COMP6247 - **Reinforcement & Online Learning (MSc)**

Online Learning

Week 5: Markov Chain Monte Carlo

Christine Evers

School of Electronics & Computer Science
University of Southampton

Spring Semester 2020/2021

Motivation: Monte Carlo Methods

Problem: Integrals for non-linear, non-Gaussian state space models

- Bayes's theorem:

$$p(\mathbf{z} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{z})\, p(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \mid \mathbf{z})\, p(\mathbf{z})\; d\mathbf{z}}$$

- Maximum *a posteriori* estimate:

$$\mathbb{E}_{p(\mathbf{z} \mid \mathbf{x})}[\mathbf{z}] = \int_{\mathcal{Z}} \mathbf{z}\, p(\mathbf{z} \mid \mathbf{x})\, d\mathbf{z}$$

- Variance

$$\mathrm{var}_{p(\mathbf{z} \mid \mathbf{x})}[\mathbf{z}] = \int_{\mathcal{Z}} \mathbf{z}^2\, p(\mathbf{z} \mid \mathbf{x})\, d\mathbf{z} - \boldsymbol{\mu}$$

$\mathbf{x}$: Data vector, $\mathbf{z}$: Latent vector

Monte Carlo methods: Algorithms that solve integrals by random sampling

- Draw samples (random variates) from a proposal distribution:

$$\mathbf{z}^{(i)} \sim q(\mathbf{z})$$

- Approximate integrals, e.g.,

$$\mathbb{E}_{p(\mathbf{z}_n \mid \mathbf{x}_{1:n})}[\mathbf{z}] \approx \frac{1}{I} \sum_{i=1}^{I} \mathbf{z}^{(i)}$$

How do we draw the random samples?

$q(\cdot)$: Proposal pdf, $p(\cdot)$: Target pdf

Lecture Overview

Week 4: Bayesian Inference

Week 5: Markov Chain Monte Carlo (MCMC)

Part 1:	Basic Sampling Methods:	Rejection Sampling
Part 2:	Markov Chain Monte Carlo:	Metropolis-Hastings
Part 3:	Markov Chain Monte Carlo:	Gibbs Sampling

Week 6: Importance Sampling & Sequential Monte Carlo

Learning Outcomes

Following this week's lecture on Markov Chain Monte Carlo methods, you should be able to:

1) **Explain how random sampling can be used to approximate integrals;**
2) **Understand how MCMC methods use Markov chains to sample from target distributions;**
3) **Apply different techniques for MCMC to real-world problems.**

Further Reading

Southampton

Textbooks:

- C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- K. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.
- D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.
- Goodfellow, Bengio, Courville, Bengio, *Deep Learning*. MIT Press, 2016.

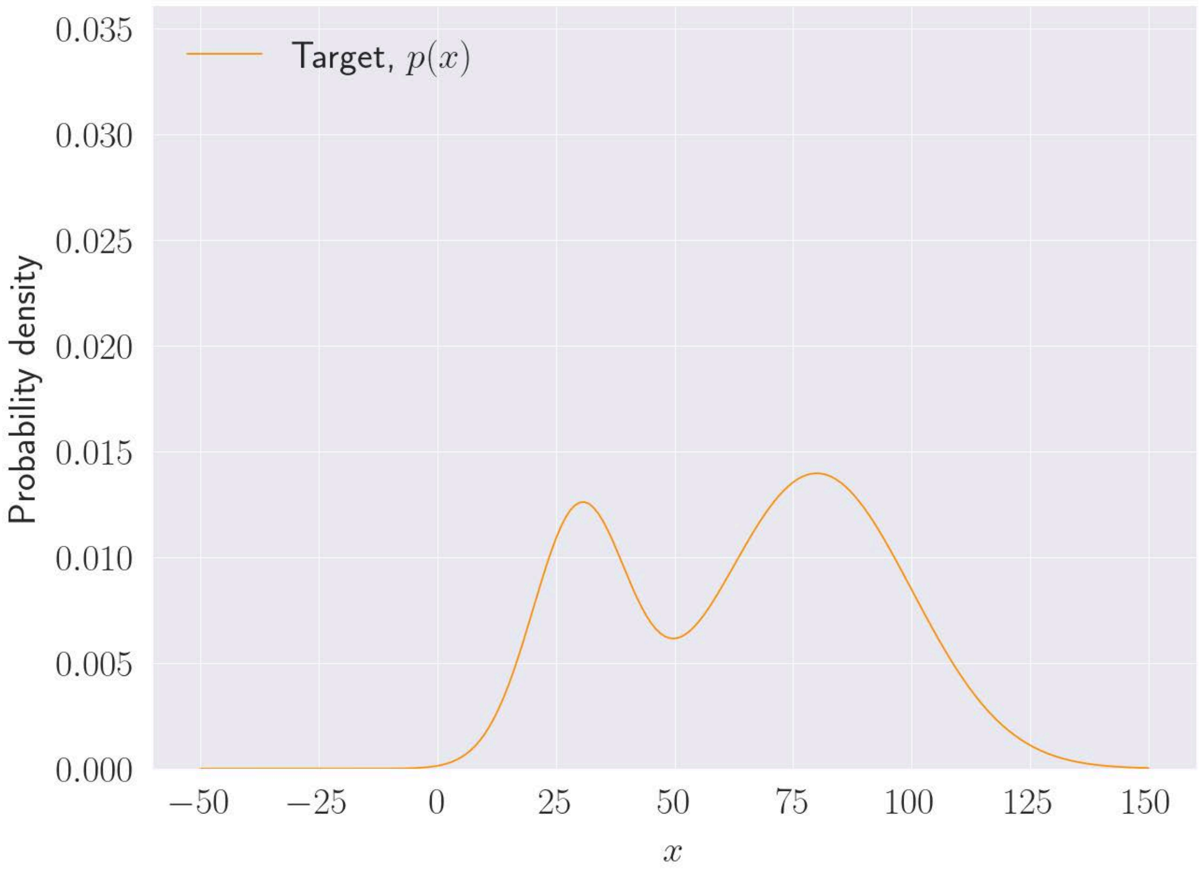
Tutorial papers:

- R. M. Neal, “Probabilistic Inference using Markov Chain Monte Carlo methods,” Technical Report CRG-TR-03-1, 1993. Online
- C. J. Geyer, “Practical Markov Chain Monte Carlo,” *in Statistical Science*, 7(4), 1992. Online

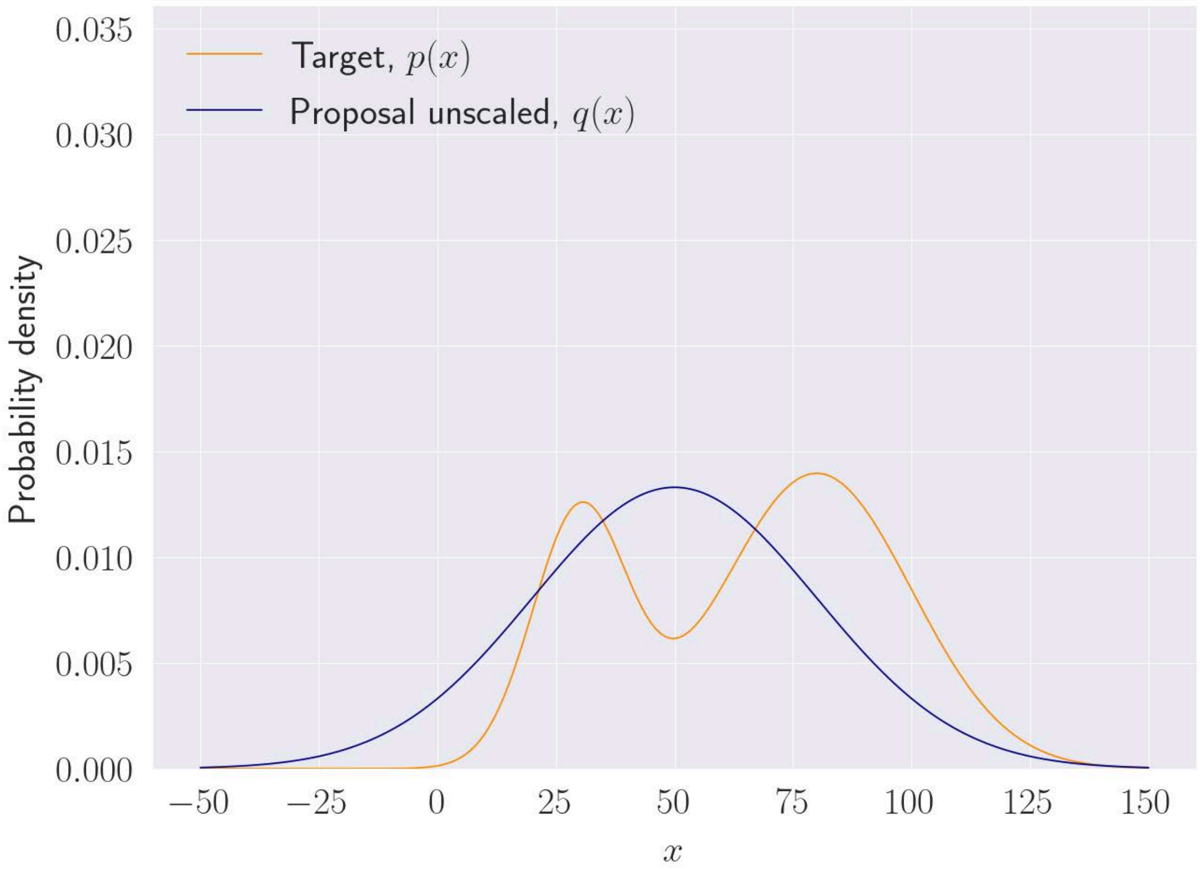
Basic Sampling Methods:
Rejection Sampling

