussian

Target

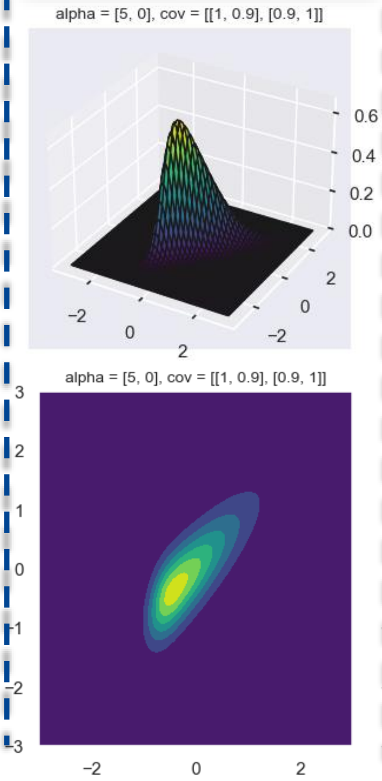


Proposal

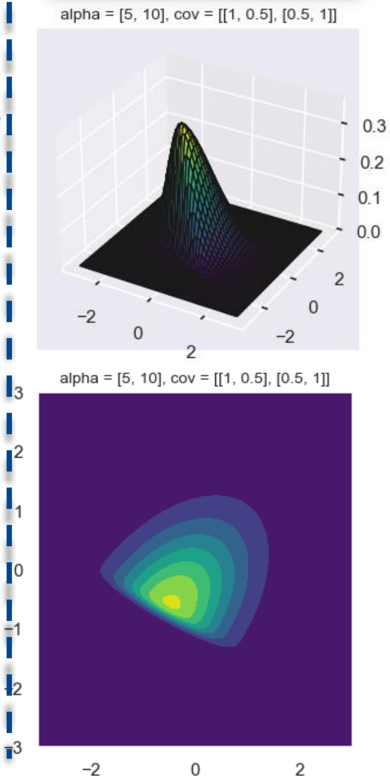
shape = [5, 0],
cov = [[1, 0], [0, 1]]



shape = [5, 0],
cov = [[1, 0.9], [0.9, 1]]

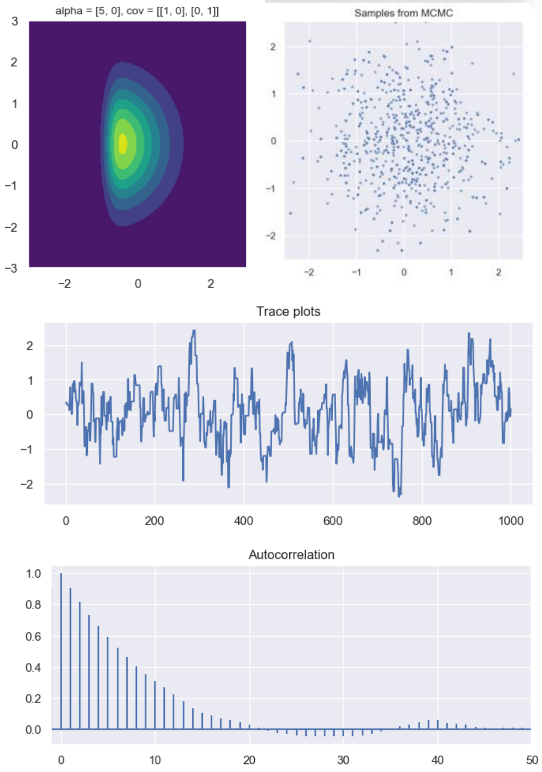


shape = [5, 10],
cov = [[1, 0.5], [0.5, 1]]



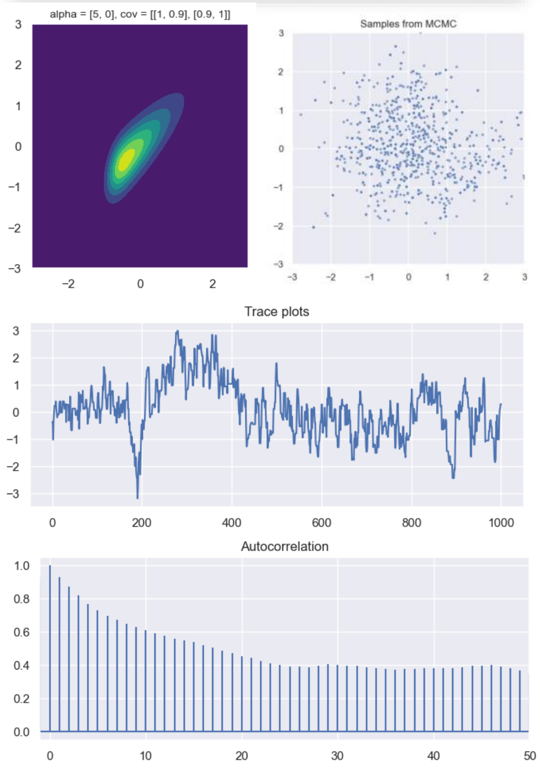
Case 2: Target is Gaussian, proposals are skew Gaussian

