CPSC 535
Metropolis-Hastings

AD

March 2007

Gibbs Sampler

- Initialization: Select deterministically or randomly $\theta = \left(\theta_1^{(0)}, ..., \theta_p^{(0)}\right)$.

Gibbs Sampler

- Initialization: Select deterministically or randomly $\theta = \left(\theta_1^{(0)}, ..., \theta_p^{(0)}\right)$.
- Iteration i; $i \geq 1$:

Gibbs Sampler

- Initialization: Select deterministically or randomly $\theta = \left(\theta_1^{(0)}, ..., \theta_p^{(0)}\right)$.
- Iteration i; $i \geq 1$:
  - For $k = 1 : p$

Gibbs Sampler

- Initialization: Select deterministically or randomly $\theta = \left(\theta_1^{(0)}, \ldots, \theta_p^{(0)}\right)$.
- Iteration i; $i \geq 1$:
  - For $k = 1 : p$
    - Sample $\theta_k^{(i)} \sim \pi\left(\theta_k \middle| \theta_{-k}^{(i)}\right)$ where $\theta_{-k}^{(i)} = \left(\theta_1^{(i)}, \ldots, \theta_{k-1}^{(i)}, \theta_{k+1}^{(i-1)}, \ldots, \theta_p^{(i-1)}\right)$.

- The Gibbs sampler requires sampling from the full conditional distributions

$$\pi\left(\theta_k \middle| \theta_{-k}\right).$$

- The Gibbs sampler requires sampling from the full conditional distributions

$$\pi\left(\theta_k \middle| \theta_{-k}\right).$$

- For many complex models, it is impossible to sample from several of these "full" conditional distributions.

- The Gibbs sampler requires sampling from the full conditional distributions

$$\pi\left(\theta_k \middle| \theta_{-k}\right).$$

- For many complex models, it is impossible to sample from several of these "full" conditional distributions.
- Even if it is possible to implement the Gibbs sampler, the algorithm might be very inefficient because the variables are very correlated or sampling from the full conditionals is extremely expensive/inefficient.

Metropolis-Hastings Algorithm

- The Metropolis-Hastings algorithm is an alternative algorithm to sample from probability distribution $\pi(\theta)$ known up to a normalizing constant.

Metropolis-Hastings Algorithm

- The Metropolis-Hastings algorithm is an alternative algorithm to sample from probability distribution $\pi(\theta)$ known up to a normalizing constant.
- This can be interpreted as the basis of all MCMC algorithm: It provides a generic way to build a Markov kernel admitting $\pi(\theta)$ as an invariant distribution.

Metropolis-Hastings Algorithm

- The Metropolis-Hastings algorithm is an alternative algorithm to sample from probability distribution $\pi(\theta)$ known up to a normalizing constant.
- This can be interpreted as the basis of all MCMC algorithm: It provides a generic way to build a Markov kernel admitting $\pi(\theta)$ as an invariant distribution.
- The Metropolis algorithm was named the "Top algorithm of the 20th century" by computer scientists, mathematicians, physicists.

- Introduce a proposal distribution/kernel $q(\theta, \theta')$, i.e.

$$\int q(\theta, \theta')\, d\theta' = 1 \text{ for any } \theta.$$

- Introduce a proposal distribution/kernel $q(\theta, \theta')$, i.e.

$$\int q(\theta, \theta')\, d\theta' = 1 \text{ for any } \theta.$$

- The basic idea of the MH algorithm is to propose a new candidate θ' based on the current state of the Markov chain θ.

- Introduce a proposal distribution/kernel $q\left(\theta,\theta'\right)$, i.e.

$$\int q\left(\theta,\theta'\right)d\theta' = 1 \text{ for any } \theta.$$

- The basic idea of the MH algorithm is to propose a new candidate θ' based on the current state of the Markov chain θ.
- We only accept this algorithm with respect to a probability $\alpha\left(\theta,\theta'\right)$ which ensures that the invariant distribution of the transition kernel is the target distribution $\pi\left(\theta\right)$.

- Initialization: Select deterministically or randomly $\theta^{(0)}$.

- Initialization: Select deterministically or randomly $\theta^{(0)}$.
- Iteration i; $i \geq 1$:

- Initialization: Select deterministically or randomly $\theta^{(0)}$.
- Iteration i; $i \geq 1$:
  - Sample $\theta^* \sim q\left(\theta^{(i-1)}, \theta^*\right)$ and compute

$$\alpha\left(\theta^{(i-1)}, \theta^*\right) = \min\left(1, \frac{\pi\left(\theta^*\right) q\left(\theta^*, \theta^{(i-1)}\right)}{\pi\left(\theta^{(i-1)}\right) q\left(\theta^{(i-1)}, \theta^*\right)}\right).$$

- Initialization: Select deterministically or randomly $\theta^{(0)}$.
- Iteration i; $i \geq 1$:
  - Sample $\theta^* \sim q\left(\theta^{(i-1)}, \theta^*\right)$ and compute

$$\alpha\left(\theta^{(i-1)}, \theta^*\right) = \min\left(1, \frac{\pi\left(\theta^*\right) q\left(\theta^*, \theta^{(i-1)}\right)}{\pi\left(\theta^{(i-1)}\right) q\left(\theta^{(i-1)}, \theta^*\right)}\right).$$

  - With probability $\alpha\left(\theta^{(i-1)}, \theta^*\right)$, set $\theta^{(i)} = \theta^*$; otherwise set $\theta^{(i)} = \theta^{(i-1)}$.