Rejection Sampling

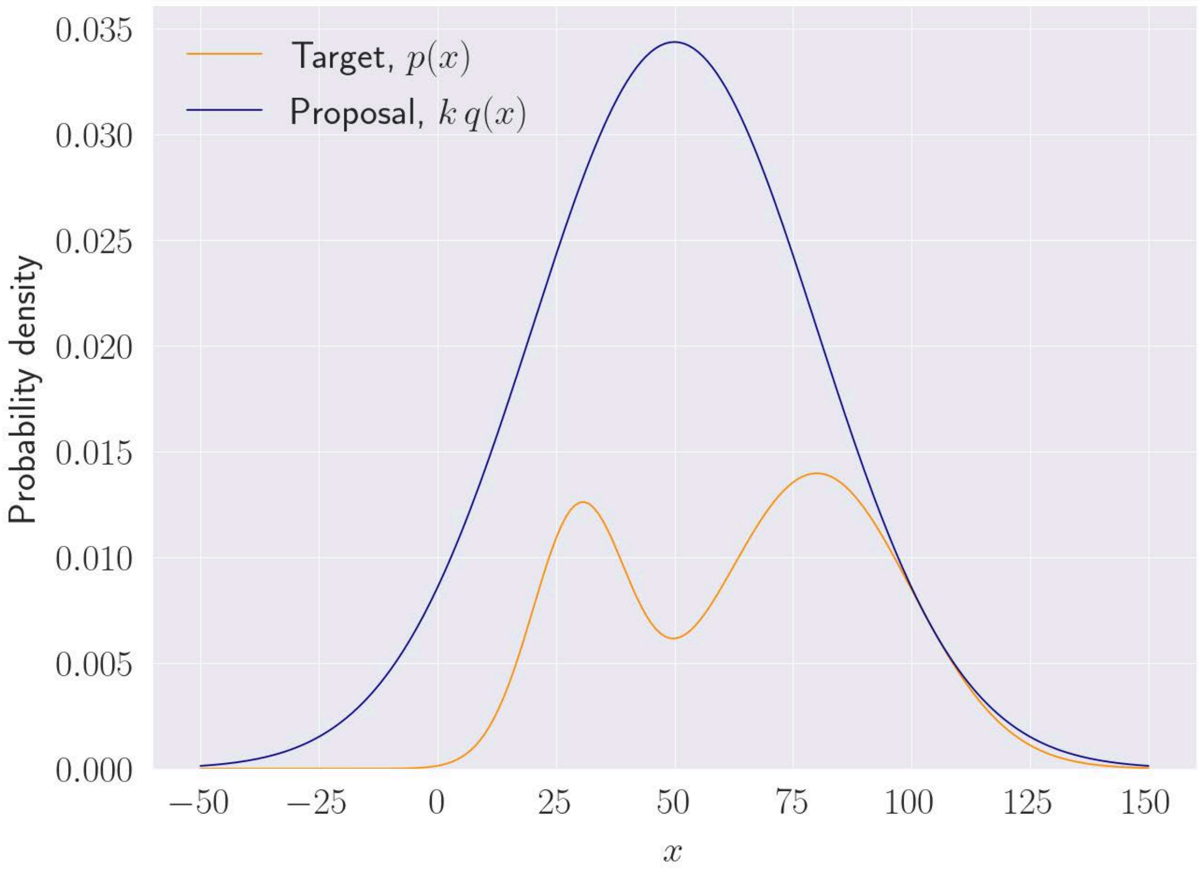
Rejection Sampling



Rejection Sampling



Rejection Sampling



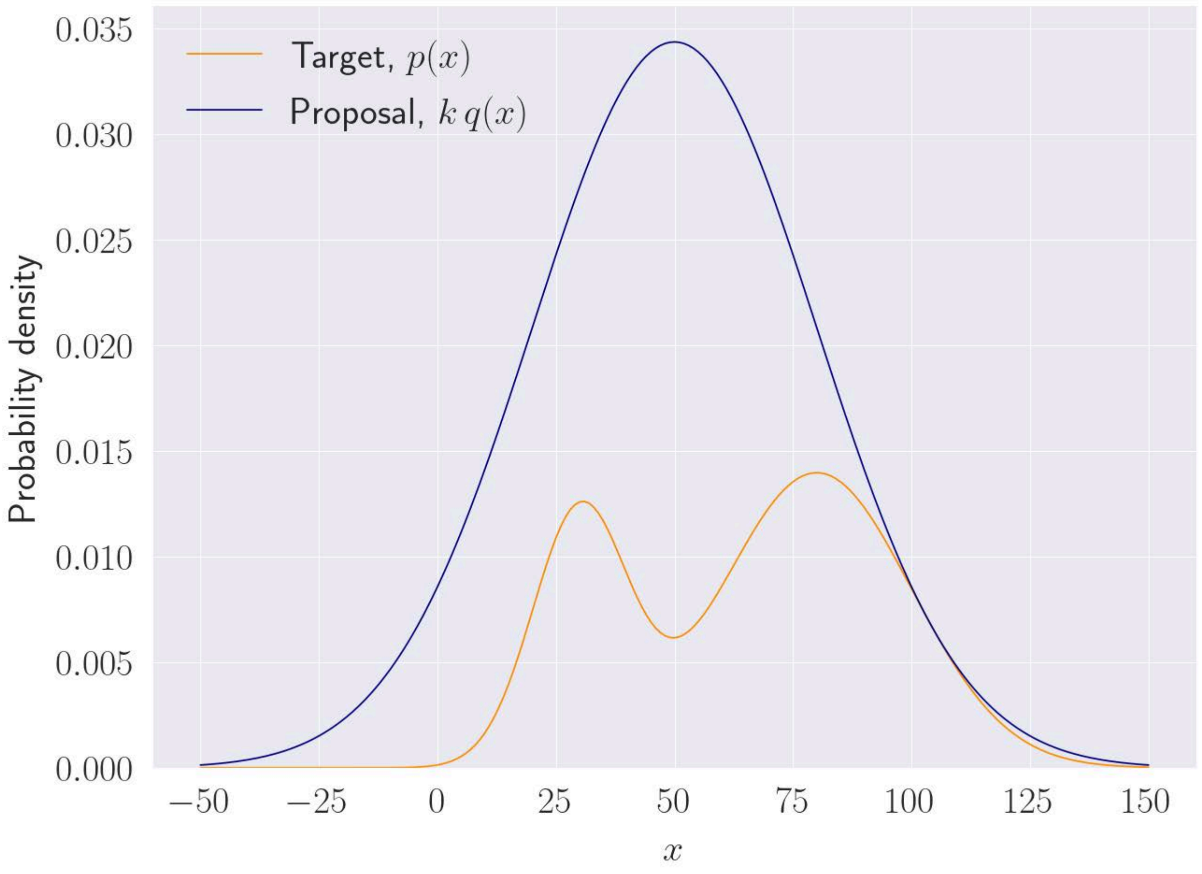
Rejection Sampling

Rejection Sampling Algorithm:

For $\ell = 1,\ldots,L$:

$p(\cdot)$: Target pdf

$q(\cdot)$: Proposal pdf, k : Scaling constant



Rejection Sampling

Rejection Sampling Algorithm:

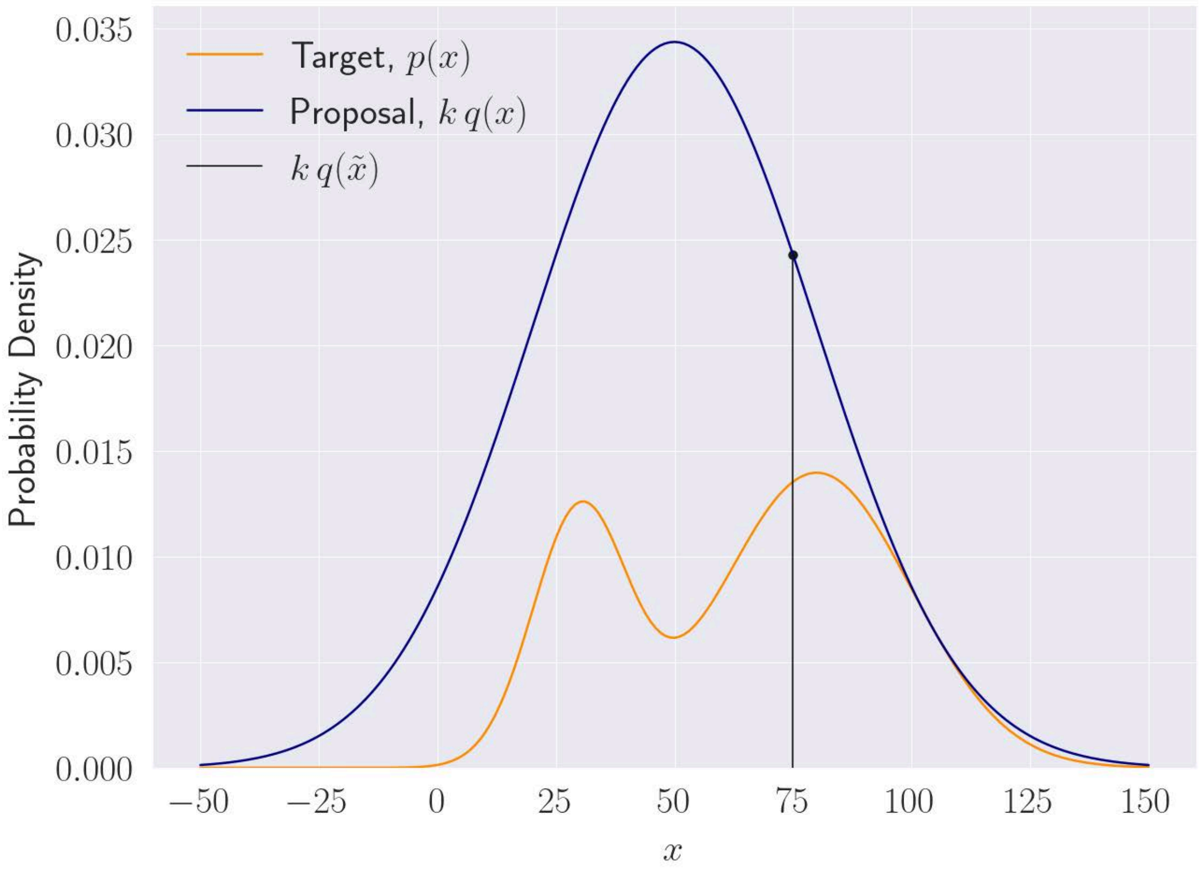
For $\ell = 1, \ldots, L$:

Repeat:

1. Draw candidate $\tilde{\mathbf{z}} \sim q(\mathbf{z})$

$p(\cdot)$: Target pdf

$q(\cdot)$: Proposal pdf, k : Scaling constant



Rejection Sampling

Rejection Sampling Algorithm:

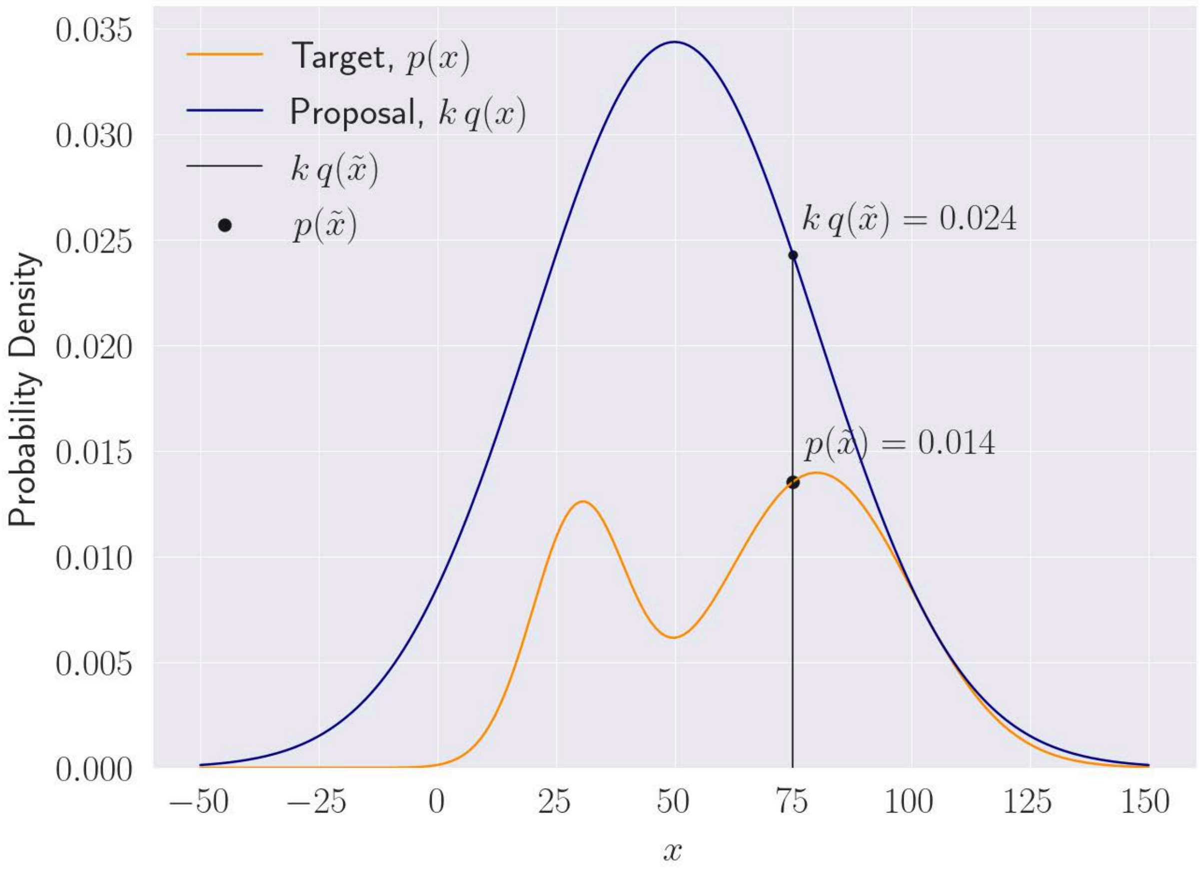
For $\ell = 1, \ldots, L$ **:**

Repeat:

1. Draw candidate $\tilde{\mathbf{z}} \sim q(\mathbf{z})$

$p(\cdot)$: Target pdf

$q(\cdot)$: Proposal pdf, k : Scaling constant



Rejection Sampling

Rejection Sampling Algorithm:

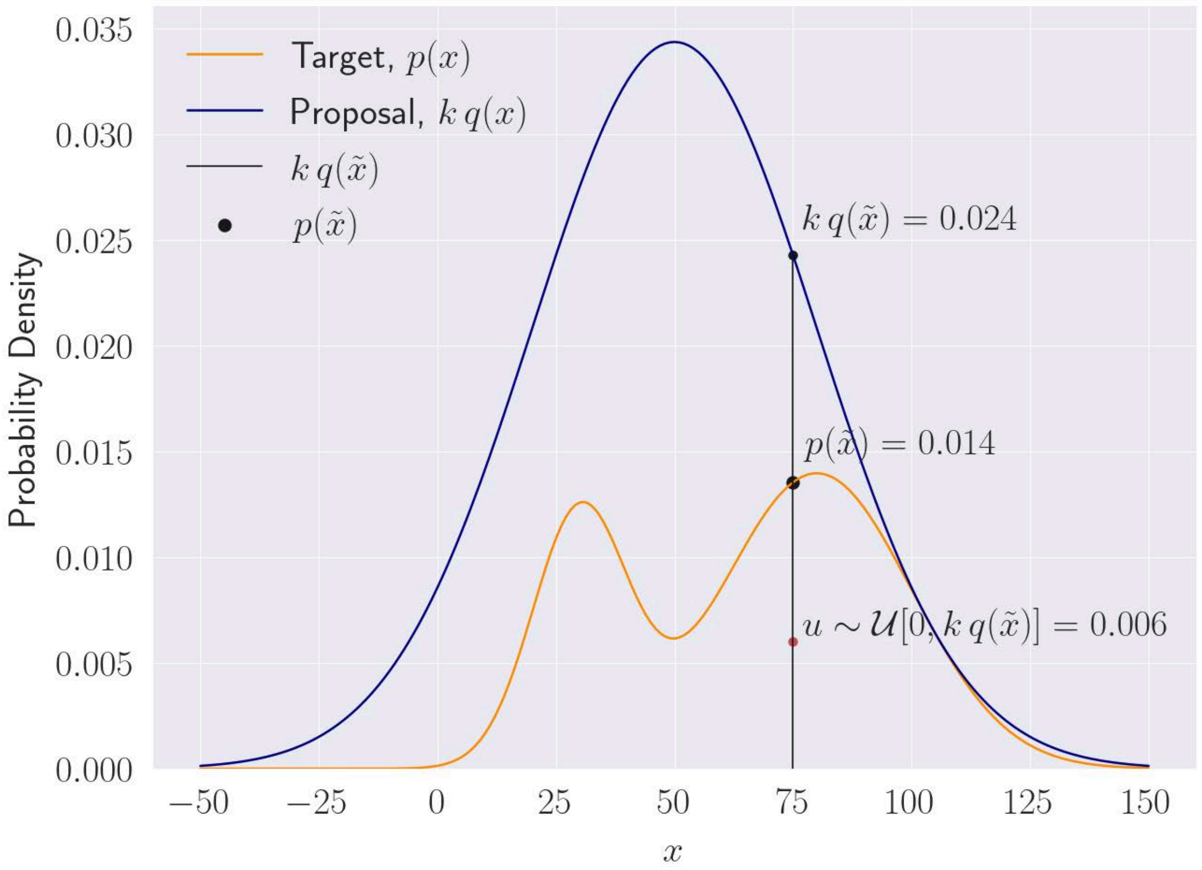
For $\ell = 1,\ldots,L$:

Repeat:

1. Draw candidate $\tilde{\mathbf{z}} \sim q(\mathbf{z})$
2. Draw $u \sim \mathcal{U}\left[0, k\, q(\tilde{\mathbf{z}})\right]$

$p(\cdot)$: Target pdf

$q(\cdot)$: Proposal pdf, k : Scaling constant



Rejection Sampling

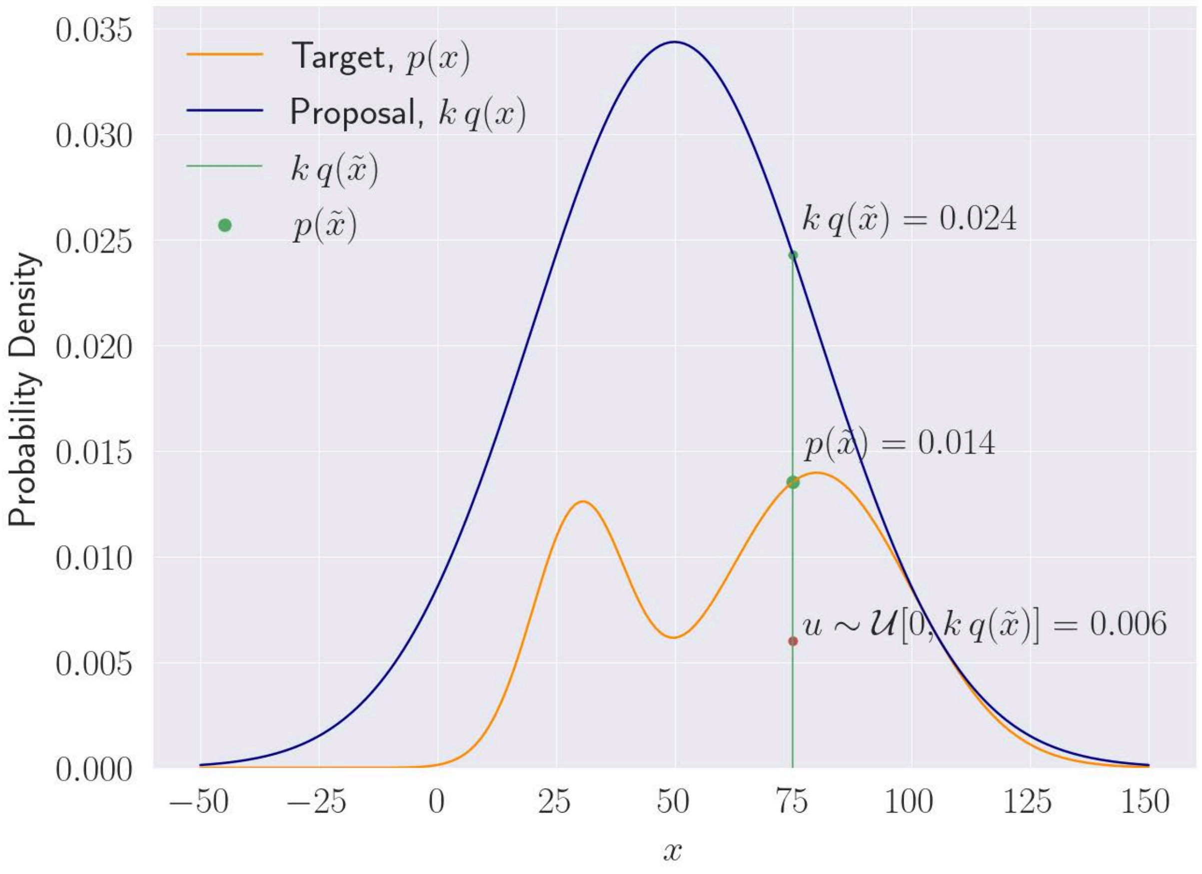
Rejection Sampling Algorithm:

For $\ell = 1,\ldots,L$:
Repeat:
1. Draw candidate $\tilde{\mathbf{z}} \sim q(\mathbf{z})$
2. Draw $u \sim \mathcal{U}\left[0, k\, q(\tilde{\mathbf{z}})\right]$

Until $u \leq p(\tilde{\mathbf{z}})$.

3. Accept $\hat{\mathbf{z}}^{(\ell)} = \tilde{\mathbf{z}}$

$p(\cdot)$: Target pdf
$q(\cdot)$: Proposal pdf, k : Scaling constant



Rejection Sampling

Rejection Sampling Algorithm:

For $\ell = 1, \ldots, L$:

Repeat:

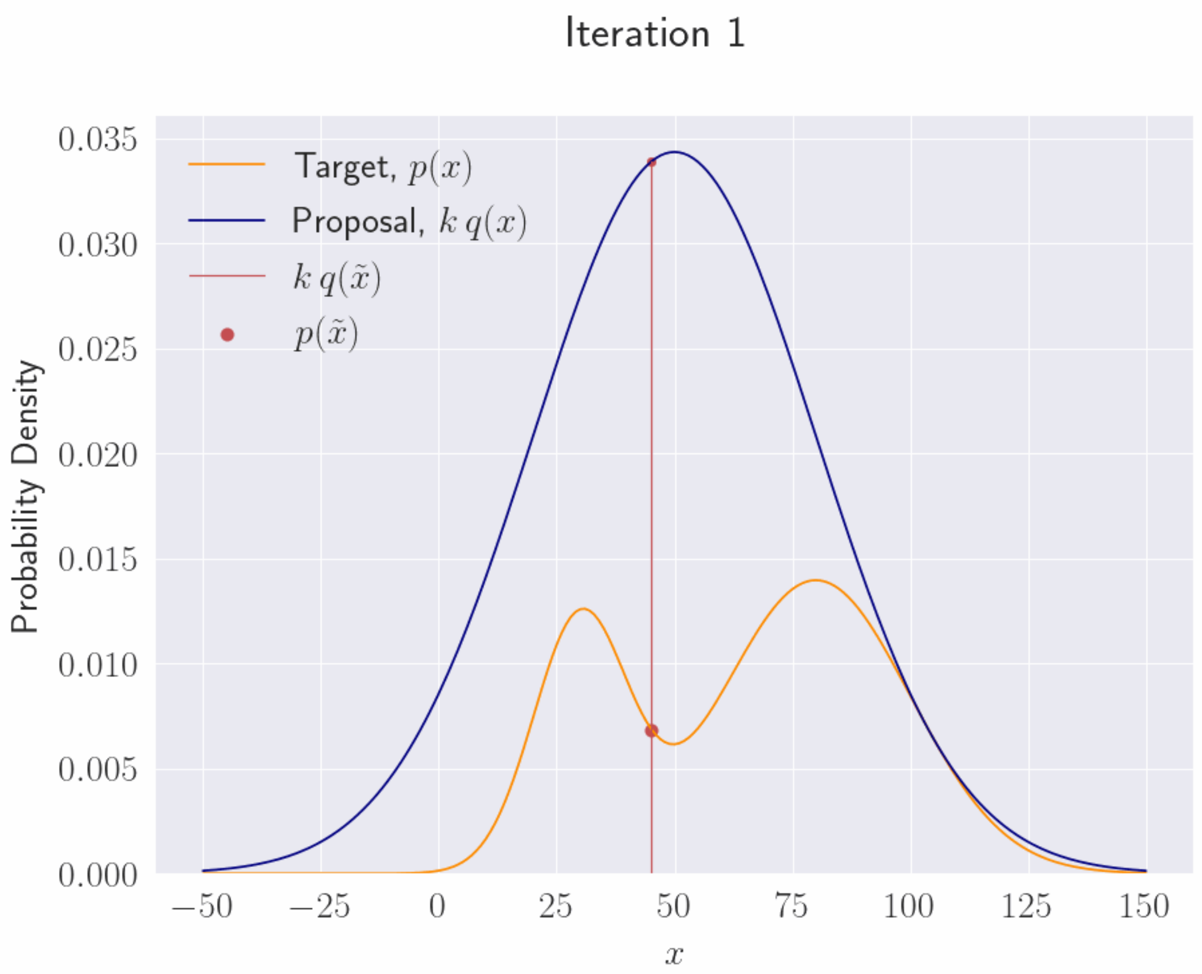
1. Draw candidate $\tilde{\mathbf{z}} \sim q(\mathbf{z})$
2. Draw $u \sim \mathcal{U}\left[0, k\, q(\tilde{\mathbf{z}})\right]$

Until $u \leq p(\tilde{\mathbf{z}})$.

