DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF COPENHAGEN

Sampling and Markov Chain Monte Carlo (MCMC) Methods, Part II

Kim Steenstrup Pedersen

Plan for today

- Last lecture:
  - Rejection sampling
  - Importance sampling
  - Sampling-Importance-Resampling (SIR)
- These methods generate independent samples

- Markov Chain Monte Carlo (MCMC) methods
  - Metropolis algorithm
  - Metropolis-Hastings algorithm
  - Gibbs sampler

- More on sampling Bayesian networks

Remember: Proposal distributions

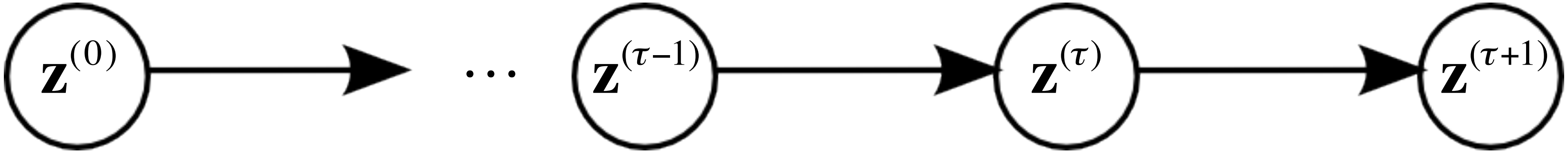
- Consider the distribution: $p(\mathbf{z}) = \tilde{p}(\mathbf{z}) \big/ Z_p$

- It may be difficult to sample from the distribution $p(\mathbf{z})$.
- Often $Z_p = \int \tilde{p}(\mathbf{z}) d\mathbf{z}$ is difficult to compute, but $\tilde{p}(\mathbf{z})$ may be evaluated for any $\boldsymbol{z}$.

- Common strategy (used in rejection sampling, importance sampling, Metropolis-Hastings, etc.):
  - Use a much simpler proposal distribution $q(z)$ from which we can sample.
  - Generate a proposal sample and evaluate an acceptance criterion for the sample.

Markov Chain Monte Carlo (MCMC) methods

A discrete time Markov chain?



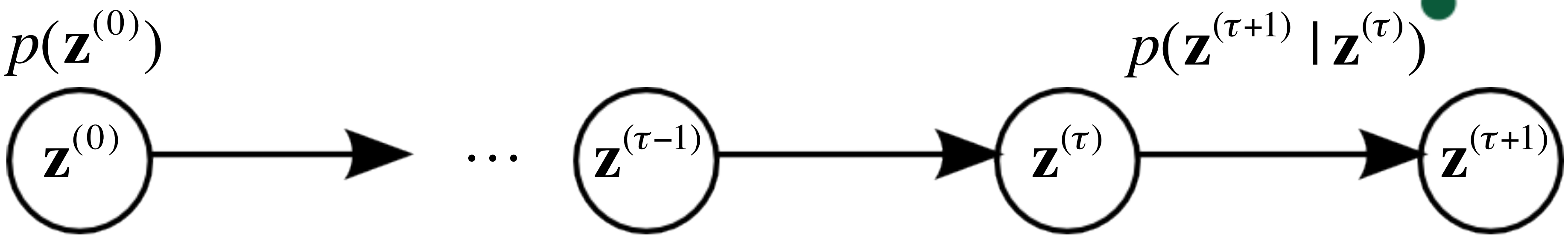
- A discrete time stochastic process is a first order Markov chain (MC), if

$$p(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)}, \mathbf{z}^{(\tau-1)}, \ldots, \mathbf{z}^{(0)}) = p_\tau(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)}) \quad \text{(Markov property)}$$

- Its Homogeneous (stationary), if $p_\tau(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)}) = p(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)})$
- Its fully specified by $p(\mathbf{z}^{(0)})$ and transition $p(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)})$
- Taking τ steps:

$$p(\mathbf{z}^{(0:\tau)}) = p(\mathbf{z}^{(0)}, \ldots, \mathbf{z}^{(\tau)}) = p(\mathbf{z}^{(0)}) \prod_{n=1}^{\tau} p(\mathbf{z}^{(n)} \mid \mathbf{z}^{(n-1)})$$

A discrete time Markov chain with discrete states



- For discrete state space $\mathbf{z}^{(\tau)} \in \Omega = \{1, 2, ..., L\}$ for all τ :

  Transition matrix $\mathbf{P} = \{p_{ij}\}_{i,j\in\Omega}$

  $$p_{ij} = p(\mathbf{z}^{(\tau+1)} = j \mid \mathbf{z}^{(\tau)} = i)$$

  Taking τ steps: $p(\mathbf{z}^{(1:\tau)}) = \mathbf{P}^{\tau} v$
  with initial state vector: $v_i = p(\mathbf{z}^{(0)} = i)$

Example:

$$\begin{array}{cc} & (j = 1 \ , \ j = 2) \\ \begin{pmatrix} i = 1 \\ i = 2 \end{pmatrix} & \begin{bmatrix} 0.2 & 0.8 \\ 0.9 & 0.1 \end{bmatrix} \end{array}$$

Discrete Markov chains as stochastic state machines – an example

- States: $\Omega = \{a, b, c\}$ and $\mathbf{z}^{(\tau)} \in \Omega$
- Initial state probabilities:

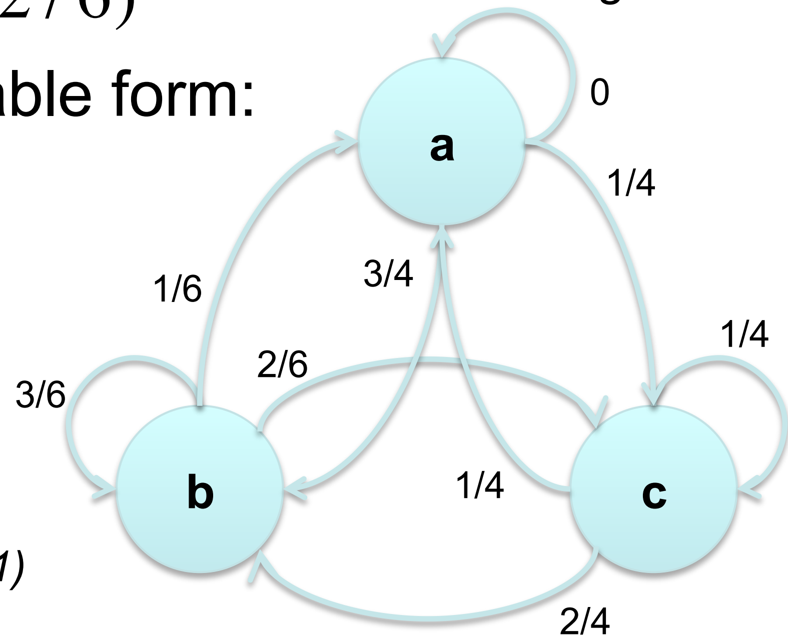
$$p(\mathbf{z}^{(0)} = a) = \frac{1}{6} \quad p(\mathbf{z}^{(0)} = b) = \frac{3}{6} \quad p(\mathbf{z}^{(0)} = c) = \frac{2}{6}$$

- Or in vector form $v = (1/6, 3/6, 2/6)^T$
- Transition probability matrix in table form:

$p(z^{(\tau+1)} \mid z^{(\tau)})$	$z^{(\tau+1)}=a$	$z^{(\tau+1)}=b$	$z^{(\tau+1)}=c$
$z^{(\tau)}=a$	0	3/4	1/4
$z^{(\tau)}=b$	1/6	3/6	2/6
$z^{(\tau)}=c$	1/4	2/4	1/4