- It is not necessary to know the normalizing constant of $\pi(\theta)$ to implement the algorithm.

- It is not necessary to know the normalizing constant of $\pi(\theta)$ to implement the algorithm.
- This algorithm is extremely general: $q(\theta, \theta')$ can be any proposal distribution. So in practice, we can select it so that it is easy to sample from it.

- It is not necessary to know the normalizing constant of $\pi(\theta)$ to implement the algorithm.
- This algorithm is extremely general: $q(\theta, \theta')$ can be any proposal distribution. So in practice, we can select it so that it is easy to sample from it.
- There is much more freedom than in the Gibbs sampler where the proposal distributions are fixed.

Random Walk Metropolis

- The original Metropolis algorithm (1953) corresponds to the following choice for $q(\theta, \theta')$

$$\theta' = \theta + Z \text{ where } Z \sim f;$$

i.e. this is a so-called *random walk proposal*.

Random Walk Metropolis

- The original Metropolis algorithm (1953) corresponds to the following choice for $q(\theta, \theta')$

$$\theta' = \theta + Z \text{ where } Z \sim f;$$

i.e. this is a so-called *random walk proposal*.

- The distribution $f(z)$ is the distribution of the random walk increments Z and

$$q(\theta, \theta') = f(\theta' - \theta) \quad \Rightarrow \quad \alpha(\theta, \theta') = \min\left(1, \frac{\pi(\theta') f(\theta - \theta')}{\pi(\theta) f(\theta' - \theta)}\right).$$

Random Walk Metropolis

- The original Metropolis algorithm (1953) corresponds to the following choice for $q(\theta, \theta')$

$$\theta' = \theta + Z \text{ where } Z \sim f;$$

  i.e. this is a so-called *random walk proposal*.

- The distribution $f(z)$ is the distribution of the random walk increments Z and

$$q(\theta, \theta') = f(\theta' - \theta) \quad \Rightarrow \alpha(\theta, \theta') = \min\left(1, \frac{\pi(\theta') f(\theta - \theta')}{\pi(\theta) f(\theta' - \theta)}\right).$$

- If $f(\theta' - \theta) = f(\theta - \theta')$ - e.g. $Z \sim \mathcal{N}(0, \Sigma)$- then

$$\alpha(\theta, \theta') = \min\left(1, \frac{\pi(\theta')}{\pi(\theta)}\right)$$

Independent Metropolis-Hastings

- The Hastings' generalization (1970) corresponds to the following choice for $q(\theta, \theta')$
$$q(\theta, \theta') = q(\theta');$$
i.e. this is a so-called *independent proposal*.

Independent Metropolis-Hastings

- The Hastings' generalization (1970) corresponds to the following choice for $q(\theta, \theta')$

$$q(\theta, \theta') = q(\theta');$$

  i.e. this is a so-called *independent proposal*.

- In this case, the acceptance probability is given by

$$\alpha(\theta, \theta') = \min\left(1, \frac{\pi(\theta')\, q(\theta)}{\pi(\theta)\, q(\theta')}\right) = \min\left(1, \frac{\pi^*(\theta')}{q^*(\theta')} \frac{q^*(\theta)}{\pi^*(\theta)}\right)$$

  where π^* and q^* are unnormalized versions of π and q.

Independent Metropolis-Hastings

- The Hastings' generalization (1970) corresponds to the following choice for $q(\theta, \theta')$

$$q(\theta, \theta') = q(\theta');$$

i.e. this is a so-called *independent proposal*.

- In this case, the acceptance probability is given by

$$\alpha(\theta, \theta') = \min\left(1, \frac{\pi(\theta')\, q(\theta)}{\pi(\theta)\, q(\theta')}\right) = \min\left(1, \frac{\pi^*(\theta')}{q^*(\theta')} \frac{q^*(\theta)}{\pi^*(\theta)}\right)$$

where π^* and q^* are unnormalized versions of π and q.

- The ratio $\pi^*(\theta) / q^*(\theta)$ appearing in the Accept/Reject and Importance Sampling methods also reappears here.

- To establish that the MH chain converges towards the required target, we need to establish that

- To establish that the MH chain converges towards the required target, we need to establish that
  - $\pi(\theta)$ is the invariant distribution of the Markov kernel associated to the MH algorithm.