3. Accept $\hat{\mathbf{z}}^{(\ell)} = \tilde{\mathbf{z}}$

$p(\cdot)$: Target pdf

$q(\cdot)$: Proposal pdf, k : Scaling constant



Rejection Sampling

Rejection Sampling Algorithm:

For $\ell = 1,\ldots,L$:

Repeat:

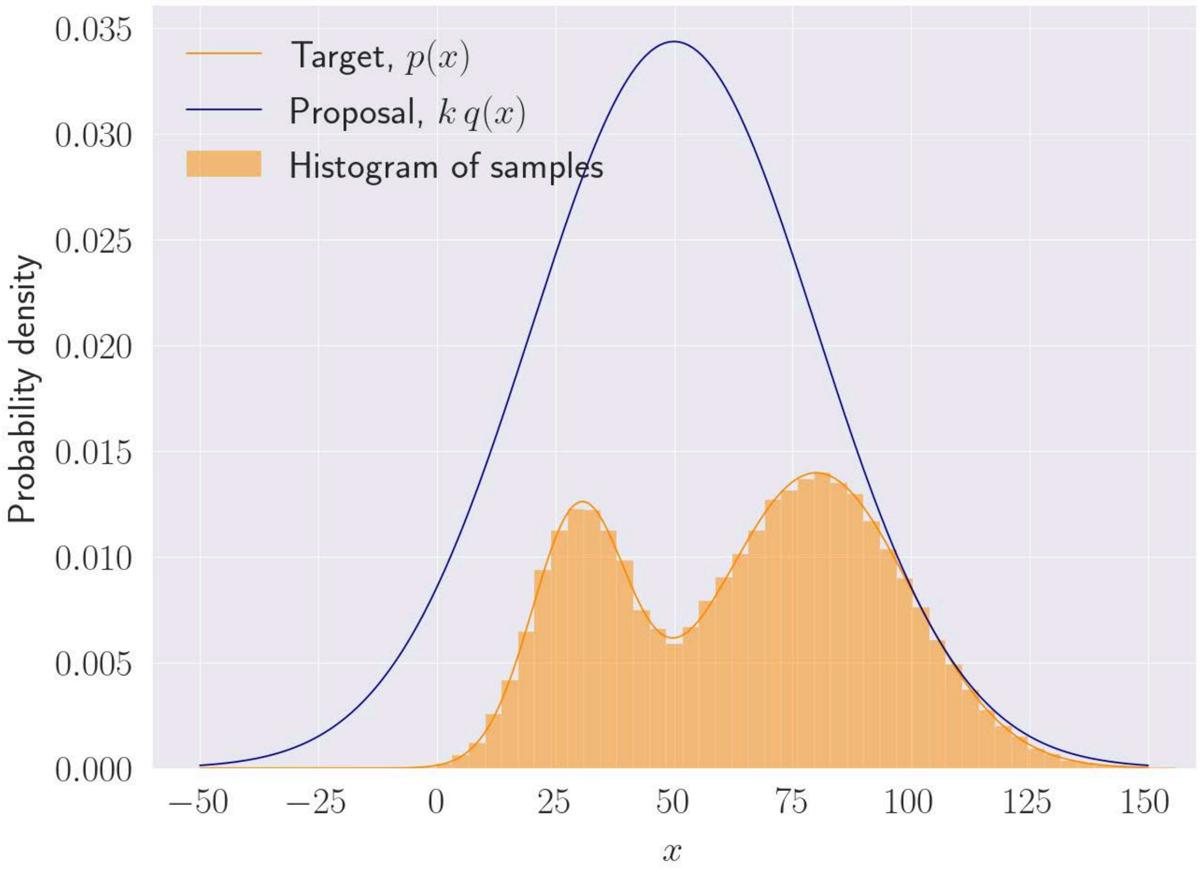
1. Draw candidate $\tilde{\mathbf{z}} \sim q(\mathbf{z})$
2. Draw $u \sim \mathcal{U}\left[0, k\, q(\tilde{\mathbf{z}})\right]$

Until $u \leq p(\tilde{\mathbf{z}})$.

3. Accept $\hat{\mathbf{z}}^{(\ell)} = \tilde{\mathbf{z}}$

$p(\cdot)$: Target pdf

$q(\cdot)$: Proposal pdf, k : Scaling constant



Rejection Sampling

UNIVERSITY OF Southampton

Acceptance Criterion:

$$\alpha = \frac{p(\tilde{\mathbf{z}})}{k\, q(\tilde{\mathbf{z}})}, \quad \text{where } \tilde{\mathbf{z}} \sim p(\mathbf{z})$$

$$p(\text{accept}) = \mathbb{E}_{q(\mathbf{z})}\left[\frac{p(\mathbf{z})}{k\, q(\mathbf{z})}\right] = \int_{\mathcal{Z}} \frac{p(\mathbf{z})}{k\, q(\mathbf{z})}\, q(\mathbf{z})\, d\mathbf{z}$$

$$= \frac{1}{k} \int_{\mathcal{Z}} p(\mathbf{z})\, d\mathbf{z}$$

Number of rejected samples $\propto$ area between curves. Keep k as small as possible, s.th. $k\, q(\mathbf{z}) \geq p(\mathbf{z})$!

Rejection Sampling

Acceptance Criterion:

$$\alpha = \frac{p(\tilde{\mathbf{z}})}{k\, q(\tilde{\mathbf{z}})}, \quad \text{where } \tilde{\mathbf{z}} \sim p(\mathbf{z})$$

Scaling constant s.th.

$k\, q(\mathbf{z}) \geq p(\mathbf{z})$

$$p(\text{accept}) = \mathbb{E}_{q(\mathbf{z})}\left[\frac{p(\mathbf{z})}{k\, q(\mathbf{z})}\right] = \int_{\mathcal{Z}} \frac{p(\mathbf{z})}{k\, q(\mathbf{z})}\, q(\mathbf{z})\, d\mathbf{z}$$

$$= \frac{1}{k} \int_{\mathcal{Z}} p(\mathbf{z})\, d\mathbf{z}$$

Number of rejected samples $\propto$ area between curves. Keep k as small as possible, s.th. $k\, q(\mathbf{z}) \geq p(\mathbf{z})$!

Rejection Sampling

Acceptance Criterion:

$$\alpha = \frac{p(\tilde{\mathbf{z}})}{\underbrace{k}_{\text{Scaling constant s.th. } k\,q(\mathbf{z}) \geq p(\mathbf{z})}\;\underbrace{q(\tilde{\mathbf{z}})}_{\text{Proposal distribution}}}, \quad \text{where } \tilde{\mathbf{z}} \sim p(\mathbf{z})$$

$$p(\text{accept}) = \mathbb{E}_{q(\mathbf{z})}\left[\frac{p(\mathbf{z})}{k\,q(\mathbf{z})}\right] = \int_{\mathcal{Z}} \frac{p(\mathbf{z})}{k\,q(\mathbf{z})}\,q(\mathbf{z})\,d\mathbf{z}$$

$$= \frac{1}{k}\int_{\mathcal{Z}} p(\mathbf{z})\,d\mathbf{z}$$

Number of rejected samples ∝ area between curves. Keep k as small as possible, s.th. $k\,q(\mathbf{z}) \geq p(\mathbf{z})$!

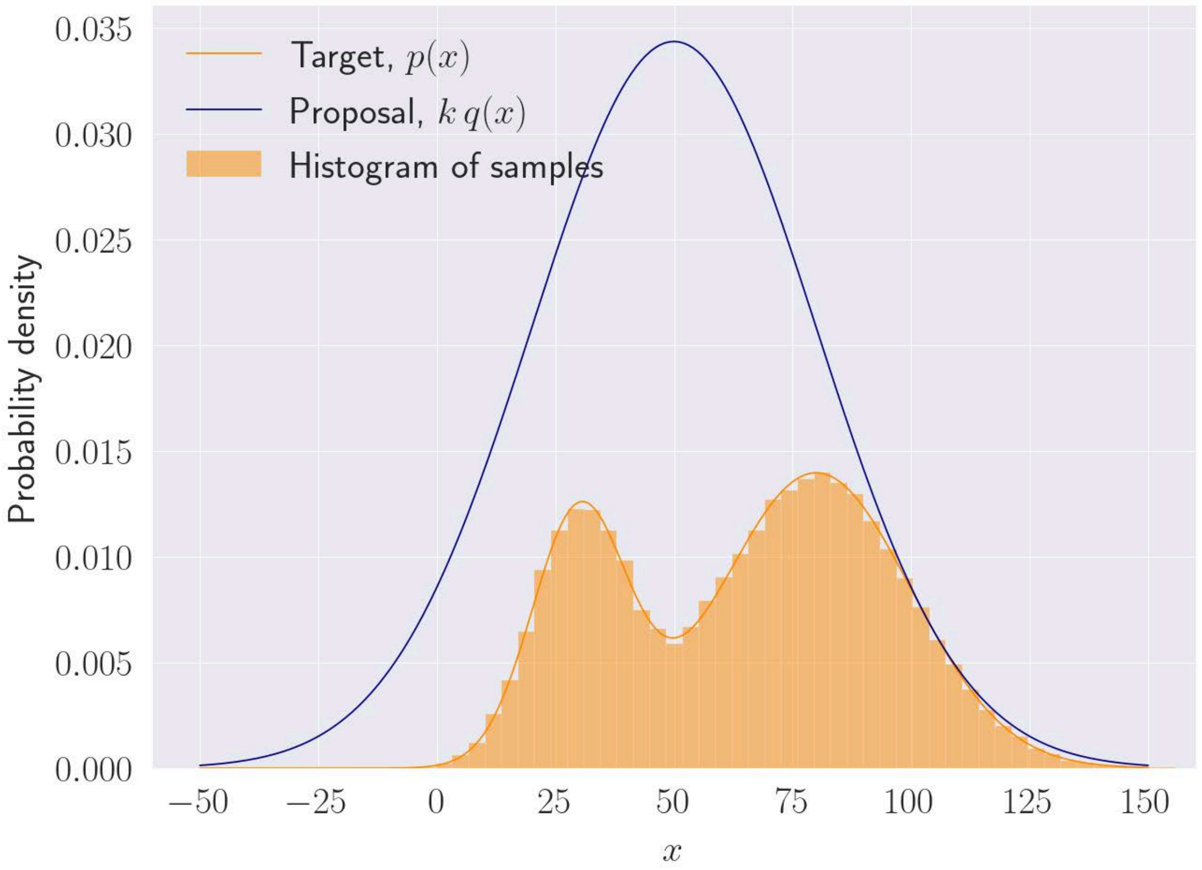
Rejection Sampling

Acceptance Criterion:

$$\alpha = \frac{p(\tilde{\mathbf{z}})}{\underbrace{k}_{\text{Scaling constant s.th. } k\,q(\mathbf{z}) \geq p(\mathbf{z})}\,\underbrace{q(\tilde{\mathbf{z}})}_{\text{Proposal distribution}}}, \quad \text{where } \tilde{\mathbf{z}} \sim p(\mathbf{z})$$

$$p(\text{accept}) = \mathbb{E}_{q(\mathbf{z})}\left[\frac{p(\mathbf{z})}{k\,q(\mathbf{z})}\right] = \int_{\mathcal{Z}} \frac{p(\mathbf{z})}{k\,q(\mathbf{z})}\,q(\mathbf{z})\,d\mathbf{z}$$

$$= \frac{1}{k}\int_{\mathcal{Z}} p(\mathbf{z})\,d\mathbf{z}$$



Number of rejected samples ∝ area between curves. Keep k as small as possible, s.th. $k\,q(\mathbf{z}) \geq p(\mathbf{z})$!

Conclusion: Rejection Sampling

Assumptions:

- Target distribution, $p(\mathbf{z})$, is difficult to sample from, but can be evaluated directly
- Sample from proposal distribution, $q(\mathbf{z})$

Advantages:

- Easy to implement + fast if $\mathbf{z}$ is one- or two-dimensional

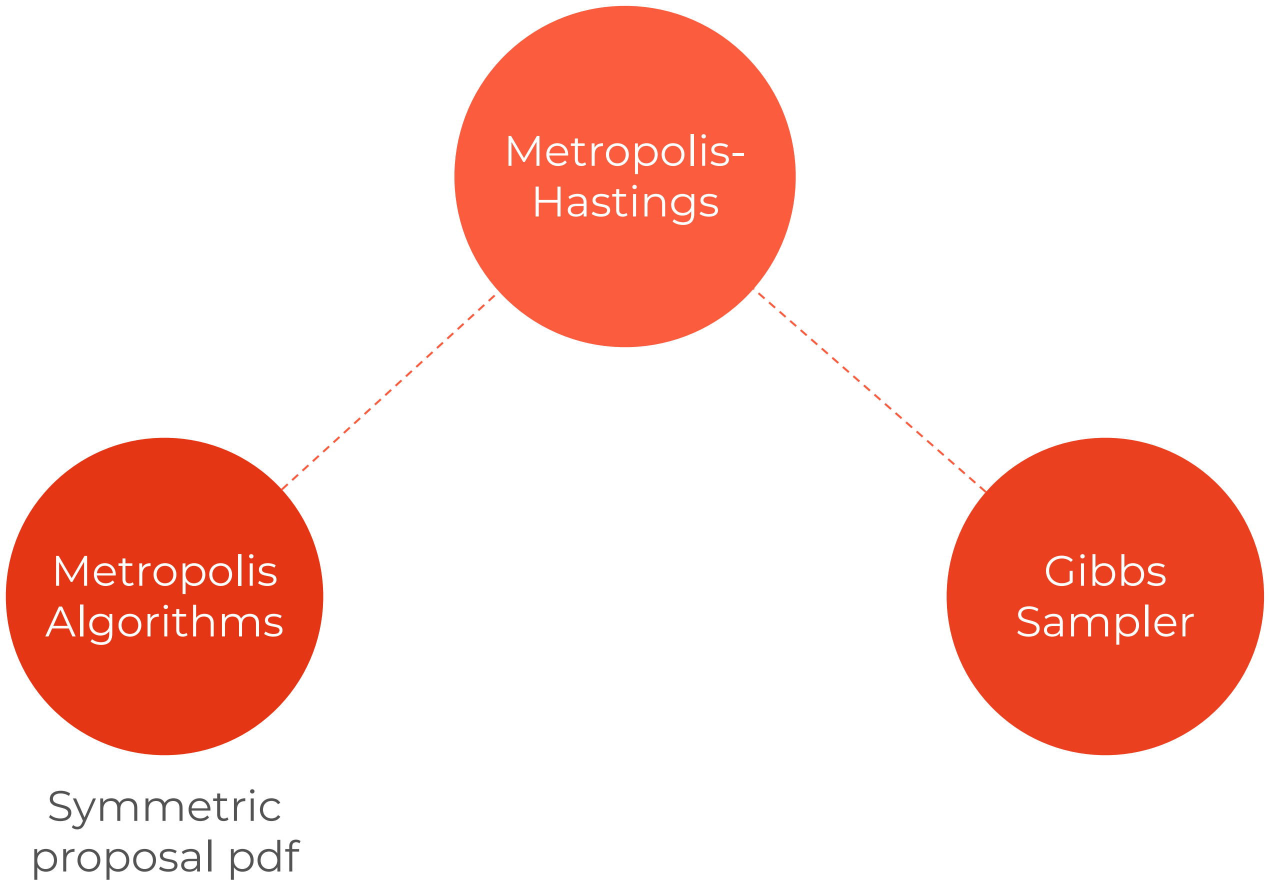
Limitations:

- Multimodal distributions: Difficult to find good proposal distribution (high rejection rate)
- Exponential decrease in acceptance rate with increasing dimensionality of $\mathbf{z}$

Markov Chain Monte Carlo:
Metropolis-Hastings

Markov Chain Monte Carlo

Principle: Use Markov chains to sample from a given distribution



Primer: Markov Chains

First-order Markov chain:

$$p(\mathbf{z}^{(m+1)} \mid \mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}) = p(\mathbf{z}^{(m+1)} \mid \mathbf{z}^{(m)}) \triangleq T_m(\mathbf{z}^{(m)} \mid \mathbf{z}^{(m+1)})$$

Transition operator

Homogeneous Markov chain: $T_1(\cdot) = \ldots = T_m(\cdot) \triangleq T(\cdot)$

Metropolis-Hastings Algorithm

Metropolis-Hastings Algorithm:

For $\ell = 1,\ldots,L$:

1. Draw candidate: $\tilde{\mathbf{z}} \sim q(\mathbf{z} \mid \mathbf{z}^{(\ell)})$
2. Draw $u \sim \mathcal{U}\left[0, 1\right]$
3. Evaluate $\alpha = \dfrac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}$

if $u \leq \alpha$.

Accept $\hat{\mathbf{z}}^{(\ell+1)} = \tilde{\mathbf{z}}$

else:

Accept $\hat{\mathbf{z}}^{(\ell+1)} = \mathbf{z}^{(\ell)}$

$p(\,\cdot\,)$: Target pdf, $q(\,\cdot\,)$: Proposal pdf

$\mathbf{z}^{(\ell)}$: Current accepted sample

W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *in Biometrika*, 57(1), 1970. DOI:10.1093/biomet/57.1.97

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})\ d\mathbf{z}}$$

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})\; d\mathbf{z}}$$

Assume $p(\mathbf{x} \mid \mathbf{z})$ and $p(\mathbf{z})$ are non-Gaussian $\rightarrow$ $p(\mathbf{z} \mid \mathbf{x})$ generally intractable

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})\; d\mathbf{z}}$$

Assume $p(\mathbf{x} \mid \mathbf{z})$ and $p(\mathbf{z})$ are non-Gaussian $\rightarrow$ $p(\mathbf{z} \mid \mathbf{x})$ generally intractable

Approximate posterior pdf by sampling from $p(\mathbf{z} \mid \mathbf{x})$:

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}} \mid \mathbf{x})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)} \mid \mathbf{x})}$$

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})\; d\mathbf{z}}$$

Assume $p(\mathbf{x} \mid \mathbf{z})$ and $p(\mathbf{z})$ are non-Gaussian $\rightarrow$ $p(\mathbf{z} \mid \mathbf{x})$ generally intractable

Approximate posterior pdf by sampling from $p(\mathbf{z} \mid \mathbf{x})$:

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}} \mid \mathbf{x})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)} \mid \mathbf{x})}$$

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \,|\, \mathbf{x}) = \frac{p(\mathbf{x} \,|\, \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \,|\, \mathbf{z})\, p\,(\mathbf{z})\; d\mathbf{z}}$$

Assume $p(\mathbf{x} \,|\, \mathbf{z})$ and $p(\mathbf{z})$ are non-Gaussian $\rightarrow$ $p(\mathbf{z} \,|\, \mathbf{x})$ generally intractable

Approximate posterior pdf by sampling from $p(\mathbf{z} \,|\, \mathbf{x})$:

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}} \,|\, \mathbf{x})}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)} \,|\, \mathbf{x})} = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, \frac{p(\mathbf{x} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{p(\mathbf{x})}}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, \frac{p(\mathbf{x} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}{p(\mathbf{x})}}$$

