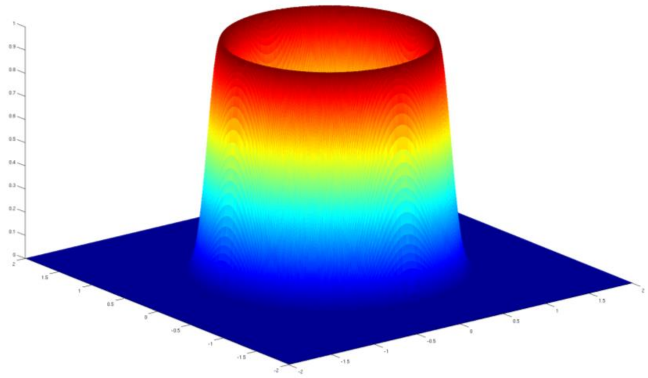
Gaussian Processes to Speed up Hamiltonian Monte Carlo

Matthieu Lê



Murray, Iain http://videolectures.net/mlss09uk_murray_mcmc/

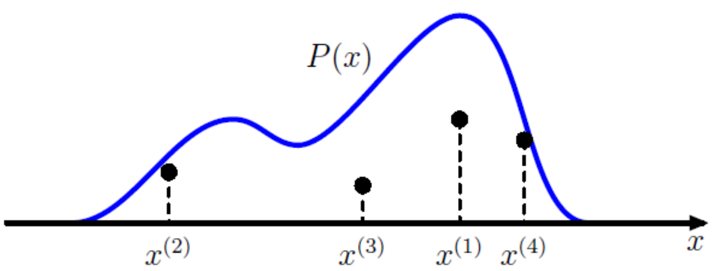
Rasmussen, Carl Edward. "Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals." *Bayesian Statistics 7: Proceedings of the 7th Valencia International Meeting*. Oxford University Press, 2003.

Neal, Radford M (2011). " MCMC Using Hamiltonian Dynamics. " In Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC.

MCMC

- Monte Carlo : Rejection sampling, Importance sampling, …
- MCMC : Markov Chain Monte Carlo
- Sampling technique to estimate a probability distribution

Draw samples from complex probability distributions



In general :

$$\int f(x)P(x)dx \approx \frac{1}{S}\sum_{s=1}^{S} f\left(x^{(s)}\right), \qquad x^{(s)} \sim P(x)$$

Objective

Bayesian Inference :

- θ : Model parameter
- x_i : Observations
- $P(\theta)$: prior on the model parameters
- $P(x_1 \ \dots x_n|\theta)$: likelyhood, potentially expensive to evaluate

$$P(\theta|x_1 \ \dots x_n) = \frac{P(x_1 \ \dots x_n|\theta)P(\theta)}{\int P(x_1 \ \dots x_n|\theta)P(\theta)d\theta}$$

$$P(\theta|x_1 \ \dots x_n) = \frac{1}{Z} P(x_1 \ \dots x_n|\theta)P(\theta)$$

MCMC : sample $P(\theta|x_1 \ \dots x_n)$ without the knowledge of Z

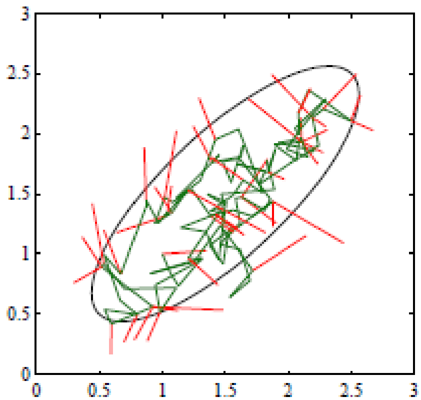
Metropolis-Hastings

We want to draw samples from $P \; \alpha \; \tilde{P}$

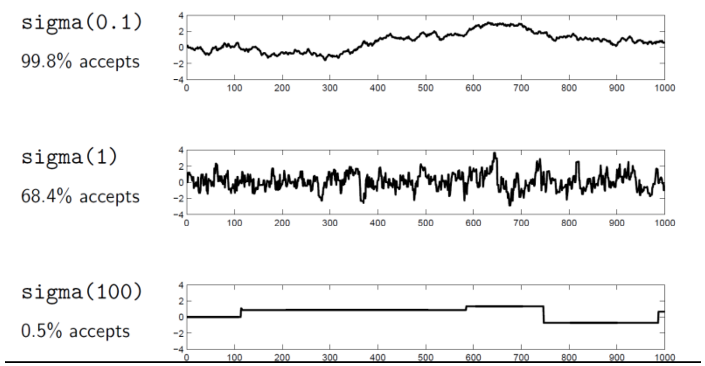
1. Propose new θ' from the transition function $Q(\theta', \theta)$ (e.g. $N(\theta, \sigma^2)$). The transition from one parameter to another is a Markov Chain

