PMR: Sampling II

Probabilistic Modelling and Reasoning

Amos Storkey

School of Informatics, University of Edinburgh

Outline

Outline

Outline

MCMC - Metropolis-Hastings Sampler

- Markov chain: Propose $Q(\theta'|\theta_t)$.
- Accept with probability

$$P(Accept) = \min\left(1, \frac{P(\theta')Q(\theta_t|\theta')}{P(\theta_t)Q(\theta'|\theta_t)}\right)$$

- If accept, set $\theta_{t+1} = \theta'$, else set $\theta_{t+1} = \theta_t$.

MCMC - Metropolis-Hastings Sampler

- Markov chain: Propose $Q(\theta'|\theta_t)$.
- Accept with probability

$$P(Accept) = \min\left(1, \frac{P(\theta')Q(\theta_t|\theta')}{P(\theta_t)Q(\theta'|\theta_t)}\right)$$

- If accept, set $\theta_{t+1} = \theta'$, else set $\theta_{t+1} = \theta_t$.

MCMC - Metropolis-Hastings Sampler

- Markov chain: Propose $Q(\theta'|\theta_t)$.
- Accept with probability

$$P(Accept) = \min\left(1, \frac{P(\theta')Q(\theta_t|\theta')}{P(\theta_t)Q(\theta'|\theta_t)}\right)$$

- If accept, set $\theta_{t+1} = \theta'$, else set $\theta_{t+1} = \theta_t$.

MCMC - Metropolis-Hastings Sampler

- Markov chain: Propose $Q(\theta'|\theta_t)$.
- Accept with probability

$$P(Accept) = \min\left(1, \frac{P(\theta')Q(\theta_t|\theta')}{P(\theta_t)Q(\theta'|\theta_t)}\right)$$

- If accept, set $\theta_{t+1} = \theta'$, else set $\theta_{t+1} = \theta_t$.

Metropolis-Hastings Transition

- Write out the full transition probability

$$P(\phi|\theta) = \frac{P(\phi)Q(\theta|\phi)}{P(\theta)Q(\phi|\theta)}Q(\phi|\theta) + \int d\phi' \left(1 - \min\left(1, \frac{P(\phi')Q(\theta|\phi')}{P(\theta)Q(\phi'|\theta)}\right)\right) Q(\phi'|\theta)\delta(\phi - \theta)$$

if $P(\theta)Q(\phi|\theta) > P(\phi)Q(\theta|\phi)$ and

$$P(\phi|\theta) = Q(\phi|\theta)$$

otherwise.

- Exercise: Prove satisfies detailed balance.

Metropolis-Hastings Transition

- Write out the full transition probability

$$P(\phi|\theta) = \frac{P(\phi)Q(\theta|\phi)}{P(\theta)Q(\phi|\theta)}Q(\phi|\theta)+$$
$$\int d\phi' \left(1 - \min\left(1, \frac{P(\phi')Q(\theta|\phi')}{P(\theta)Q(\phi'|\theta)}\right)\right) Q(\phi'|\theta)\delta(\phi - \theta)$$

if $P(\theta)Q(\phi|\theta) > P(\phi)Q(\theta|\phi)$ and

$$P(\phi|\theta) = Q(\phi|\theta)$$

otherwise.

- Exercise: Prove satisfies detailed balance.

Metropolis-Hastings Transition

- Write out the full transition probability

$$P(\phi|\theta) = \frac{P(\phi)Q(\theta|\phi)}{P(\theta)Q(\phi|\theta)} Q(\phi|\theta) + \int d\phi' \left(1 - \min\left(1, \frac{P(\phi')Q(\theta|\phi')}{P(\theta)Q(\phi'|\theta)}\right)\right) Q(\phi'|\theta)\delta(\phi - \theta)$$

  if $P(\theta)Q(\phi|\theta) > P(\phi)Q(\theta|\phi)$ and

$$P(\phi|\theta) = Q(\phi|\theta)$$

  otherwise.
- Exercise: Prove satisfies detailed balance.