Rows contain probabilities for $z^{(\tau+1)}$

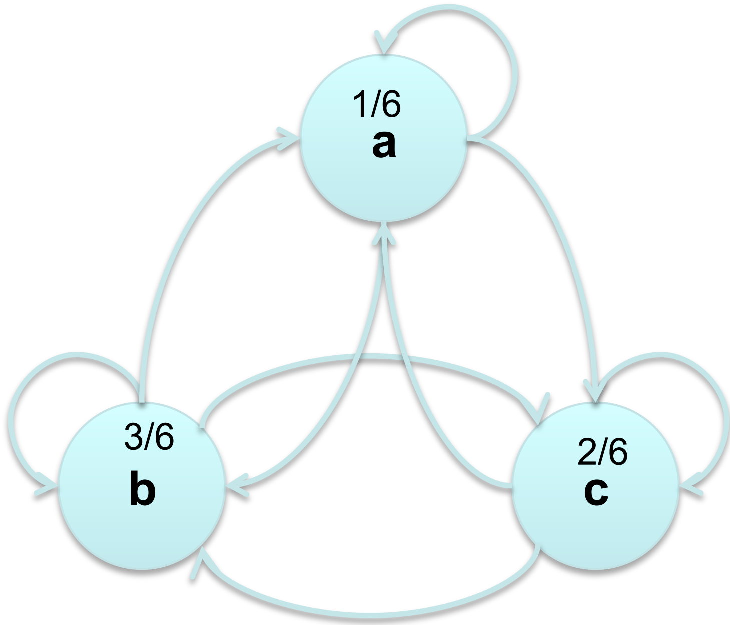
State diagram



Simulation of a Markov chain – an example

- Sample an initial state from $q(X_0)$

$q(X_0)$	X_0=a	X_0=b	X_0=c
	1/6	3/6	2/6

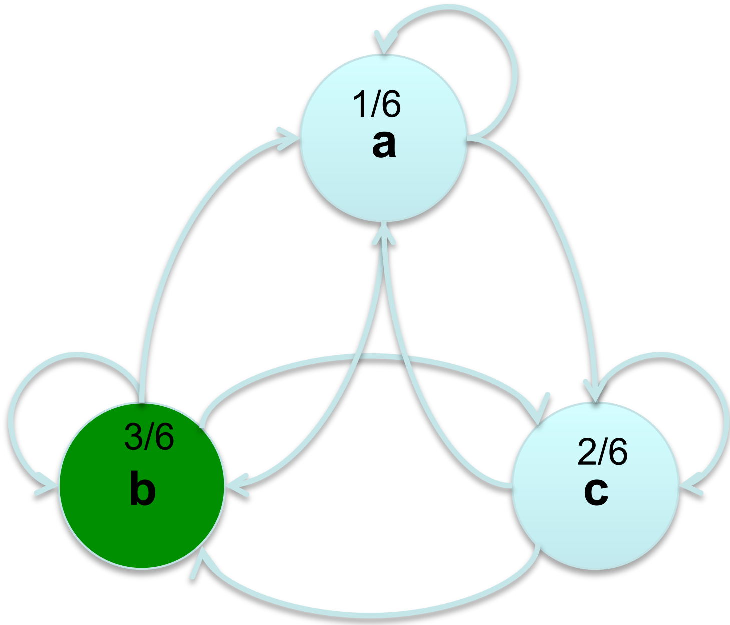


Result:

Simulation of a Markov chain – an example

- Sample an initial state from $q(X_0)$

$q(X_0)$	X_0=a	X_0=b	X_0=c
	1/6	3/6	2/6



Result:
b

Simulation of a Markov chain – an example

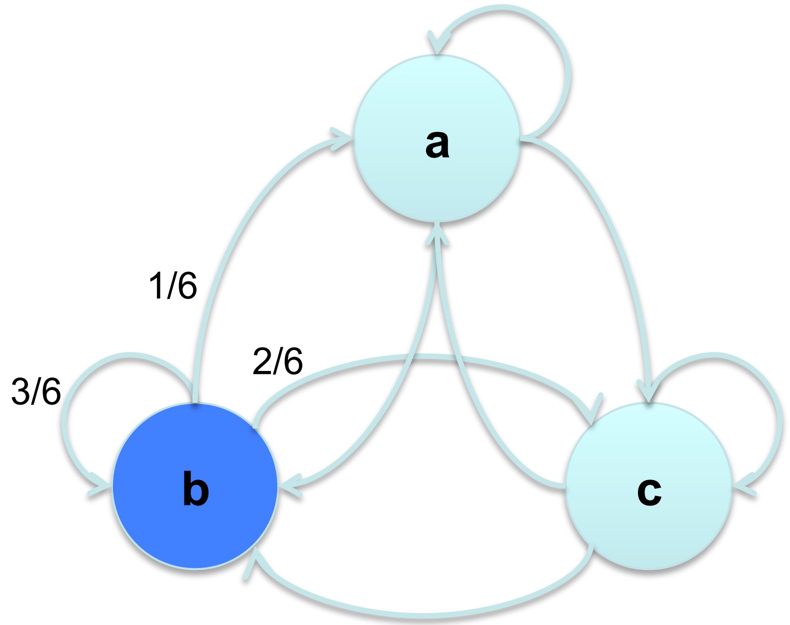
- Sample the next state from the transition probability $p(X_1 \mid X_0 = \mathbf{b})$

$p(X_t \mid X_{t-1})$	X_t=a	X_t=b	X_t=c
X_{t-1}=a	0	3/4	1/4
X_{t-1}=b	1/6	3/6	2/6
X_{t-1}=c	1/4	2/4	1/4

$p(X_1 \mid X_0)$	X_1=a	X_1=b	X_1=c
X_0=b	1/6	3/6	2/6

Result:

b

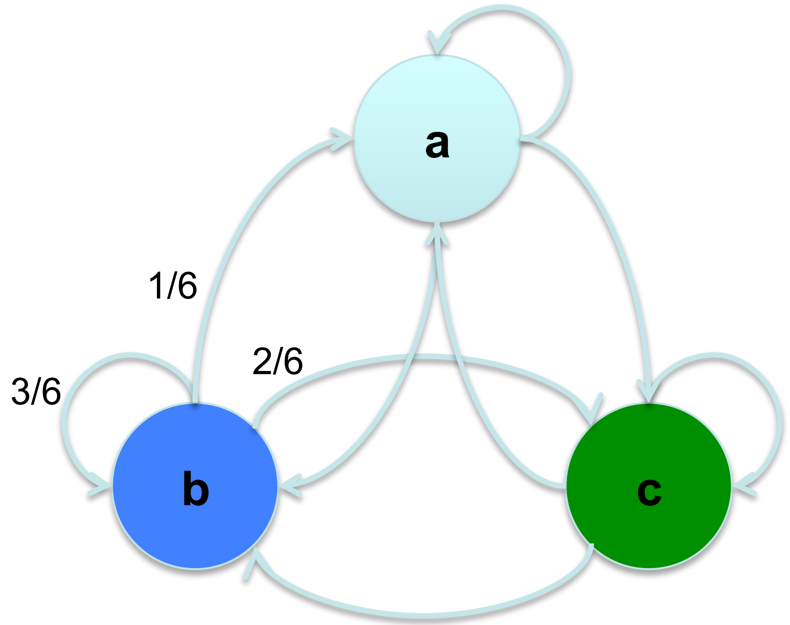


Simulation of a Markov chain – an example

- Sample the next state from the transition probability $p(X_1 \mid X_0 = \mathbf{b})$