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})\; d\mathbf{z}}$$

Assume $p(\mathbf{x} \mid \mathbf{z})$ and $p(\mathbf{z})$ are non-Gaussian $\rightarrow$ $p(\mathbf{z} \mid \mathbf{x})$ generally intractable

Approximate posterior pdf by sampling from $p(\mathbf{z} \mid \mathbf{x})$:

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}} \mid \mathbf{x})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)} \mid \mathbf{x})} = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, \frac{p(\mathbf{x} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{p(\mathbf{x})}}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, \frac{p(\mathbf{x} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}{p(\mathbf{x})}}$$

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \mid \mathbf{z})\, p\,(\mathbf{z})\; d\mathbf{z}}$$

Assume $p(\mathbf{x} \mid \mathbf{z})$ and $p(\mathbf{z})$ are non-Gaussian $\rightarrow$ $p(\mathbf{z} \mid \mathbf{x})$ generally intractable

Approximate posterior pdf by sampling from $p(\mathbf{z} \mid \mathbf{x})$:

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}} \mid \mathbf{x})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)} \mid \mathbf{x})} = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, \frac{p(\mathbf{x} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{p(\mathbf{x})}}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, \frac{p(\mathbf{x} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}{p(\mathbf{x})}} = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\mathbf{x} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{x} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}$$

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \,|\, \mathbf{x}) = \frac{p(\mathbf{x} \,|\, \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \,|\, \mathbf{z})\, p\,(\mathbf{z})\; d\mathbf{z}}$$

Assume $p(\mathbf{x} \,|\, \mathbf{z})$ and $p(\mathbf{z})$ are non-Gaussian $\rightarrow$ $p(\mathbf{z} \,|\, \mathbf{x})$ generally intractable

Approximate posterior pdf by sampling from $p(\mathbf{z} \,|\, \mathbf{x})$:

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}} \,|\, \mathbf{x})}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)} \,|\, \mathbf{x})} = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, \frac{p(\mathbf{x} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{p(\mathbf{x})}}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, \frac{p(\mathbf{x} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}{p(\mathbf{x})}} = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, p(\mathbf{x} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{x} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}$$

Application: Intractable Posterior

Bayes's Theorem:

$$p(\mathbf{z} \,|\, \mathbf{x}) = \frac{p(\mathbf{x} \,|\, \mathbf{z})\, p\,(\mathbf{z})}{\int_{\mathcal{Z}} p(\mathbf{x} \,|\, \mathbf{z})\, p\,(\mathbf{z})\; d\mathbf{z}}$$

Assume $p(\mathbf{x} \,|\, \mathbf{z})$ and $p(\mathbf{z})$ are non-Gaussian $\rightarrow$ $p(\mathbf{z} \,|\, \mathbf{x})$ generally intractable

Approximate posterior pdf by sampling from $p(\mathbf{z} \,|\, \mathbf{x})$:

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}} \,|\, \mathbf{x})}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)} \,|\, \mathbf{x})} = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}}) \frac{p(\mathbf{x} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{p(\mathbf{x})}}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)}) \frac{p(\mathbf{x} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}{p(\mathbf{x})}} = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, p(\mathbf{x} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{x} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}$$

Metropolis-Hastings allows us to approximate analytically intractable pdfs!

Application: Intractable Posterior

Metropolis-Hastings Algorithm:

For $\ell = 1, \ldots, L$:

1. Draw candidate: $\tilde{z} \sim q(z \mid z^{(\ell)})$
2. Draw $u \sim \mathcal{U}\left[0, 1\right]$
3. Evaluate $\alpha = \dfrac{q(z^{(\ell)} \mid \tilde{z})\, p(\mathbf{x} \mid \tilde{z})\, p(\tilde{z})}{q(\tilde{z} \mid z^{(\ell)})\, p(\mathbf{x} \mid z^{(\ell)})\, p(z^{(\ell)})}$

if $u \leq \alpha$.

Accept $\hat{z}^{(\ell+1)} = \tilde{z}$

else:

Accept $\hat{z}^{(\ell+1)} = z^{(\ell)}$

$p(\cdot)$: Target pdf, $q(\cdot)$: Proposal pdf

$\mathbf{z}^{(\ell)}$: Current accepted sample

State-Space Model

Dataset of i.i.d. datapoints: $p(\mathbf{x} \mid z) = \prod_{n=1}^{N} p(x_n \mid z)$

Likelihood: $p(x_n \mid z) = \mathcal{M}\left(x_n \mid z, \kappa_l\right)$

Prior: $p(z) = \mathcal{M}\left(z \mid \mu_0, \kappa_0\right)$

Posterior: $p(z \mid \mathbf{x}) = ?$

Proposal: $q(z \mid z^{(\ell)}) = \mathcal{M}\left(z \mid z^{(\ell)}, \kappa_p\right)$

Von Mises Distribution:

$$p(x \mid \mathbf{z}) = \frac{\exp\left\{\kappa \cos(x - \mu)\right\}}{2\pi\, I_0(\kappa)}$$

$-\pi \leq x \leq \pi$: Angle, $\kappa > 0$: Concentration,

$I_0(\cdot)$: Modified Bessel function of order 0

Application: Intractable Posterior

Metropolis-Hastings Algorithm:

For $\ell = 1, \ldots, L$:

1. Draw candidate: $\tilde{z} \sim q(z \mid z^{(\ell)})$
2. Draw $u \sim \mathcal{U}\left[0, 1\right]$
3. Evaluate $\alpha = \dfrac{q(z^{(\ell)} \mid \tilde{z})\, p(\mathbf{x} \mid \tilde{z})\, p(\tilde{z})}{q(\tilde{z} \mid z^{(\ell)})\, p(\mathbf{x} \mid z^{(\ell)})\, p(z^{(\ell)})}$

if $u \leq \alpha$.

Accept $\hat{z}^{(\ell+1)} = \tilde{z}$

else:

Accept $\hat{z}^{(\ell+1)} = z^{(\ell)}$

$p(\cdot)$: Target pdf, $q(\cdot)$: Proposal pdf

$\mathbf{z}^{(\ell)}$: Current accepted sample

Metropolis Algorithm

Metropolis-Hastings: Acceptance Criterion

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}, \quad \text{where } \tilde{\mathbf{z}} \sim q(\mathbf{z} \,|\, \mathbf{z}^{(\ell)})$$

$\mathbf{z}^{(\ell)}$: Current accepted sample, $q(\,\cdot\,)$: Proposal, $p(\,\cdot\,)$: Target

N. Metropolis *et al.*, "Equation of State Calculations by Fast Computing Machines," *in J. Chemical Physics*, 21(6), 1953. DOI:10.1063/1.1699114

Metropolis Algorithm

UNIVERSITY OF Southampton

Metropolis-Hastings: Acceptance Criterion

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}, \quad \text{where } \tilde{\mathbf{z}} \sim q(\mathbf{z} \mid \mathbf{z}^{(\ell)})$$

$\mathbf{z}^{(\ell)}$: Current accepted sample, $q(\cdot)$: Proposal, $p(\cdot)$: Target

If the proposal is a symmetric distribution: $q(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = q(\mathbf{z}^{(\ell)} \mid \mathbf{z})$

N. Metropolis *et al.*, "Equation of State Calculations by Fast Computing Machines," *in J. Chemical Physics*, 21(6), 1953. DOI:10.1063/1.1699114

Metropolis Algorithm

University of Southampton

Metropolis-Hastings: Acceptance Criterion

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}, \quad \text{where } \tilde{\mathbf{z}} \sim q(\mathbf{z} \,|\, \mathbf{z}^{(\ell)})$$

$\mathbf{z}^{(\ell)}$: Current accepted sample, $q(\,\cdot\,)$: Proposal, $p(\,\cdot\,)$: Target

If the proposal is a symmetric distribution: $q(\mathbf{z} \,|\, \mathbf{z}^{(\ell)}) = q(\mathbf{z}^{(\ell)} \,|\, \mathbf{z})$

Example:

$$p(\mathbf{z} \,|\, \mathbf{z}^{(\ell)}) = \frac{1}{(2\pi)^{\frac{Q}{2}} \,|\, \boldsymbol{\Sigma} \,|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\mathbf{z} - \mathbf{z}^{(\ell)}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{z} - \mathbf{z}^{(\ell)}\right) \right\}$$

N. Metropolis *et al.*, "Equation of State Calculations by Fast Computing Machines," *in J. Chemical Physics*, 21(6), 1953. DOI:10.1063/1.1699114

Metropolis Algorithm

Metropolis-Hastings: Acceptance Criterion

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \,|\, \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \,|\, \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}, \quad \text{where } \tilde{\mathbf{z}} \sim q(\mathbf{z} \,|\, \mathbf{z}^{(\ell)})$$

$\mathbf{z}^{(\ell)}$: Current accepted sample, $q(\,\cdot\,)$: Proposal, $p(\,\cdot\,)$: Target

If the proposal is a symmetric distribution: $q(\mathbf{z} \,|\, \mathbf{z}^{(\ell)}) = q(\mathbf{z}^{(\ell)} \,|\, \mathbf{z})$