Metropolis-Hastings Demo

- Matlab Demo

Metropolis-Hastings Demo

- Matlab Demo

Metropolis-Hastings Problems

- Setting the right proposal to provide good mixing, but reasonable acceptance probability.
- Try to get acceptance rate to be 0.234 (various arguments provide conditions for this to be optimal). Can vary width.
- Random walk behaviour in high dimensions: proposal is a diffusion process. Can take a long time to get anywhere.

Metropolis-Hastings Problems

- Setting the right proposal to provide good mixing, but reasonable acceptance probability.
- Try to get acceptance rate to be 0.234 (various arguments provide conditions for this to be optimal). Can vary width.
- Random walk behaviour in high dimensions: proposal is a diffusion process. Can take a long time to get anywhere.

Metropolis-Hastings Problems

- Setting the right proposal to provide good mixing, but reasonable acceptance probability.
- Try to get acceptance rate to be 0.234 (various arguments provide conditions for this to be optimal). Can vary width.
- Random walk behaviour in high dimensions: proposal is a diffusion process. Can take a long time to get anywhere.

Metropolis-Hastings Problems

- Setting the right proposal to provide good mixing, but reasonable acceptance probability.
- Try to get acceptance rate to be 0.234 (various arguments provide conditions for this to be optimal). Can vary width.
- Random walk behaviour in high dimensions: proposal is a diffusion process. Can take a long time to get anywhere.

MCMC - Gibbs Sampler

- Markov chain: Adapt θ_i keeping all $\theta_{j \neq i}$ fixed. i.e.
- Choose i uniformly from $i = 1, 2, \ldots, D$. Set $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$. Then sample $\theta_{t+1,i}$ from the conditional probability $P(\theta_{t+1,i}|\theta_{t+1,\neq i})$ where $\theta_{t+1,\neq i}$ denotes the set $\{\theta_{t+1,j}|j \neq i\}$.
- Repeat.
- Can cycle through i either (this is not reversible, but can be shown to have a unique equilibrium distribution)

MCMC - Gibbs Sampler

- Markov chain: Adapt θ_i keeping all $\theta_{j \neq i}$ fixed. i.e.
- Choose i uniformly from $i = 1, 2, \ldots, D$. Set $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$. Then sample $\theta_{t+1,i}$ from the conditional probability $P(\theta_{t+1,i} | \theta_{t+1,\neq i})$ where $\theta_{t+1,\neq i}$ denotes the set $\{\theta_{t+1,j} | j \neq i\}$.
- Repeat.
- Can cycle through i either (this is not reversible, but can be shown to have a unique equilibrium distribution)

Gibbs Demo

- Matlab Demo: Gaussian
- Matlab Demo: Lattice

MCMC - Block Sampler

- Gibbs sampler suffers from self reinforcement problem: frustrated systems.
- Instead of updating one variable at a time it may be possible to update a whole block of variables in one go. Can help a bit. But need to have joint distribution for block.

MCMC - Block Sampler

- Gibbs sampler suffers from self reinforcement problem: frustrated systems.
- Instead of updating one variable at a time it may be possible to update a whole block of variables in one go. Can help a bit. But need to have joint distribution for block.

MCMC - Block Sampler

- Gibbs sampler suffers from self reinforcement problem: frustrated systems.
- Instead of updating one variable at a time it may be possible to update a whole block of variables in one go. Can help a bit. But need to have joint distribution for block.

MCMC - Mixed Sampler

- Can mix ergodic sampling steps from different samplers: still satisfies detailed balance.
- Can helps to overcome the disadvantages of one method by incorporating another.
- Often helpful to add in specific steps to help with mixing: if a sampler gets stuck in one potential well (a region surrounded by low probability regions), need a means of getting it to transition to another.