2. Accept the new parameter θ' with a probability $\min(1, \frac{\tilde{P}(\theta')Q(\theta,\theta')}{\tilde{P}(\theta)Q(\theta',\theta)})$

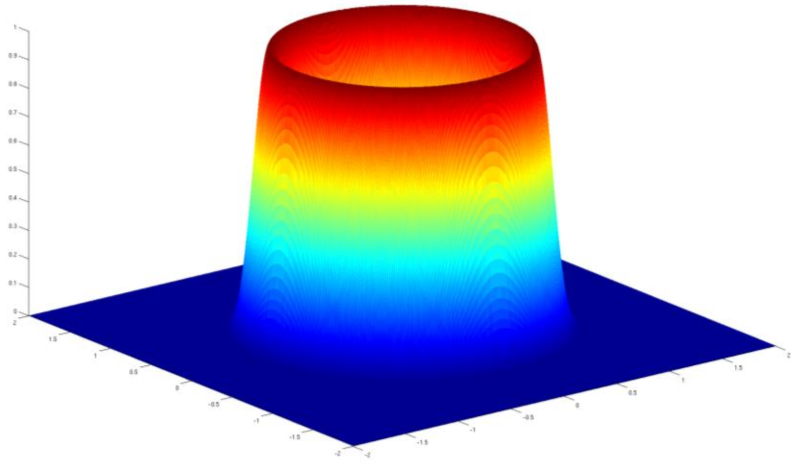
Q must be chosen to fulfill some technical requirements.
Samples are not independent.

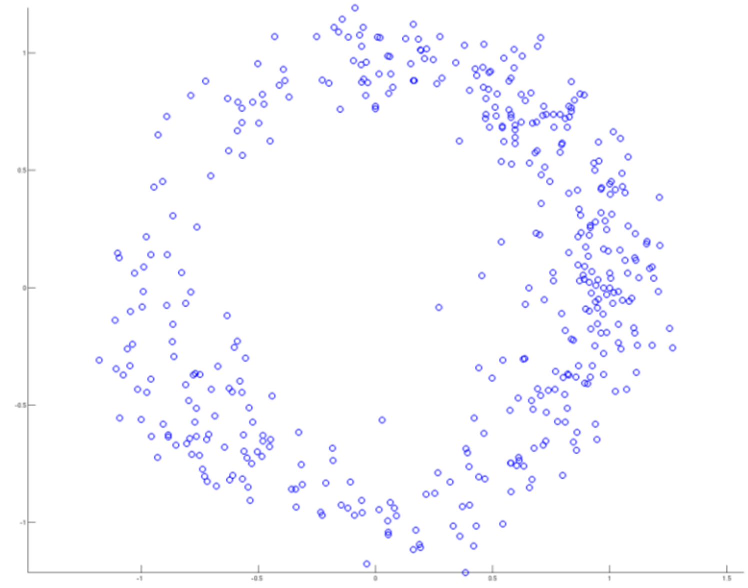


Sample $N(0,1\ ))$



Metropolis-Hastings





Metropolis-Hastings: 1000 samples, step 0.1
Rejection rate: 44%

Problem : The proposed samples come from a Gaussian. There is possibly an important rejection rate.

Hamiltonian Monte Carlo

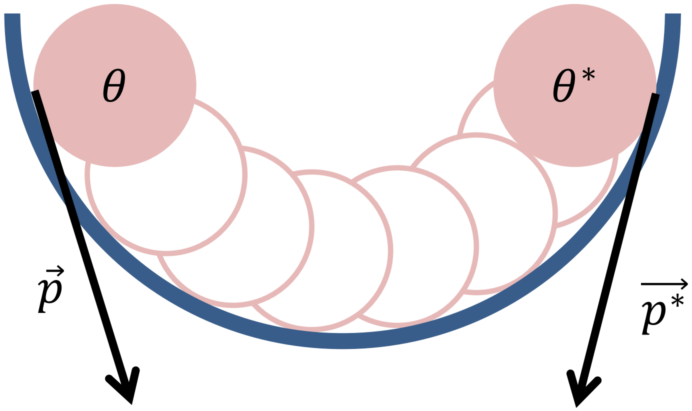
Same idea as Metropolis-Hastings **BUT** the proposed samples now come from the Hamiltonian dynamics :