$p(X_t \mid X_{t-1})$	X_t=a	X_t=b	X_t=c
X_{t-1}=a	0	3/4	1/4
X_{t-1}=b	1/6	3/6	2/6
X_{t-1}=c	1/4	2/4	1/4

$p(X_1 \mid X_0)$	X_1=a	X_1=b	X_1=c
X_0=b	1/6	3/6	2/6



Result:
bc

Simulation of a Markov chain – an example

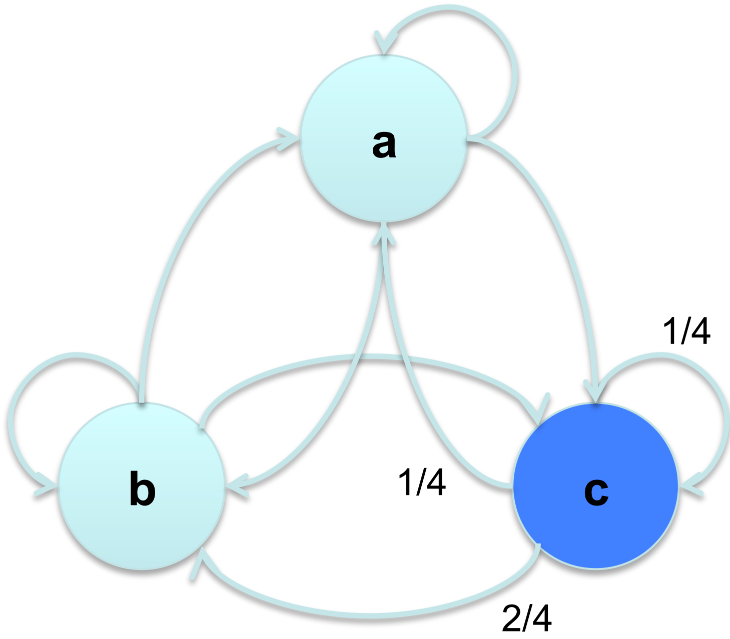
- Sample the next state from the transition probability $p(X_2 \mid X_1 = \mathbf{c})$

$p(X_t \mid X_{t-1})$	$X_t=a$	$X_t=b$	$X_t=c$
$X_{t-1}=a$	0	3/4	1/4
$X_{t-1}=b$	1/6	3/6	2/6
$X_{t-1}=c$	1/4	2/4	1/4

$p(X_2 \mid X_1)$	X_2=a	X_2=b	X_2=c
$X_1=c$	1/4	2/4	1/4

Result:

bc



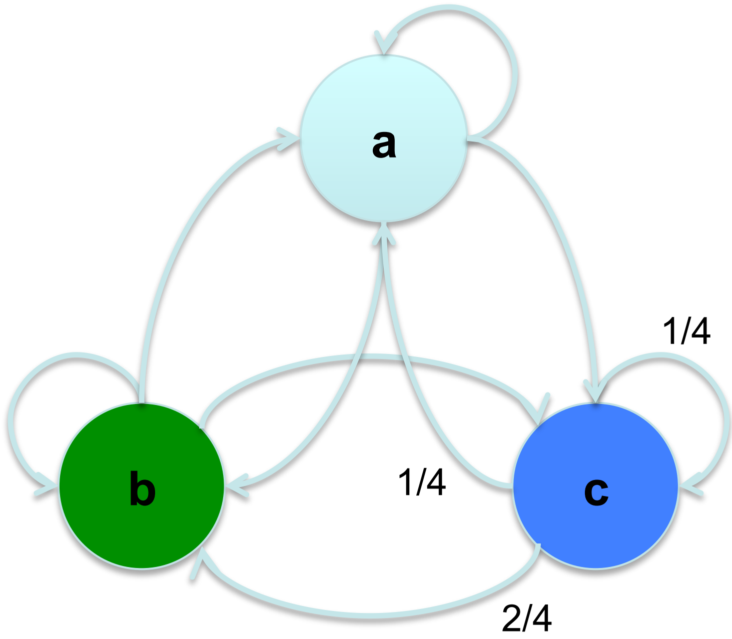
Simulation of a Markov chain – an example

- Sample the next state from the transition probability $p(X_2 \mid X_1 = \mathbf{c})$

$p(X_t \mid X_{t-1})$	X_t=a	X_t=b	X_t=c
X_{t-1}=a	0	3/4	1/4
X_{t-1}=b	1/6	3/6	2/6
X_{t-1}=c	1/4	2/4	1/4

$p(X_2 \mid X_1)$	X_2=a	X_2=b	X_2=c
X_1=c	1/4	2/4	1/4

Result:
bcb



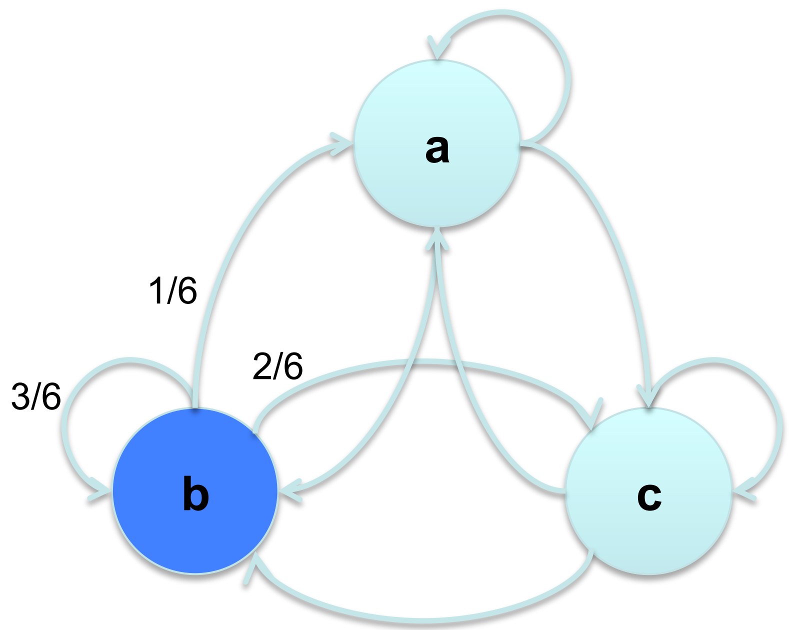
Simulation of a Markov chain – an example

- Sample the next state from the transition probability
  $p(X_3 \mid X_2 = \mathbf{b})$

$p(X_t \mid X_{t-1})$	X_t=a	X_t=b	X_t=c
X_{t-1}=a	0	3/4	1/4
X_{t-1}=b	1/6	3/6	2/6
X_{t-1}=c	1/4	2/4	1/4