MCMC - Mixed Sampler

- Can mix ergodic sampling steps from different samplers: still satisfies detailed balance.
- Can helps to overcome the disadvantages of one method by incorporating another.
- Often helpful to add in specific steps to help with mixing: if a sampler gets stuck in one potential well (a region surrounded by low probability regions), need a means of getting it to transition to another.

MCMC - Mixed Sampler

- Can mix ergodic sampling steps from different samplers: still satisfies detailed balance.
- Can helps to overcome the disadvantages of one method by incorporating another.
- Often helpful to add in specific steps to help with mixing: if a sampler gets stuck in one potential well (a region surrounded by low probability regions), need a means of getting it to transition to another.

MCMC - Mixed Sampler

- Can mix ergodic sampling steps from different samplers: still satisfies detailed balance.
- Can helps to overcome the disadvantages of one method by incorporating another.
- Often helpful to add in specific steps to help with mixing: if a sampler gets stuck in one potential well (a region surrounded by low probability regions), need a means of getting it to transition to another.

Augmentation Methods

- Suppose we have a big sampling problem, that is hard.
- Solution?
- Turn it into an even bigger problem by adding additional variables. Solve that.
- Can get sample from the original problem by just throwing unneeded variables away.
- If samples (ψ_i, θ_i) are from joint $P(\psi, \theta)$ then samples are also samples of $P(\psi|\theta)P(\theta)$ as this is the same. Hence samples θ_i must be from $P(\theta)$ as $\int d\psi P(\psi|\theta) = 1$ whatever θ is.
- Examples: Hamiltonian Monte-Carlo. Swendsen Wang.

Augmentation Methods

- Suppose we have a big sampling problem, that is hard.
- Solution?
- Turn it into an even bigger problem by adding additional variables. Solve that.
- Can get sample from the original problem by just throwing unneeded variables away.
- If samples (ψ_i, θ_i) are from joint $P(\psi, \theta)$ then samples are also samples of $P(\psi|\theta)P(\theta)$ as this is the same. Hence samples θ_i must be from $P(\theta)$ as $\int d\psi P(\psi|\theta) = 1$ whatever θ is.
- Examples: Hamiltonian Monte-Carlo. Swendsen Wang.

Augmentation Methods

- Suppose we have a big sampling problem, that is hard.
- Solution?
- Turn it into an even bigger problem by adding additional variables. Solve that.
- Can get sample from the original problem by just throwing unneeded variables away.
- If samples (ψ_i, θ_i) are from joint $P(\psi, \theta)$ then samples are also samples of $P(\psi|\theta)P(\theta)$ as this is the same. Hence samples θ_i must be from $P(\theta)$ as $\int d\psi P(\psi|\theta) = 1$ whatever θ is.
- Examples: Hamiltonian Monte-Carlo. Swendsen Wang.

Augmentation Methods

- Suppose we have a big sampling problem, that is hard.
- Solution?
- Turn it into an even bigger problem by adding additional variables. Solve that.
- Can get sample from the original problem by just throwing unneeded variables away.
- If samples (ψ_i, θ_i) are from joint $P(\psi, \theta)$ then samples are also samples of $P(\psi|\theta)P(\theta)$ as this is the same. Hence samples θ_i must be from $P(\theta)$ as $\int d\psi P(\psi|\theta) = 1$ whatever θ is.
- Examples: Hamiltonian Monte-Carlo. Swendsen Wang.

Augmentation Methods

- Suppose we have a big sampling problem, that is hard.
- Solution?
- Turn it into an even bigger problem by adding additional variables. Solve that.
- Can get sample from the original problem by just throwing unneeded variables away.
- If samples (ψ_i, θ_i) are from joint $P(\psi, \theta)$ then samples are also samples of $P(\psi|\theta)P(\theta)$ as this is the same. Hence samples θ_i must be from $P(\theta)$ as $\int d\psi P(\psi|\theta) = 1$ whatever θ is.
- Examples: Hamiltonian Monte-Carlo. Swendsen Wang.