shape = [5, 0], cov = [[1, 0], [0, 1]]



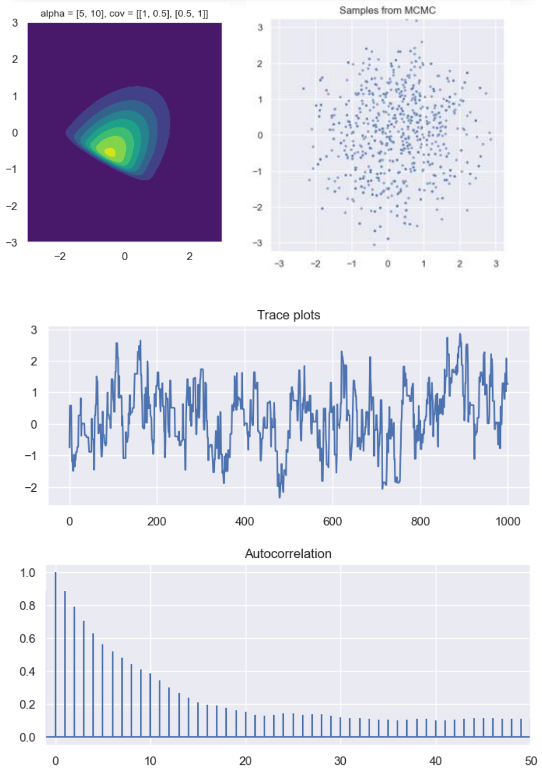
Acceptance rate: 56%

shape = [5, 0], cov = [[1, 0.9], [0.9, 1]]



Acceptance rate: 62%

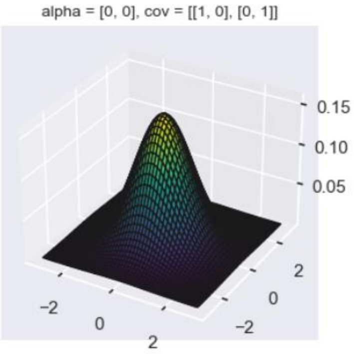
shape = [5, 10], cov = [[1, 0.5], [0.5, 1]]

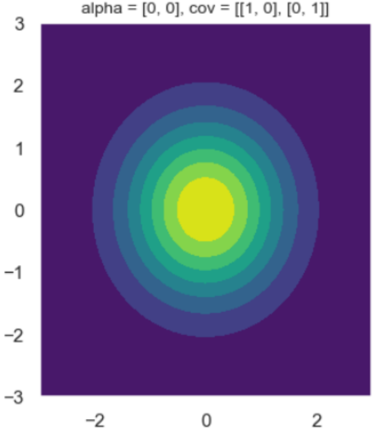


Acceptance rate: 56%

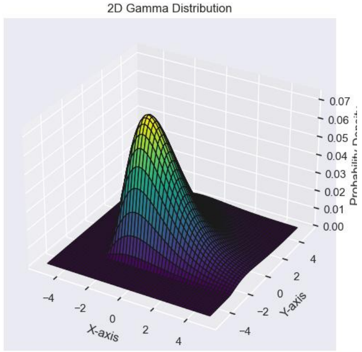
Case 3: Target is Gaussian, proposals are gama / log