$p(X_3 \mid X_2)$	X_3=a	X_3=b	X_3=c
X_2=b	1/6	3/6	2/6

Result:
bcb



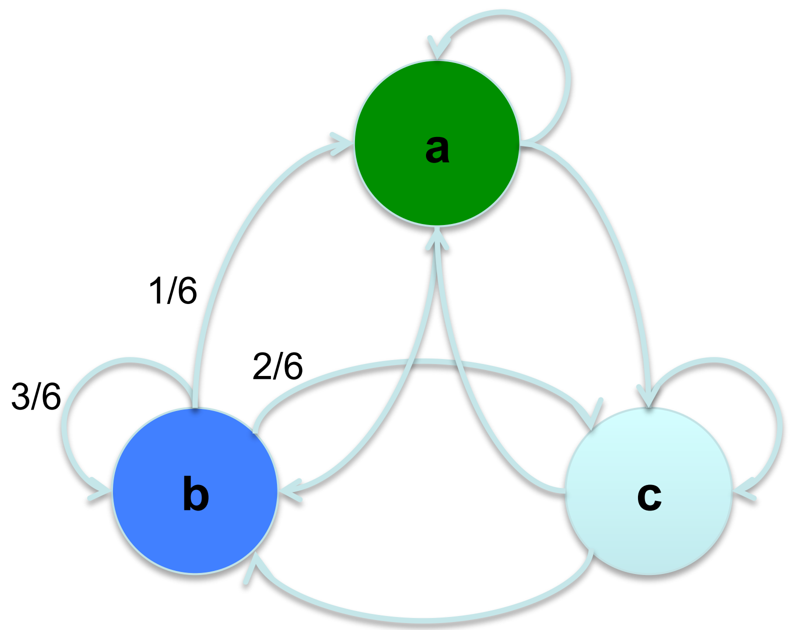
Simulation of a Markov chain – an example

- Sample the next state from the transition probability $p(X_3 \mid X_2 = \mathbf{b})$

$p(X_t \mid X_{t-1})$	$X_t=a$	$X_t=b$	$X_t=c$
$X_{t-1}=a$	0	3/4	1/4
$X_{t-1}=b$	1/6	3/6	2/6
$X_{t-1}=c$	1/4	2/4	1/4

$p(X_3 \mid X_2)$	X_3=a	X_3=b	X_3=c
$X_2=b$	1/6	3/6	2/6

Result:
bcba



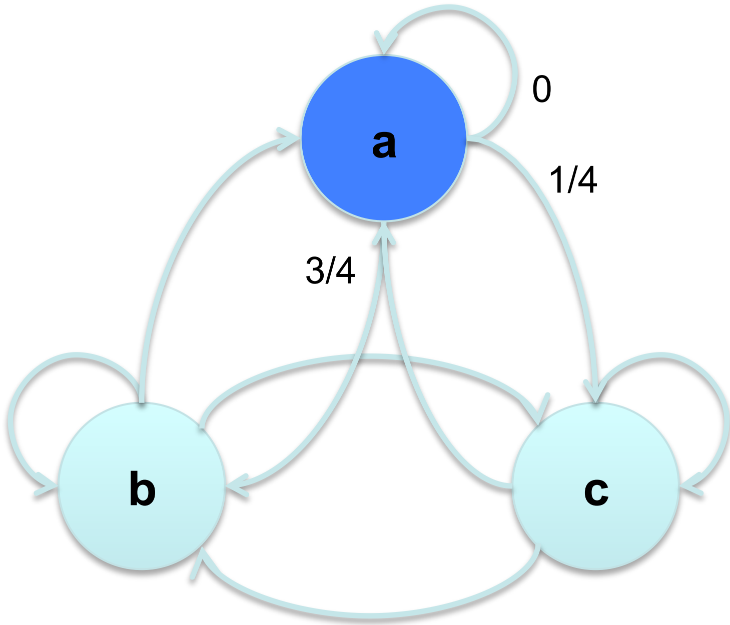
Simulation of a Markov chain – an example

- Sample the next state from the transition probability $p(X_4 \mid X_3 = \mathbf{a})$

$p(X_t \mid X_{t-1})$	X_t=a	X_t=b	X_t=c
X_{t-1}=a	0	3/4	1/4
X_{t-1}=b	1/6	3/6	2/6
X_{t-1}=c	1/4	2/4	1/4

$p(X_4 \mid X_3)$	X_4=a	X_4=b	X_4=c
X_3=a	0	3/4	1/4

Result:
bcba



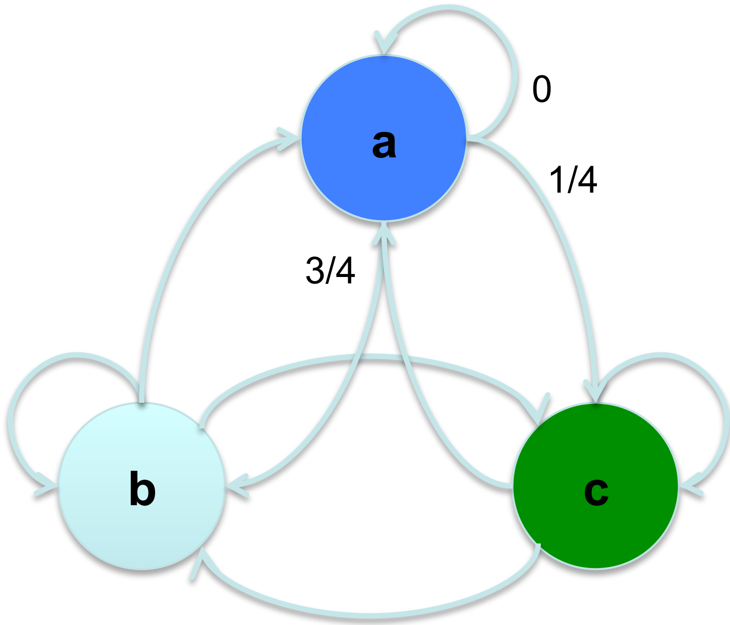
Simulation of a Markov chain – an example

- Sample the next state from the transition probability $p(X_4 \mid X_3 = \mathbf{a})$

$p(X_t \mid X_{t-1})$	X_t=a	X_t=b	X_t=c
X_{t-1}=a	0	3/4	1/4
X_{t-1}=b	1/6	3/6	2/6
X_{t-1}=c	1/4	2/4	1/4

$p(X_4 \mid X_3)$	X_4=a	X_4=b	X_4=c
X_3=a	0	3/4	1/4

Result:
bcbac



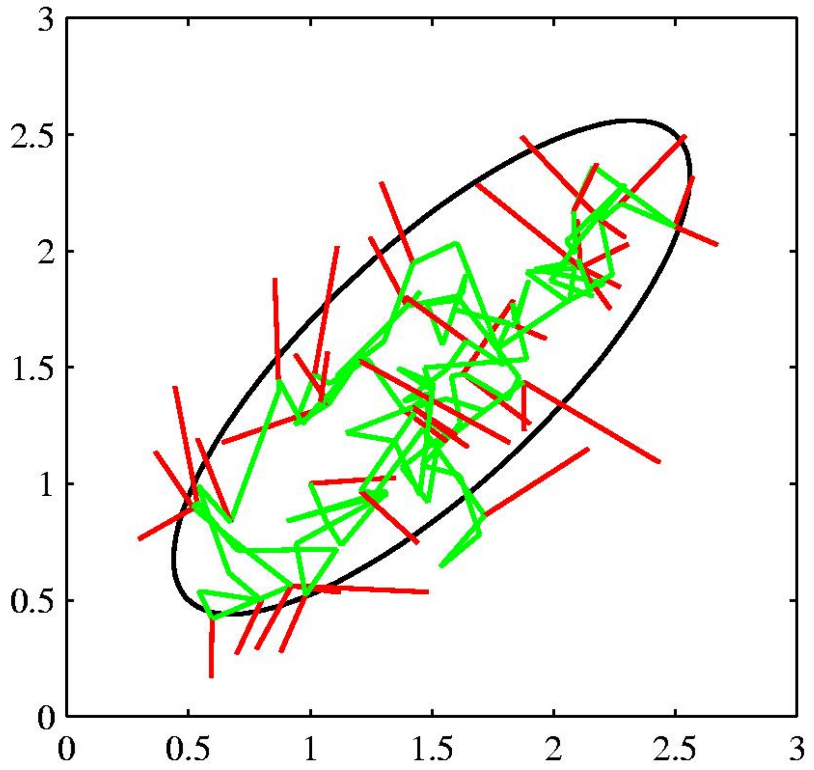
Sampling methods using Markov chains as proposal distribution

Markov Chain Monte Carlo (MCMC) sampling

- Let the proposal distribution be dependent on the current state $\mathbf{z}^{(\tau)}$, $q(\mathbf{z} \mid \mathbf{z}^{(\tau)})$.
- Samples $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \mathbf{z}^{(3)}, \ldots$ form a 1st order Markov chain with

$$q(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(\tau)}) = q(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)})$$

- MCMC methods explore state space by a random walk.
- Scales well with dimensionality.
- Consequence: Samples are not independent, but we can select subset of samples that are independent.