Augmentation Methods

- Suppose we have a big sampling problem, that is hard.
- Solution?
- Turn it into an even bigger problem by adding additional variables. Solve that.
- Can get sample from the original problem by just throwing unneeded variables away.
- If samples (ψ_i, θ_i) are from joint $P(\psi, \theta)$ then samples are also samples of $P(\psi|\theta)P(\theta)$ as this is the same. Hence samples θ_i must be from $P(\theta)$ as $\int d\psi P(\psi|\theta) = 1$ whatever θ is.
- Examples: Hamiltonian Monte-Carlo. Swendsen Wang.

Augmentation Methods

- Suppose we have a big sampling problem, that is hard.
- Solution?
- Turn it into an even bigger problem by adding additional variables. Solve that.
- Can get sample from the original problem by just throwing unneeded variables away.
- If samples (ψ_i, θ_i) are from joint $P(\psi, \theta)$ then samples are also samples of $P(\psi|\theta)P(\theta)$ as this is the same. Hence samples θ_i must be from $P(\theta)$ as $\int d\psi P(\psi|\theta) = 1$ whatever θ is.
- Examples: Hamiltonian Monte-Carlo. Swendsen Wang.

MCMC - Hamiltonian (or Hybrid) Monte-Carlo

- Problem of Metropolis Hastings is random walk behaviour: slow diffusion to cover the space.
- Hamiltonion Monte-Carlo reduces this by augmenting each variable in the original space with another random variable.
- Now can do contour walks for each of these variables, in addition to Gibbs sampling steps in the joint distribution of the augmented variables.
- Related to Hamiltonian systems in physics: maintain constant energy by swapping kinetic energy for potential energy. Augmented variables are momentum variables.
- See Mackay Chapter 30.

MCMC - Hamiltonian (or Hybrid) Monte-Carlo

- Problem of Metropolis Hastings is random walk behaviour: slow diffusion to cover the space.
- Hamiltonion Monte-Carlo reduces this by augmenting each variable in the original space with another random variable.
- Now can do contour walks for each of these variables, in addition to Gibbs sampling steps in the joint distribution of the augmented variables.
- Related to Hamiltonian systems in physics: maintain constant energy by swapping kinetic energy for potential energy. Augmented variables are momentum variables.
- See Mackay Chapter 30.

MCMC - Hamiltonian (or Hybrid) Monte-Carlo

- Problem of Metropolis Hastings is random walk behaviour: slow diffusion to cover the space.
- Hamiltonion Monte-Carlo reduces this by augmenting each variable in the original space with another random variable.
- Now can do contour walks for each of these variables, in addition to Gibbs sampling steps in the joint distribution of the augmented variables.
- Related to Hamiltonian systems in physics: maintain constant energy by swapping kinetic energy for potential energy. Augmented variables are momentum variables.
- See Mackay Chapter 30.

MCMC - Hamiltonian (or Hybrid) Monte-Carlo

- Problem of Metropolis Hastings is random walk behaviour: slow diffusion to cover the space.
- Hamiltonion Monte-Carlo reduces this by augmenting each variable in the original space with another random variable.
- Now can do contour walks for each of these variables, in addition to Gibbs sampling steps in the joint distribution of the augmented variables.
- Related to Hamiltonian systems in physics: maintain constant energy by swapping kinetic energy for potential energy. Augmented variables are momentum variables.
- See Mackay Chapter 30.

MCMC - Hamiltonian (or Hybrid) Monte-Carlo

- Problem of Metropolis Hastings is random walk behaviour: slow diffusion to cover the space.
- Hamiltonion Monte-Carlo reduces this by augmenting each variable in the original space with another random variable.
- Now can do contour walks for each of these variables, in addition to Gibbs sampling steps in the joint distribution of the augmented variables.
- Related to Hamiltonian systems in physics: maintain constant energy by swapping kinetic energy for potential energy. Augmented variables are momentum variables.
- See Mackay Chapter 30.