Example:

$$p(\mathbf{z} \,|\, \mathbf{z}^{(\ell)}) = \frac{1}{(2\pi)^{\frac{Q}{2}} \,|\, \boldsymbol{\Sigma} \,|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\mathbf{z} - \mathbf{z}^{(\ell)}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{z} - \mathbf{z}^{(\ell)}\right) \right\} = \frac{1}{(2\pi)^{\frac{Q}{2}} \,|\, \boldsymbol{\Sigma} \,|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\mathbf{z}^{(\ell)} - \mathbf{z}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{z}^{(\ell)} - \mathbf{z}\right) \right\} = p(\mathbf{z}^{(\ell)} \,|\, \mathbf{z})$$

N. Metropolis *et al.*, “Equation of State Calculations by Fast Computing Machines,” *in J. Chemical Physics*, 21(6), 1953. DOI:10.1063/1.1699114

Metropolis Algorithm

Metropolis-Hastings: Acceptance Criterion

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}, \quad \text{where } \tilde{\mathbf{z}} \sim q(\mathbf{z} \mid \mathbf{z}^{(\ell)})$$

$\mathbf{z}^{(\ell)}$: Current accepted sample, $q(\,\cdot\,)$: Proposal, $p(\,\cdot\,)$: Target

If the proposal is a symmetric distribution: $q(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = q(\mathbf{z}^{(\ell)} \mid \mathbf{z})$

Example:

$$p(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = \frac{1}{(2\pi)^{\frac{Q}{2}} \mid \boldsymbol{\Sigma} \mid^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\mathbf{z} - \mathbf{z}^{(\ell)}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{z} - \mathbf{z}^{(\ell)}\right) \right\} = \frac{1}{(2\pi)^{\frac{Q}{2}} \mid \boldsymbol{\Sigma} \mid^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\mathbf{z}^{(\ell)} - \mathbf{z}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{z}^{(\ell)} - \mathbf{z}\right) \right\} = p(\mathbf{z}^{(\ell)} \mid \mathbf{z})$$

N. Metropolis *et al.*, “Equation of State Calculations by Fast Computing Machines,” *in J. Chemical Physics*, 21(6), 1953. DOI:10.1063/1.1699114

Metropolis Algorithm

University of Southampton

Metropolis-Hastings: Acceptance Criterion

$$\alpha = \frac{q(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}})\, p(\tilde{\mathbf{z}})}{q(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}, \quad \text{where } \tilde{\mathbf{z}} \sim q(\mathbf{z} \mid \mathbf{z}^{(\ell)})$$

$\mathbf{z}^{(\ell)}$: Current accepted sample, $q(\cdot)$: Proposal, $p(\cdot)$: Target

Metropolis: Acceptance Criterion

$$\alpha = \frac{p(\tilde{\mathbf{z}})}{p(\mathbf{z}^{(\ell)})}, \quad \text{where } \tilde{\mathbf{z}} \sim q(\mathbf{z} \mid \mathbf{z}^{(\ell)})$$

$\mathbf{z}^{(\ell)}$: Current accepted sample, $q(\cdot)$: Proposal, $p(\cdot)$: Target

If the proposal is a symmetric distribution: $q(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = q(\mathbf{z}^{(\ell)} \mid \mathbf{z})$

Example:

$$p(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = \frac{1}{(2\pi)^{\frac{Q}{2}} \mid \boldsymbol{\Sigma} \mid^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\mathbf{z} - \mathbf{z}^{(\ell)}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{z} - \mathbf{z}^{(\ell)}\right) \right\} = \frac{1}{(2\pi)^{\frac{Q}{2}} \mid \boldsymbol{\Sigma} \mid^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\mathbf{z}^{(\ell)} - \mathbf{z}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{z}^{(\ell)} - \mathbf{z}\right) \right\} = p(\mathbf{z}^{(\ell)} \mid \mathbf{z})$$

N. Metropolis *et al.*, “Equation of State Calculations by Fast Computing Machines,” *in J. Chemical Physics*, 21(6), 1953. DOI:10.1063/1.1699114

Markov Chain Monte Carlo:
Gibbs Sampling

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})$:

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}^{(\ell)}_{\setminus k})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right) \, p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right) \, p\left(\mathbf{z}^{(\ell)}\right)}$$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right) p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right) p\left(\mathbf{z}^{(\ell)}\right)}$$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right)\ p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right)\ p\left(\mathbf{z}^{(\ell)}\right)} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{\mathbf{z}})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}$$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right)\, p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right)\, p\left(\mathbf{z}^{(\ell)}\right)} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{\mathbf{z}})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}^{(\ell)})}$$

Chain rule: $p(\mathbf{z}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\backslash k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\backslash k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\backslash k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right)\ p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right)\ p\left(\mathbf{z}^{(\ell)}\right)} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\tilde{\mathbf{z}})}{p(\tilde{z}_k \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\mathbf{z}^{(\ell)})} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\tilde{z}_k \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\mathbf{z}_{\backslash k}^{(\ell)})}{p(\tilde{z}_k \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(z_k^{(\ell)} \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\mathbf{z}_{\backslash k}^{(\ell)})}$$

Gibbs Sampling

UNIVERSITY OF Southampton

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right)\, p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right)\, p\left(\mathbf{z}^{(\ell)}\right)} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{\mathbf{z}})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}^{(\ell)})} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})}$$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \dots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \dots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \dots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \dots, z_{k-1}, z_{k+1}, \dots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right)\, p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right)\, p\left(\mathbf{z}^{(\ell)}\right)} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{\mathbf{z}})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}^{(\ell)})} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})}$$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\backslash k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\backslash k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\backslash k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right)\, p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right)\, p\left(\mathbf{z}^{(\ell)}\right)} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\tilde{\mathbf{z}})}{p(\tilde{z}_k \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\mathbf{z}^{(\ell)})} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\tilde{z}_k \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\mathbf{z}_{\backslash k}^{(\ell)})}{p(\tilde{z}_k \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(z_k^{(\ell)} \mid \mathbf{z}_{\backslash k}^{(\ell)})\, p(\mathbf{z}_{\backslash k}^{(\ell)})}$$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right)\, p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right)\, p\left(\mathbf{z}^{(\ell)}\right)} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{\mathbf{z}})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}^{(\ell)})} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})} = 1$$

Gibbs Sampling

Aim:

Sample from a high-dimensional joint pdf, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$

Assumptions:

- Joint distribution, $p(\mathbf{z}) = p(z_1, \ldots, z_K)$ cannot be sampled from
- Conditional distributions, $p(z_k \mid \mathbf{z}_{\setminus k})$, can be sampled from easily

where $\mathbf{z} = [z_1, \ldots, z_K]^T$ and $\mathbf{z}_{\setminus k} = [z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K]^T$

Metropolis-Hastings:

Sample z_k from $q_k(\mathbf{z} \mid \mathbf{z}^{(\ell)}) = p(z_k \mid \mathbf{z}_{\setminus k}^{(\ell)})$:

$$\alpha = \frac{q_k\left(\mathbf{z}^{(\ell)} \mid \tilde{\mathbf{z}}\right)\, p\left(\tilde{\mathbf{z}}\right)}{q_k\left(\tilde{\mathbf{z}} \mid \mathbf{z}^{(\ell)}\right)\, p\left(\mathbf{z}^{(\ell)}\right)} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{\mathbf{z}})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}^{(\ell)})} = \frac{p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})}{p(\tilde{z}_k \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(z_k^{(\ell)} \mid \mathbf{z}_{\setminus k}^{(\ell)})\, p(\mathbf{z}_{\setminus k}^{(\ell)})} = 1$$

By sampling from conditionals, the Metropolis-Hastings steps are always accepted.

Gibbs Sampler

UNIVERSITY OF Southampton

Gibbs Sampler:

For $\ell = 1,\ldots,L$:

1. Sample $z_1^{(\ell+1)} \sim p(z_1 \,|\, z_2^{(\ell)}, \ldots, z_n^{(\ell)})$

2. Sample $z_2^{(\ell+1)} \sim p(z_2 \,|\, z_1^{(\ell+1)}, z_3^{(\ell)}, \ldots, z_n^{(\ell)})$

3. …

4. Sample $z_K^{(\ell+1)} \sim p(z_K \,|\, z_1^{(\ell+1)}, , \ldots, z_{K-1}^{(\ell+1)})$

Conclusion: MCMC

UNIVERSITY OF Southampton

Assumptions:

- Target distribution, $p(\mathbf{z})$, is difficult to sample from (and may not be known in closed form)
- Sample from proposal pdf, $q(\mathbf{z} \mid \mathbf{z}^{(\ell)}) \rightarrow$ Sequence of samples forms a Markov chain

Advantages:

- Allows sampling from a large class of distributions
- Scales well with the dimensionality of the sample space

Limitations:

- Successive samples are highly correlated
  - Retain only each M^{th} sample and discard rest of sequence, where $M \gg 1$
- 'Burn-in' period

Learning Outcomes

Following this week's lecture on Markov Chain Monte Carlo methods, you should be able to:

1) **Explain how random sampling can be used to approximate integrals;**
2) **Understand how MCMC methods use Markov chains to sample from target distributions;**
3) **Apply different techniques for MCMC to real-world problems.**

What's Next

Tuesday (2 March):
Q&A Markov Chain Monte Carlo
Lab Worksheet Markov Chain Monte Carlo

Week 6: Sequential Monte Carlo
Online learning for sequential data