What properties would we like of the Markov chain?

- For any choice of distribution $q(\mathbf{z}^{(0)})$ of the starting point $\mathbf{z}^{(0)}$, we want the sample distribution to converge to the distribution $p(\mathbf{z})$ as $\tau \rightarrow \infty$.

- We also want to make sure that we can visit all parts of state space where $p(\mathbf{z}) > 0$. Otherwise MCMC will not be able to produce samples distributed according to $p(\mathbf{z})$.

More properties of Markov chains
(Lets assume discrete states for simplicity)

- *Transition probability* for a *homogeneous Markov chain*

$$p(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)})$$

- *Marginal* probability distribution:

$$p(\mathbf{z}^{(\tau+1)}) = \sum_{\mathbf{z}^{(\tau)}} p(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)}) p(\mathbf{z}^{(\tau)}) = \sum_{\mathbf{z}^{(\tau)}} p(\mathbf{z}^{(\tau+1)}, \mathbf{z}^{(\tau)})$$

- For an *invariant distribution* $p^*(\mathbf{z})$ it holds that: $p^*(\mathbf{z}) = \sum_{\mathbf{z}'} p(\mathbf{z} \mid \mathbf{z}') p^*(\mathbf{z}')$

- A sufficient condition for ensuring $p^*(\mathbf{z})$ is an invariant distribution of the Markov chain is given by choosing $p(\mathbf{z} \mid \mathbf{z}')$ to satisfy the *detailed balance* equation $p^*(\mathbf{z}) p(\mathbf{z}' \mid \mathbf{z}) = p^*(\mathbf{z}') p(\mathbf{z} \mid \mathbf{z}')$

$$\text{Proof}: \sum_{\mathbf{z}'} p^*(\mathbf{z}') p(\mathbf{z} \mid \mathbf{z}') = \sum_{\mathbf{z}'} p^*(\mathbf{z}) p(\mathbf{z}' \mid \mathbf{z}) = p^*(\mathbf{z}) \sum_{\mathbf{z}'} p(\mathbf{z}' \mid \mathbf{z}) = p^*(\mathbf{z})$$

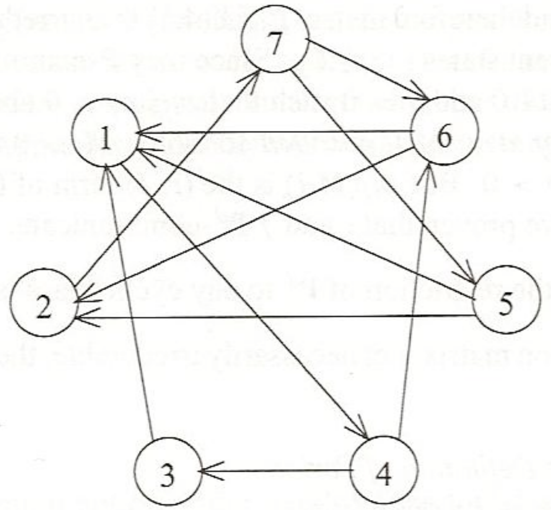
More properties of Markov chains
What we need for MCMC to work

- We want a Markov chain such that $p(\mathbf{z})$ is an invariant.
- We also require that $p(\mathbf{z}^{(\tau)})$ converge to $p(\mathbf{z})$ as $\tau \rightarrow \infty$, irrespectively of the choice of initial distribution $q(\mathbf{z}^{(0)})$.
- Some initial burn-in time is to be expected before convergence is achieved.
- Such a Markov chain is called an *ergodic* Markov chain.

- A sufficient condition for ergodicity is that the conditional distributions $p(\mathbf{z}^{(\tau+1)} | \mathbf{z}^{(\tau)})$ are nowhere zero, $p(\mathbf{z}^{(\tau+1)} | \mathbf{z}^{(\tau)}) > 0$ for all $\mathbf{z}^{(\tau+1)},\mathbf{z}^{(\tau)}$.
- This ensures that any state can be reached from any other state in a finite number of steps.

Ergodicity: $p(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)}) > 0$, a sufficient condition, but not a necessary condition

- Here is an example of a state transition diagram which leads to an ergodic Markov chain:



- With cyclic state classes:

$$C_0 = \{1,2\}\ ,\ C_1 = \{4,7\}\ ,\ C_2 = \{3,5,6\}$$

- And transition probability matrix:

$$\mathbf{P} = \begin{array}{c} \\ C_0 \\ C_1 \\ C_2 \end{array} \begin{array}{c} \begin{array}{ccc} C_0 & C_1 & C_2 \end{array} \\ \begin{bmatrix} 0 & A_0 & 0 \\ 0 & 0 & A_1 \\ A_2 & 0 & 0 \end{bmatrix} \end{array}$$

Some more properties of Markov chains
What we need for MCMC to work

- We want a Markov chain such that $p(\mathbf{z})$ is an invariant.
- We also require that $p(\mathbf{z}^{(\tau)})$ converge to $p(\mathbf{z})$ as $\tau \to \infty$, irrespectively of the choice of initial distribution $q(\mathbf{z}^{(0)})$.
- Some initial burn-in time is to be expected before convergence is achieved.

- Such a Markov chain is called an *ergodic* Markov chain.

- A sufficient condition for ergodicity is that the conditional distributions $p(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)})$ are nowhere zero, $p(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)}) > 0$ for all $\mathbf{z}^{(\tau+1)}, \mathbf{z}^{(\tau)}$.
- This ensures that any state can be reached from any other state in a finite number of steps.

The Metropolis-Hastings sampler
A MCMC method

1. Pick an initial sample state $\mathbf{z}^{(0)}$ for the Markov chain
2. Generate a sample $\mathbf{z}^*$ from the proposal distribution $q(\mathbf{z} \mid \mathbf{z}^{(\tau)})$
3. Evaluate the acceptance probability

$$A(\mathbf{z}^*, \mathbf{z}^{(\tau)}) = \min\left(1, \frac{\tilde{p}(\mathbf{z}^*)q(\mathbf{z}^{(\tau)} \mid \mathbf{z}^*)}{\tilde{p}(\mathbf{z}^{(\tau)})q(\mathbf{z}^* \mid \mathbf{z}^{(\tau)})}\right)$$

4. Pick a uniform random number, $u \sim \mathcal{U}(u \mid 0,1)$
5. Accept the sample $\mathbf{z}^*$ if $A(\mathbf{z}^*, \mathbf{z}^{(\tau)}) > u$ and set $\mathbf{z}^{(\tau+1)} = \mathbf{z}^*$,
   otherwise reuse current sample and set $\mathbf{z}^{(\tau+1)} = \mathbf{z}^{(\tau)}$
6. Repeat 2.) for $q(\mathbf{z} \mid \mathbf{z}^{(\tau+1)})$