$$H = E_{pot} + E_{kin}$$

$$E_{pot} = -\log\big(P(\theta)\big), \qquad E_{kin} = \frac{p^2}{2m}$$

$$\frac{\partial \theta}{\partial t} = \frac{\partial E_{kin}}{\partial p} = \frac{p}{m}$$

$$\frac{\partial p}{\partial t} = -\frac{\partial E_{pot}}{\partial \theta}$$



Hamiltonian Monte Carlo

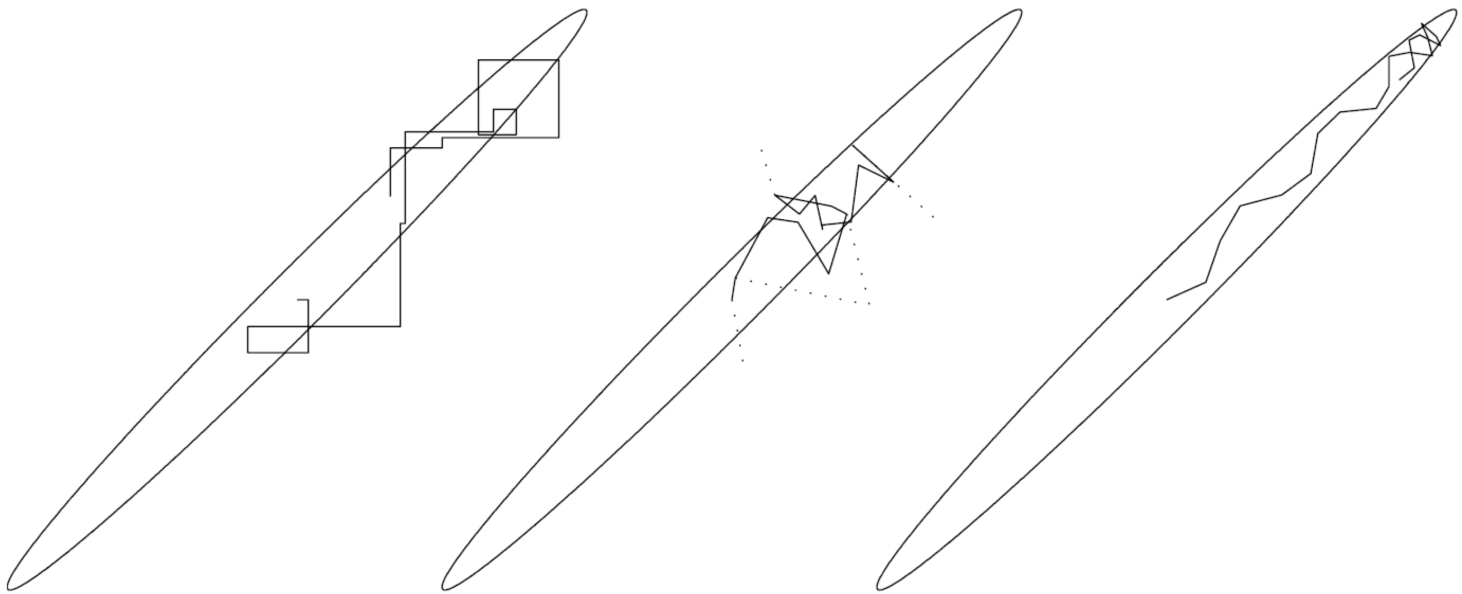
Algorithm :

- Sample p according to its known distribution
- Run the Hamiltonian dynamics during a time T
- Accept the new sample with probability :

$$\min(1, \exp\left(-E^*_{pot} + E_{pot}\right) \exp(-E^*_{kin} + E_{kin}))$$

The Energy is conserved so the acceptance probability should theoretically be 1. Because of the numerical precision, we need the Metropolis-Hastings type decision in the end.