Hamiltonian Monte-Carlo Procedure

- Original problem $P(\boldsymbol{\theta})$. Add augmented Gaussian $P(\mathbf{v})$.
- Step 1: Sample from Gaussian $P(\mathbf{v})$.
- Step 2: Choose a direction ($b = \pm 1$) (to maintain reversibility).
- Walk along Hamiltonian path.

$$\dot{\theta}_i = -b \frac{\partial}{\partial v_i} \log P(\mathbf{v})$$
$$\dot{v}_i = b \frac{\partial}{\partial \theta_i} \log P(\boldsymbol{\theta})$$

- Repeat.
- Actually have to put in a few fixes to deal with fact that can only run differential system using finite steps. Need to use leapfrog steps and run bidirectionally use a proposal/acceptance approach to ensure detailed balance.

Hamiltonian Monte-Carlo Procedure

- Original problem $P(\boldsymbol{\theta})$. Add augmented Gaussian $P(\mathbf{v})$.
- Step 1: Sample from Gaussian $P(\mathbf{v})$.
- Step 2: Choose a direction ($b = \pm 1$) (to maintain reversibility).
- Walk along Hamiltonian path.

$$\dot{\theta}_i = -b \frac{\partial}{\partial v_i} \log P(\mathbf{v})$$

$$\dot{v}_i = b \frac{\partial}{\partial \theta_i} \log P(\boldsymbol{\theta})$$

- Repeat.
- Actually have to put in a few fixes to deal with fact that can only run differential system using finite steps. Need to use leapfrog steps and run bidirectionally use a proposal/acceptance approach to ensure detailed balance.

Hamiltonian Monte-Carlo Procedure

- Original problem $P(\boldsymbol{\theta})$. Add augmented Gaussian $P(\mathbf{v})$.
- Step 1: Sample from Gaussian $P(\mathbf{v})$.
- Step 2: Choose a direction ($b = \pm 1$) (to maintain reversibility).
- Walk along Hamiltonian path.

$$\dot{\theta}_i = -b \frac{\partial}{\partial v_i} \log P(\mathbf{v})$$

$$\dot{v}_i = b \frac{\partial}{\partial \theta_i} \log P(\boldsymbol{\theta})$$

- Repeat.
- Actually have to put in a few fixes to deal with fact that can only run differential system using finite steps. Need to use leapfrog steps and run bidirectionally use a proposal/acceptance approach to ensure detailed balance.

Hamiltonian Monte-Carlo Procedure

- Original problem $P(\boldsymbol{\theta})$. Add augmented Gaussian $P(\mathbf{v})$.
- Step 1: Sample from Gaussian $P(\mathbf{v})$.
- Step 2: Choose a direction ($b = \pm 1$) (to maintain reversibility).
- Walk along Hamiltonian path.

$$\dot{\theta}_i = -b \frac{\partial}{\partial v_i} \log P(\mathbf{v})$$

$$\dot{v}_i = b \frac{\partial}{\partial \theta_i} \log P(\boldsymbol{\theta})$$

- Repeat.
- Actually have to put in a few fixes to deal with fact that can only run differential system using finite steps. Need to use leapfrog steps and run bidirectionally use a proposal/acceptance approach to ensure detailed balance.

Hamiltonian Monte-Carlo Procedure

- Original problem $P(\boldsymbol{\theta})$. Add augmented Gaussian $P(\mathbf{v})$.
- Step 1: Sample from Gaussian $P(\mathbf{v})$.
- Step 2: Choose a direction ($b = \pm 1$) (to maintain reversibility).
- Walk along Hamiltonian path.

$$\dot{\theta}_i = -b \frac{\partial}{\partial v_i} \log P(\mathbf{v})$$

$$\dot{v}_i = b \frac{\partial}{\partial \theta_i} \log P(\boldsymbol{\theta})$$

- Repeat.
- Actually have to put in a few fixes to deal with fact that can only run differential system using finite steps. Need to use leapfrog steps and run bidirectionally use a proposal/acceptance approach to ensure detailed balance.