Convergence of The Metropolis-Hastings (M-H) sampler

- We just need to check that $p(\mathbf{z})$ is invariant with respect to the Markov chain formed by the M-H sampler, by verifying that the transition probability

$$T(\mathbf{z}^{(\tau)},\mathbf{z}^{(\tau+1)}) \equiv q(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)})A(\mathbf{z}^{(\tau+1)},\mathbf{z}^{(\tau)})$$

- satisfy the detailed balance equation

$$p(\mathbf{z})T(\mathbf{z},\mathbf{z}') = p(\mathbf{z})q(\mathbf{z}' \mid \mathbf{z})A(\mathbf{z}',\mathbf{z}) = \min\big(p(\mathbf{z})q(\mathbf{z}' \mid \mathbf{z}), p(\mathbf{z}')q(\mathbf{z} \mid \mathbf{z}')\big)$$

$$= \min\big(p(\mathbf{z}')q(\mathbf{z} \mid \mathbf{z}'), p(\mathbf{z})q(\mathbf{z}' \mid \mathbf{z})\big) = p(\mathbf{z}')q(\mathbf{z} \mid \mathbf{z}')A(\mathbf{z},\mathbf{z}') = p(\mathbf{z}')T(\mathbf{z}',\mathbf{z})$$

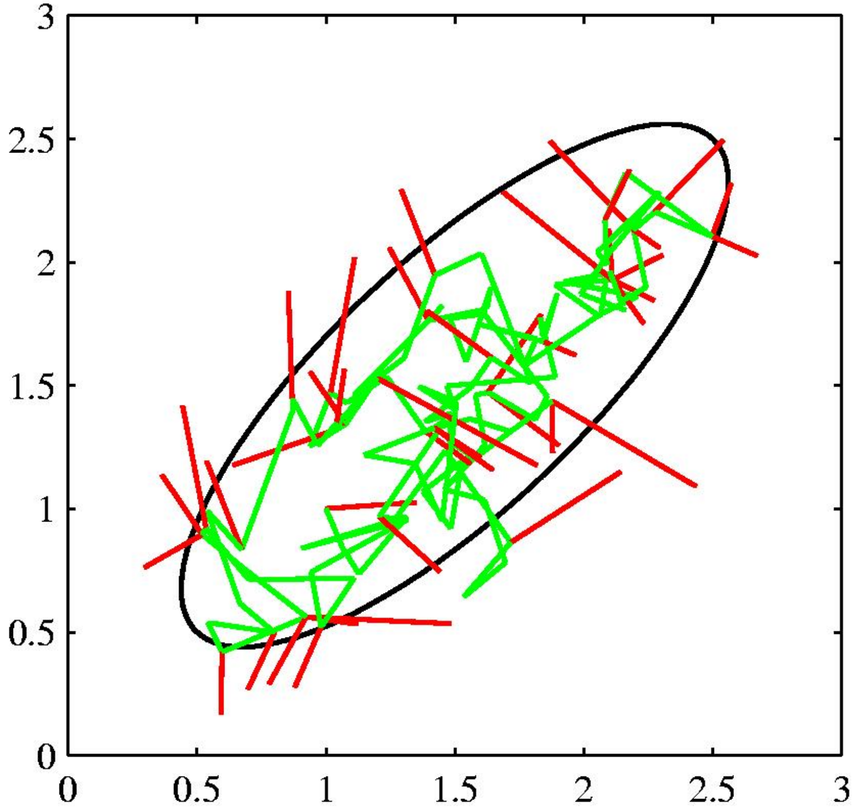
$$A(\mathbf{z}',\mathbf{z}) = \min\left(1, \frac{p(\mathbf{z}')q(\mathbf{z} \mid \mathbf{z}')}{p(\mathbf{z})q(\mathbf{z}' \mid \mathbf{z})}\right) \qquad A(\mathbf{z},\mathbf{z}') = \min\left(1, \frac{p(\mathbf{z})q(\mathbf{z}' \mid \mathbf{z})}{p(\mathbf{z}')q(\mathbf{z} \mid \mathbf{z}')}\right)$$

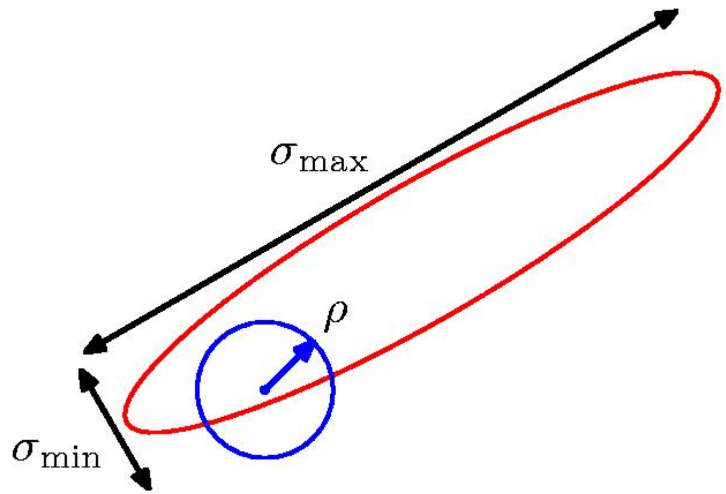
- However convergence rate and correlation between samples will dependent on the choice of proposal distribution.

The Metropolis-Hastings applied to a Gaussian

$p(\mathbf{z}) = \mathcal{N}\left(\mathbf{z} \mid \mu, \Sigma\right)$

$$q(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)}) = \mathcal{N}\left(\mathbf{z}^{(\tau+1)} \mid \mathbf{z}^{(\tau)}, \rho^2 = 0.2^2\mathbf{I}\right)$$





A good choice: $\rho = \sigma_{\min}$

Samples are independent approximately in order of steps: $\left(\dfrac{\sigma_{\max}}{\sigma_{\min}}\right)^2$

Correlation of the M-H sampler (Sampling a 2D Gaussian)

- 1-D autocorrelation with lag *k*=1, 2, …

$$\hat{R}(k) = \frac{1}{(L-k)\hat{\sigma}^2} \sum_{l=1}^{L-k} (y^{(l)} - \hat{\mu})(y^{(l+k)} - \hat{\mu})$$

- Sample mean $\hat{\mu}$ and sample variance $\hat{\sigma}^2$
- Perfect correlation/anti-correlation $\hat{R} = \pm 1$
- De-correlated $\hat{R} = 0$

- Start Matlab and run demo!

Metropolis is a special case of the M-H sampler

- Assuming the proposal distribution is symmetric

$$q(\mathbf{z}_A \mid \mathbf{z}_B) = q(\mathbf{z}_B \mid \mathbf{z}_A)$$

we get the Metropolis sampler

$$A(\mathbf{z}^*, \mathbf{z}^{(\tau)}) = \min\left(1, \frac{\tilde{p}(\mathbf{z}^*)}{\tilde{p}(\mathbf{z}^{(\tau)})}\right)$$

Gibbs Sampler for $p(\mathbf{z}) = p(z_1,\ldots,z_M)$
(Sample one variable at each step)