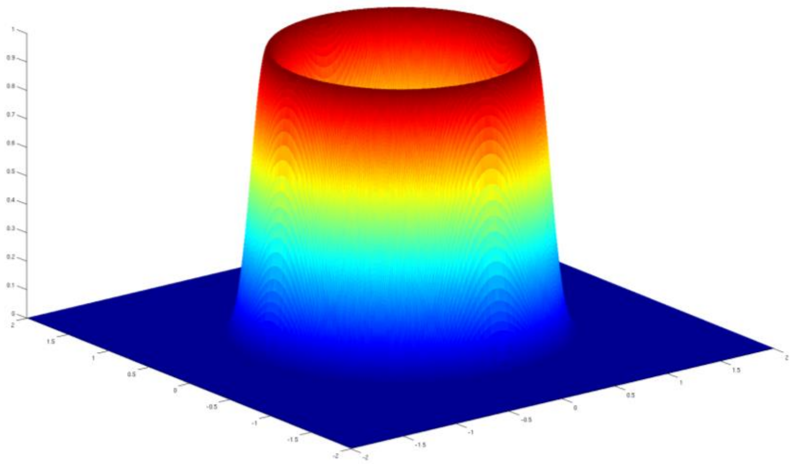
Hamiltonian Monte Carlo

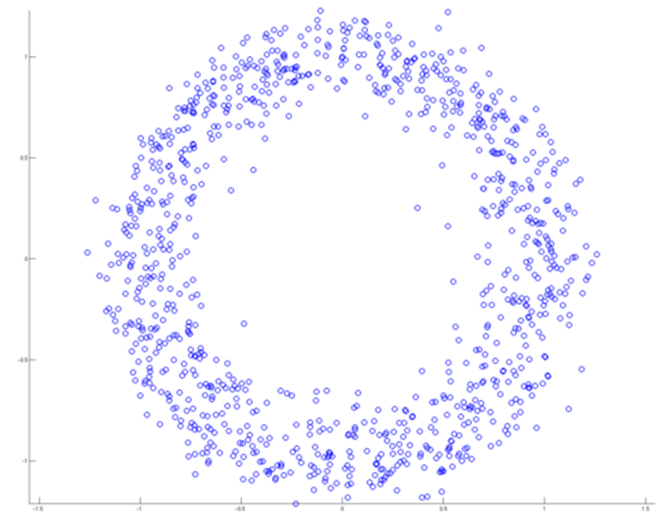


Gibbs Sampling Metropolis-Hastings Hamiltonian Monte Carlo

Hamiltonian Monte Carlo

Advantage : The Hamiltonian stays (approximately) constant during the dynamic, hence lower rejection rate !





1000 samples, L = 200, ϵ =0,01
Rejection rate = 0%

Problem : Computing the Hamiltonian dynamic requires computing the model partial derivatives, high number of simulation evaluation !

Neal, Radford M (2011). " MCMC Using Hamiltonian Dynamics. " In Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC.

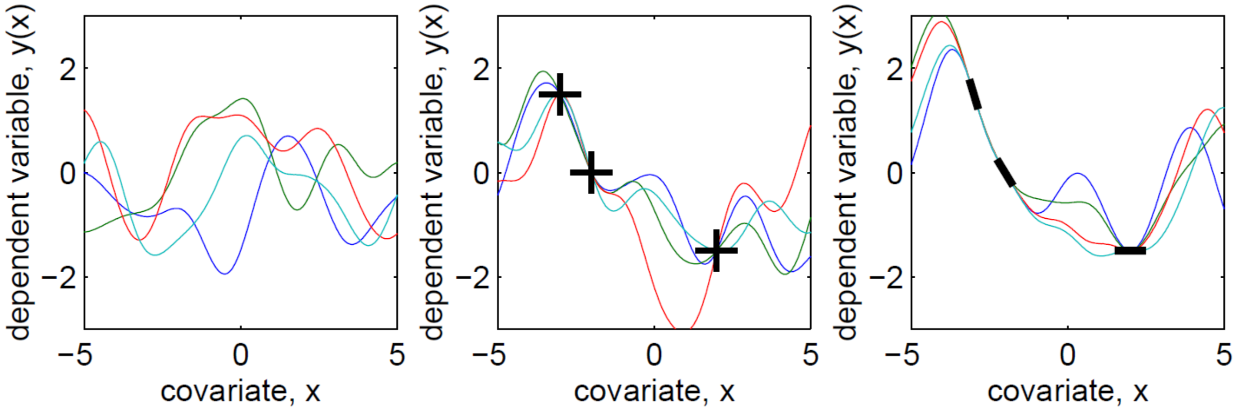
Gaussian Process HMC

Same algorithm as HMC **BUT** the Hamiltonian dynamic is computed using Gaussian process simulating E_{pot}

$$\frac{\partial p}{\partial t} = -\frac{\partial E_{pot}}{\partial \theta}$$

Gaussian process = distribution over smooth function to approximate E_{pot} :

$$P(E_{pot}|\theta) \sim N(0, \Sigma), \qquad \Sigma_{\mathrm{pq}} = \omega_0 \exp(-\frac{1}{2}\sum_{d=1}^{D}\left(x_d^p - x_d^q\right)^2/\omega_d^2)$$