Hamiltonian Monte-Carlo Procedure

- Original problem $P(\boldsymbol{\theta})$. Add augmented Gaussian $P(\mathbf{v})$.
- Step 1: Sample from Gaussian $P(\mathbf{v})$.
- Step 2: Choose a direction ($b = \pm 1$) (to maintain reversibility).
- Walk along Hamiltonian path.

$$\dot{\theta}_i = -b \frac{\partial}{\partial v_i} \log P(\mathbf{v})$$

$$\dot{v}_i = b \frac{\partial}{\partial \theta_i} \log P(\boldsymbol{\theta})$$

- Repeat.
- Actually have to put in a few fixes to deal with fact that can only run differential system using finite steps. Need to use leapfrog steps and run bidirectionally use a proposal/acceptance approach to ensure detailed balance.

Hamiltonian Monte-Carlo Procedure

- Original problem $P(\boldsymbol{\theta})$. Add augmented Gaussian $P(\mathbf{v})$.
- Step 1: Sample from Gaussian $P(\mathbf{v})$.
- Step 2: Choose a direction ($b = \pm 1$) (to maintain reversibility).
- Walk along Hamiltonian path.

$$\dot{\theta}_i = -b \frac{\partial}{\partial v_i} \log P(\mathbf{v})$$
$$\dot{v}_i = b \frac{\partial}{\partial \theta_i} \log P(\boldsymbol{\theta})$$

- Repeat.
- Actually have to put in a few fixes to deal with fact that can only run differential system using finite steps. Need to use leapfrog steps and run bidirectionally use a proposal/acceptance approach to ensure detailed balance.

Hamiltonian Monte-Carlo Comments

- Most commonly used in large continuous systems.
- Hamiltonian Monte-Carlo is recommended for many typical unsupervised settings.

Hamiltonian Monte-Carlo Comments

- Most commonly used in large continuous systems.
- Hamiltonian Monte-Carlo is recommended for many typical unsupervised settings.

Hamiltonian Monte-Carlo Comments

- Most commonly used in large continuous systems.
- Hamiltonian Monte-Carlo is recommended for many typical unsupervised settings.

Swendsen Wang

- Problem: Gibbs sampling mixes poorly due to self-reinforcement.
- Add in bond variables between highly aligned variables.
- Bonds can be in ‘connected’ or ‘disconnected’ states.
- Ensure marginal distribution is original problem.
- Conditioned on the states, bonds are independent. Can randomly cut strong bonds.
- Better mixing comes from fact that there are now fewer reinforcing influences.
- See `http://www.inference.phy.cam.ac.uk/mackay/itila/swendsen.pdf`

Swendsen Wang

- Problem: Gibbs sampling mixes poorly due to self-reinforcement.
- Add in bond variables between highly aligned variables.
- Bonds can be in ‘connected’ or ‘disconnected’ states.
- Ensure marginal distribution is original problem.
- Conditioned on the states, bonds are independent. Can randomly cut strong bonds.
- Better mixing comes from fact that there are now fewer reinforcing influences.
- See `http://www.inference.phy.cam.ac.uk/mackay/itila/swendsen.pdf`

Swendsen Wang

- Problem: Gibbs sampling mixes poorly due to self-reinforcement.
- Add in bond variables between highly aligned variables.
- Bonds can be in ‘connected’ or ‘disconnected’ states.
- Ensure marginal distribution is original problem.
- Conditioned on the states, bonds are independent. Can randomly cut strong bonds.
- Better mixing comes from fact that there are now fewer reinforcing influences.
- See `http://www.inference.phy.cam.ac.uk/mackay/itila/swendsen.pdf`