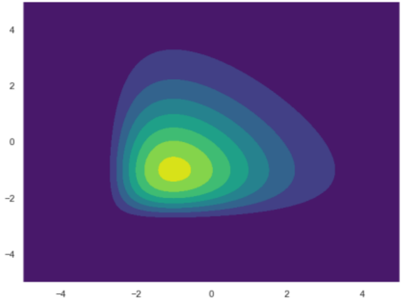
Target

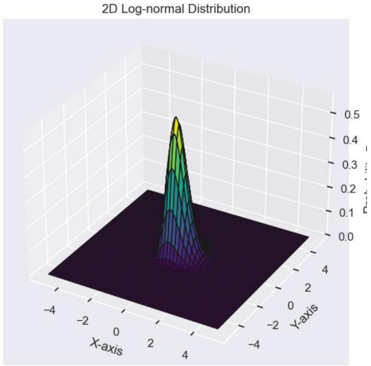


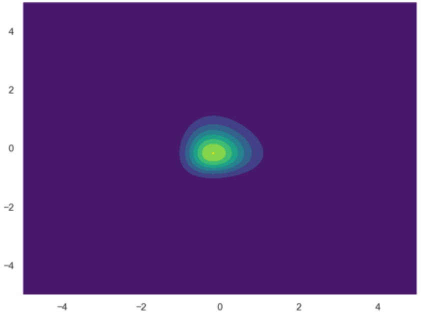


Proposal





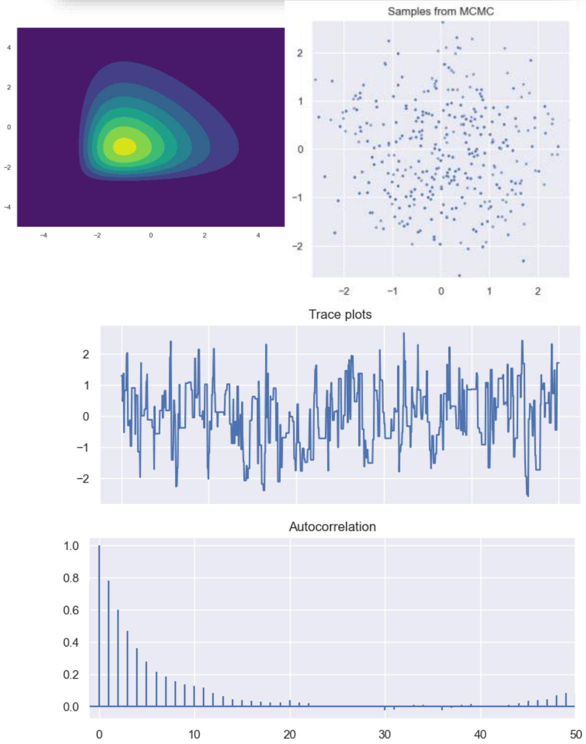




NTNU

Case 3: Target is Gaussian, proposals are gama / log

Gama distribution

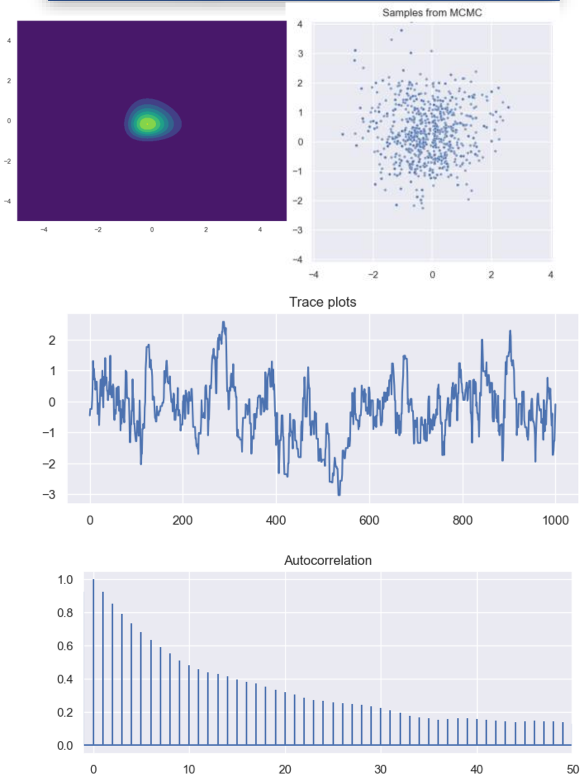


Acceptance rate: 37%

Note that the support of gama and log normal are not the same as Gaussian.

But still, they give reasonable result.

Lognormal distribution



Acceptance rate: 66%

The effect of proposal selection in MCMC

Given a target distribution,

- What is the difference we select one proposal over another?

-> It seems that the more “similar” to the target the proposal is, the better the samples are (in term of autocorrelation)

- Is there a “better” proposal distribution over some others?

-> Choose the best-knowledge proposal that is similar to target , i.e. the prior distribution?