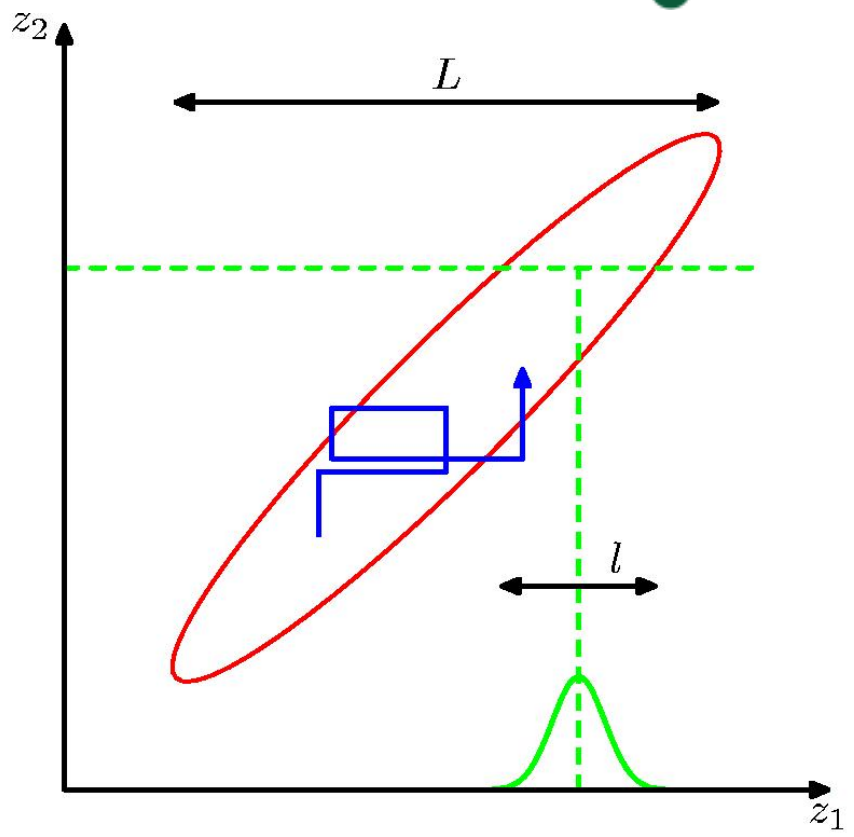
- Sample one variable at each step conditioned on the other variables

$$p(z_i \mid \mathbf{z}_{\setminus i}) = p(z_i \mid z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_M)$$

1. Initialize $\{z_i : i = 1,\ldots,M\}$
2. For τ in $1,\ldots,T$:
   1. Visit all variables i in $1,\ldots,M$:
   – Sample $z_i^{(\tau+1)}$ from

   $$p(z_i \mid z_1^{(\tau+1)},\ldots,z_{i-1}^{(\tau+1)},z_{i+1}^{(\tau)},\ldots,z_M^{(\tau)})$$

3. Repeat 2.



Gibbs sampler is a special case of the Metropolis-Hastings (M-H) sampler

- Each Gibbs step changes one variable z_k leaving the rest unchanged, such that $\mathbf{z}^*_{\backslash k} = \mathbf{z}_{\backslash k}$ in M-H steps.
- Transition probability of M-H step $q_k(\mathbf{z}^* | \mathbf{z}) = p(z^*_k | \mathbf{z}_{\backslash k})$
- The acceptance probability is

$$A(\mathbf{z}^*, \mathbf{z}) = \frac{p(\mathbf{z}^*) q_k(\mathbf{z} | \mathbf{z}^*)}{p(\mathbf{z}) q_k(\mathbf{z}^* | \mathbf{z})} = \frac{p(z^*_k | \mathbf{z}^*_{\backslash k}) p(\mathbf{z}^*_{\backslash k}) p(z_k | \mathbf{z}^*_{\backslash k})}{p(z_k | \mathbf{z}_{\backslash k}) p(\mathbf{z}_{\backslash k}) p(z^*_k | \mathbf{z}_{\backslash k})} = 1$$

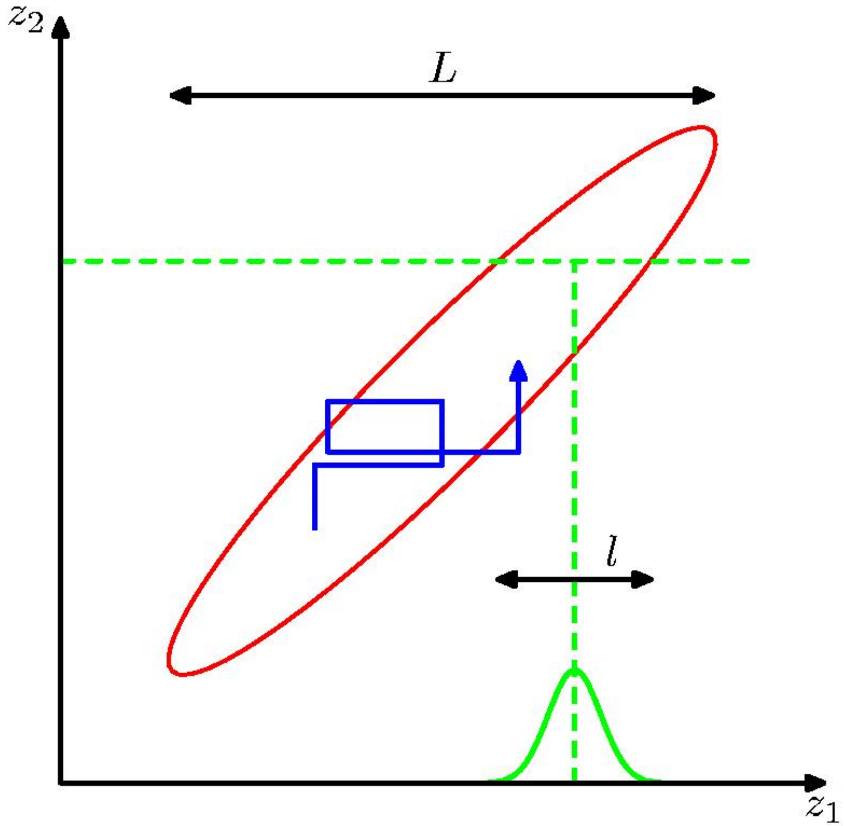
Using $p(\mathbf{z}^*) = p(z^*_k | \mathbf{z}^*_{\backslash k}) p(\mathbf{z}^*_{\backslash k})$, $p(\mathbf{z}) = p(z_k | \mathbf{z}_{\backslash k}) p(\mathbf{z}_{\backslash k})$ and $\mathbf{z}^*_{\backslash k} = \mathbf{z}_{\backslash k}$

- Hence the M-H steps are always accepted, so the Gibbs step always results in new samples (contrary to M-H).
- Since M-H converges, the Gibbs sampler also converges.

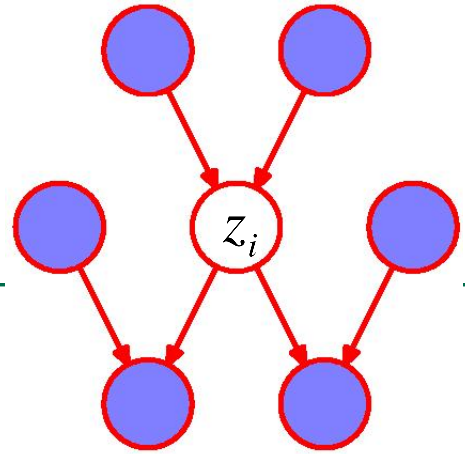
Gibbs sampler applied to a Gaussian distribution

As for M-H sampler we have:

- Size of conditional distribution and step size is *l*
- Size of marginal distribution *L*
- We need approximately in the order of $(L/l)^2$ number of steps to obtain independent samples



Gibbs sampling a Bayesian network



- We can use Markov blankets to specify

$$p(z_i \mid \mathbf{z}_{\setminus i}) = p(z_i \mid z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_M)$$