Gaussian Process HMC

Once the Gaussian process is defined with a covariance matrix, we can predict new values :

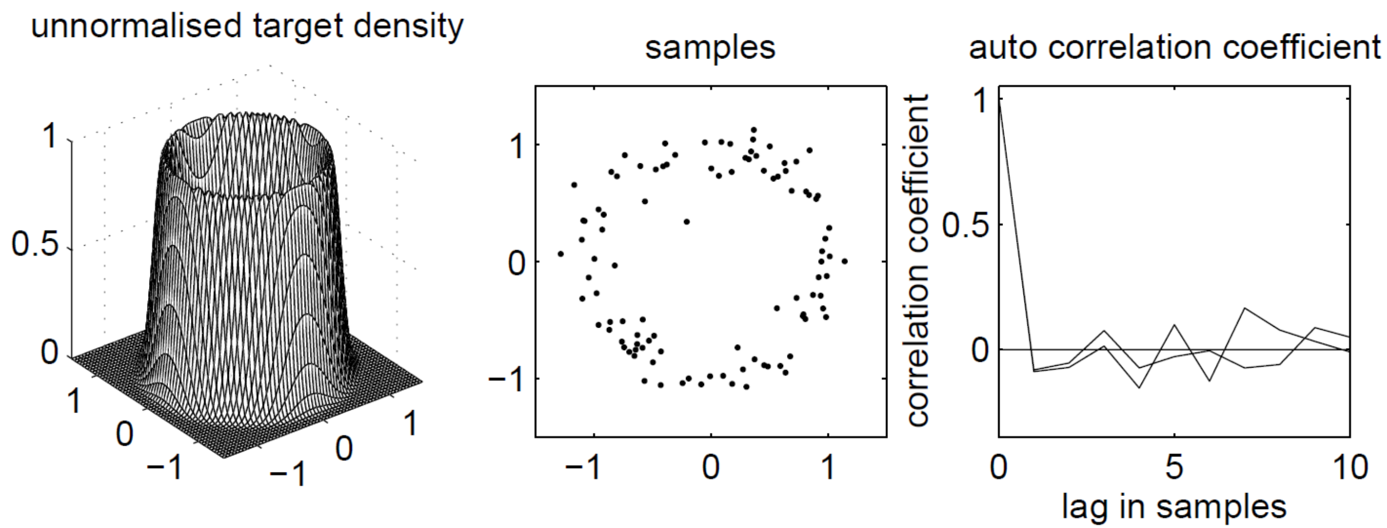
$$P\left(E_{pot}^{*} \middle| \boldsymbol{\theta}, \boldsymbol{E_{pot}}, \theta^{*}\right) \sim N(\mu, \sigma^{2}),$$

If the Gaussian process is "good", $\mu(\theta^{*}) \approx$ target density

Algorithm :

1. Initialization :
   - Evaluate the target density at D random points to define the Gaussian process.
2. Exploratory phase :
   - HMC with $E_{pot} = \mu - \sigma$: evaluation of points with high target value and high uncertainty. Evaluate the real target density at the end of each iteration.
3. Sampling phase :
   - HMC with $E_{pot} = \mu$.

Gaussian Process HMC



Rasmussen, Carl Edward. "Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals." *Bayesian Statistics 7: Proceedings of the 7th Valencia International Meeting*. Oxford University Press, 2003.

Conclusion

- Metropolis-Hastings : few model evaluation per iteration but important rejection rate
- Hamiltonian Monte Carlo : a lot of model evaluation per iteration but low rejection rate
- GPHMC : few model evaluation per iteration and low rejection rate
  - BUT : Initialization requires model evaluations to define a “good” Gaussian process
  - BUT : Exploratory phase requires one model evaluation per iteration


That's all Folks!