Swendsen Wang

- Problem: Gibbs sampling mixes poorly due to self-reinforcement.
- Add in bond variables between highly aligned variables.
- Bonds can be in ‘connected’ or ‘disconnected’ states.
- Ensure marginal distribution is original problem.
- Conditioned on the states, bonds are independent. Can randomly cut strong bonds.
- Better mixing comes from fact that there are now fewer reinforcing influences.
- See `http://www.inference.phy.cam.ac.uk/mackay/itila/swendsen.pdf`

Swendsen Wang

- Problem: Gibbs sampling mixes poorly due to self-reinforcement.
- Add in bond variables between highly aligned variables.
- Bonds can be in ‘connected’ or ‘disconnected’ states.
- Ensure marginal distribution is original problem.
- Conditioned on the states, bonds are independent. Can randomly cut strong bonds.
- Better mixing comes from fact that there are now fewer reinforcing influences.
- See `http://www.inference.phy.cam.ac.uk/mackay/itila/swendsen.pdf`

Swendsen Wang

- Problem: Gibbs sampling mixes poorly due to self-reinforcement.
- Add in bond variables between highly aligned variables.
- Bonds can be in ‘connected’ or ‘disconnected’ states.
- Ensure marginal distribution is original problem.
- Conditioned on the states, bonds are independent. Can randomly cut strong bonds.
- Better mixing comes from fact that there are now fewer reinforcing influences.
- See `http://www.inference.phy.cam.ac.uk/mackay/itila/swendsen.pdf`

Swendsen Wang

- Problem: Gibbs sampling mixes poorly due to self-reinforcement.
- Add in bond variables between highly aligned variables.
- Bonds can be in ‘connected’ or ‘disconnected’ states.
- Ensure marginal distribution is original problem.
- Conditioned on the states, bonds are independent. Can randomly cut strong bonds.
- Better mixing comes from fact that there are now fewer reinforcing influences.
- See `http://www.inference.phy.cam.ac.uk/mackay/itila/swendsen.pdf`

Swendsen Wang

- Problem: Gibbs sampling mixes poorly due to self-reinforcement.
- Add in bond variables between highly aligned variables.
- Bonds can be in ‘connected’ or ‘disconnected’ states.
- Ensure marginal distribution is original problem.
- Conditioned on the states, bonds are independent. Can randomly cut strong bonds.
- Better mixing comes from fact that there are now fewer reinforcing influences.
- See `http://www.inference.phy.cam.ac.uk/mackay/itila/swendsen.pdf`

Convergence?

- No guaranteed way to test convergence in general.
- Main tests involve tests of coalescence: converged when it has forgotten past.
- Multiple chains: do they end up in the same place?

Convergence?

- No guaranteed way to test convergence in general.
- Main tests involve tests of coalescence: converged when it has forgotten past.
- Multiple chains: do they end up in the same place?

Convergence?

- No guaranteed way to test convergence in general.
- Main tests involve tests of coalescence: converged when it has forgotten past.
- Multiple chains: do they end up in the same place?

Convergence?

- No guaranteed way to test convergence in general.
- Main tests involve tests of coalescence: converged when it has forgotten past.
- Multiple chains: do they end up in the same place?

Real Machine Learning?

- But how are these used in a real probabilistic modelling context?
- For sampling from the posterior distribution of machine learning methods.

Real Machine Learning?

- But how are these used in a real probabilistic modelling context?
- For sampling from the posterior distribution of machine learning methods.

Real Machine Learning?

- But how are these used in a real probabilistic modelling context?
- For sampling from the posterior distribution of machine learning methods.

To Do

Examinable Reading

Mackay Chapter 29, 30

Preparatory Reading

Mackay Chapter 45

Extra Reading

Any papers of Radford Neal that take your fancy. Iain Murray's tutorial slides.