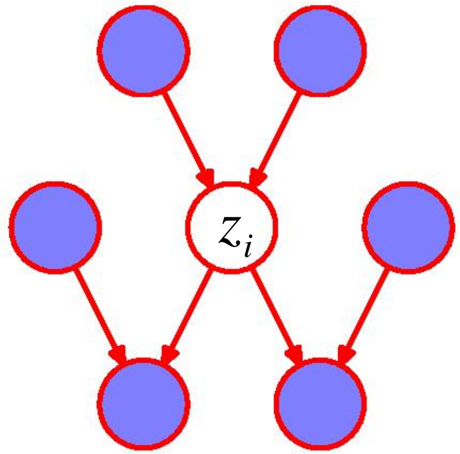
- Recall $p(z_i \mid \mathbf{z}_{\setminus i}) = \dfrac{p(z_1, \ldots, z_M)}{\int p(z_1, \ldots, z_M)\, dz_i} = \dfrac{\prod_k p(\mathbf{z}_k \mid \mathrm{pa}_k)}{\int \prod_k p(\mathbf{z}_k \mid \mathrm{pa}_k)\, dz_i}$

- all terms except those involving z_i cancels

$$p(z_i \mid \mathbf{z}_{\setminus i}) = \frac{p(\mathbf{z}_i \mid \mathrm{pa}_i) \prod_{k \in \mathrm{children}(\mathbf{z}_i)} p(\mathbf{z}_k \mid \mathrm{pa}_k) \cancel{\prod_{k \notin \mathrm{children}(\mathbf{z}_i) \cup i} p(\mathbf{z}_k \mid \mathrm{pa}_k)}}{\cancel{\prod_{k \notin \mathrm{children}(\mathbf{z}_i) \cup i} p(\mathbf{z}_k \mid \mathrm{pa}_k)} \int p(\mathbf{z}_i \mid \mathrm{pa}_i) \prod_{k \in \mathrm{children}(\mathbf{z}_i)} p(\mathbf{z}_k \mid \mathrm{pa}_k)\, dz_i}$$

Gibbs sampling a Bayesian network

- Hence we are left with the Markov blanket

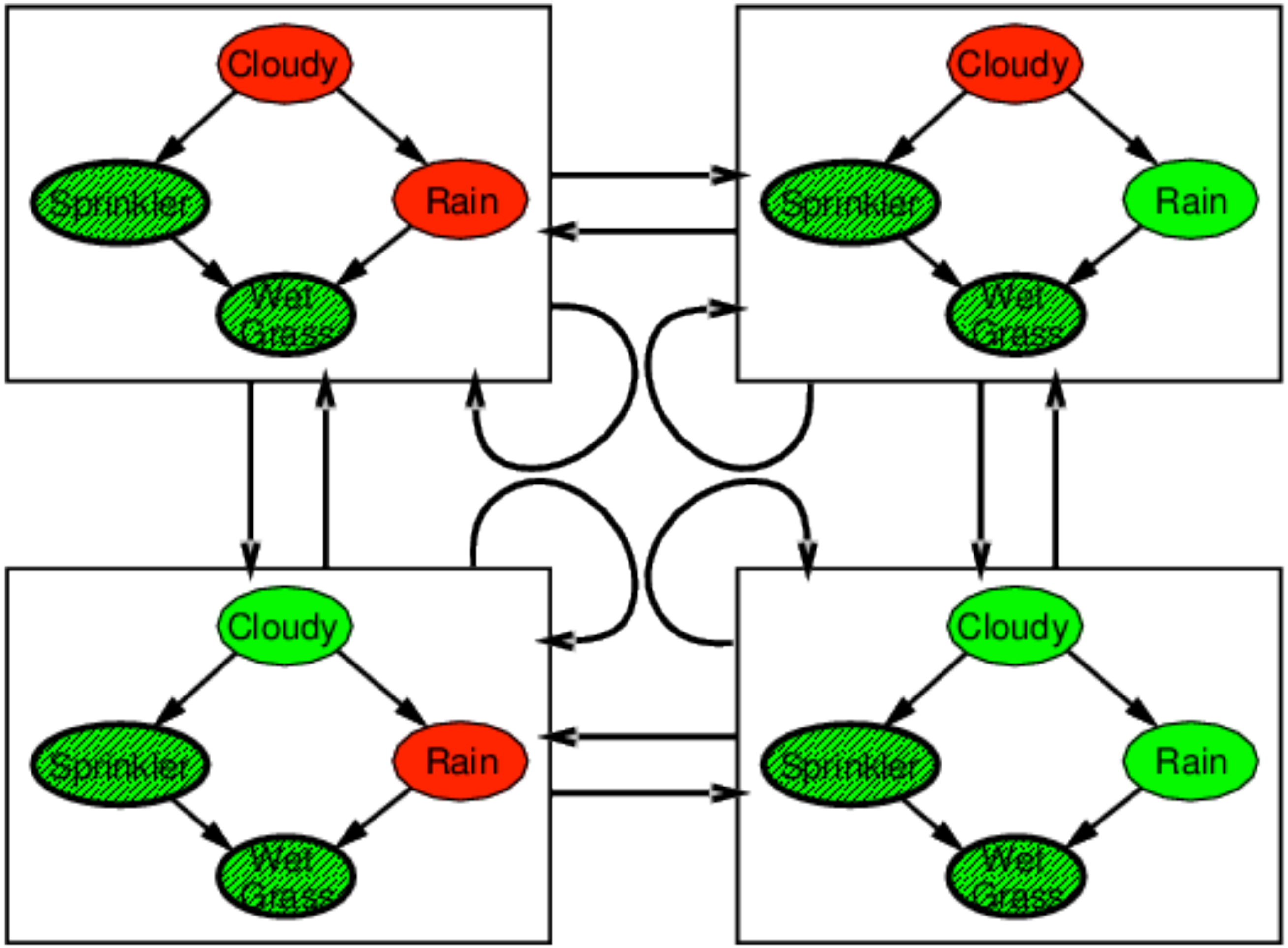


$$p(z_i \mid \mathbf{z}_{\setminus i}) = \frac{p(\mathbf{z}_i \mid \mathrm{pa}_i) \prod_{k \in \mathrm{children}(\mathbf{z}_i)} p(\mathbf{z}_k \mid \mathrm{pa}_k)}{\int p(\mathbf{z}_i \mid \mathrm{pa}_i) \prod_{k \in \mathrm{children}(\mathbf{z}_i)} p(\mathbf{z}_k \mid \mathrm{pa}_k) dz_i}$$

- We are left with the terms:
  - The conditional on z_i given its parents
  - The conditionals for the children of z_i
  - (Markov blanket: Parents, children, children's parents)
- Now we can apply the Gibbs sampling algorithm to each variable in the BN using its Markov blanket

Returning to the wet grass example

- Given that we observe Sprinkler = True and Wet Grass=True there are four states left:



Summary

- Markov Chain Monte Carlo (MCMC) methods
  - Metropolis algorithm
  - Metropolis-Hastings algorithm
  - Gibbs sampler
- More on sampling for Bayesian networks

Literature

- Basic sampling: CB Sec. 11. – 11.1.5
- Ancestral sampling: CB Sec. 8.1.2
- Likelihood weighted sampling: CB Sec. 11.1.4
- MCMC methods: CB Sec. 11.2 – 11.3

- Suggestions for further reading on MCMC:
  - Pierre Brémaud. Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues. Springer, 1999.
  - Gerhard Winkler. Image Analysis, Random Fields and Markov Chain Monte Carlo Methods – A Mathematical Introduction. Springer, 2nd edition, 2003.