- To establish that the MH chain converges towards the required target, we need to establish that
  - $\pi(\theta)$ is the invariant distribution of the Markov kernel associated to the MH algorithm.
  - The Markov chain is irreducible; i.e. one can reach any set A such that $\pi(A) > 0$.

- To establish that the MH chain converges towards the required target, we need to establish that
  - $\pi(\theta)$ is the invariant distribution of the Markov kernel associated to the MH algorithm.
  - The Markov chain is irreducible; i.e. one can reach any set A such that $\pi(A) > 0$.
  - The Markov chain is aperiodic; i.e. one does not visit in a periodic way the state-space.

Invariance of the MH kernel

- The transition kernel associated to the MH algorithm can be rewritten as

$$K(\theta, \theta') = \alpha(\theta, \theta')\, q(\theta, \theta') + \underbrace{\left(1 - \int \alpha(\theta, u)\, q(\theta, u)\, du\right)}_{\text{rejection probability}} \delta_\theta(\theta')$$

Invariance of the MH kernel

- The transition kernel associated to the MH algorithm can be rewritten as

$$K(\theta,\theta') = \alpha(\theta,\theta')\, q(\theta,\theta') + \underbrace{\left(1 - \int \alpha(\theta,u)\, q(\theta,u)\, du\right)}_{\text{rejection probability}} \delta_\theta(\theta')$$

- **Remark**: This is a lose notation for

$$K(\theta,d\theta') = \alpha(\theta,\theta')\, q(\theta,\theta')\, d\theta' + \left(1 - \int \alpha(\theta,u)\, q(\theta,u)\, du\right) \delta_\theta(d\theta').$$

Invariance of the MH kernel

- The transition kernel associated to the MH algorithm can be rewritten as

$$K(\theta, \theta') = \alpha(\theta, \theta') \, q(\theta, \theta') + \underbrace{\left(1 - \int \alpha(\theta, u) \, q(\theta, u) \, du\right)}_{\text{rejection probability}} \delta_\theta(\theta')$$

- **Remark**: This is a lose notation for

$$K(\theta, d\theta') = \alpha(\theta, \theta') \, q(\theta, \theta') \, d\theta' + \left(1 - \int \alpha(\theta, u) \, q(\theta, u) \, du\right) \delta_\theta(d\theta') \,.$$

- Clearly we have

$$\begin{aligned} \int K(\theta, \theta') \, d\theta' &= \int \alpha(\theta, \theta') \, q(\theta, \theta') \, d\theta' \\ &\quad + \left(1 - \int \alpha(\theta, u) \, q(\theta, u) \, du\right) \int \delta_\theta(\theta') \, d\theta' \\ &= 1. \end{aligned}$$

- We want to show that

$$\int \pi(\theta) K(\theta, \theta') \, d\theta = \pi(\theta').$$

- We want to show that

$$\int \pi(\theta) K(\theta, \theta') d\theta = \pi(\theta') .$$

- Note that this condition is satisfied if the *reversibility property* is satisfied: For all θ, θ'

$$\pi(\theta) K(\theta, \theta') = \pi(\theta') K(\theta', \theta) ;$$

i.e. the probability of being in A and moving to B is equal to the probability of being in B and moving to A.

- We want to show that

$$\int \pi(\theta) K(\theta, \theta') \, d\theta = \pi(\theta').$$

- Note that this condition is satisfied if the *reversibility property* is satisfied: For all θ, θ'

$$\pi(\theta) K(\theta, \theta') = \pi(\theta') K(\theta', \theta);$$

  i.e. the probability of being in A and moving to B is equal to the probability of being in B and moving to A.

- Indeed the reversibility condition implies that

$$\begin{aligned}\int \pi(\theta) K(\theta, \theta') \, d\theta &= \int \pi(\theta') K(\theta', \theta) \, d\theta \\ &= \pi(\theta') \int K(\theta', \theta) \, d\theta \\ &= \pi(\theta')\end{aligned}$$

- Be careful: If a kernel is $\pi-$reversible then it is $\pi-$invariant but the reverse is not true.

- Be careful: If a kernel is $\pi-$reversible then it is $\pi-$invariant but the reverse is not true.
- The deterministic scan Gibbs sampler is not $\pi-$reversible as

$$\begin{aligned} & \pi(\theta_1, \theta_2)\, \pi\left(\theta_2' \middle| \theta_1\right) \pi\left(\theta_1' \middle| \theta_2'\right) \\ \neq \; & \pi\left(\theta_1', \theta_2'\right) \pi\left(\theta_2 \middle| \theta_1'\right) \pi\left(\theta_2' \middle| \theta_1'\right). \end{aligned}$$

- By definition of the kernel, we have

$$\begin{aligned}\pi(\theta) K(\theta, \theta') &= \pi(\theta) \alpha(\theta, \theta') q(\theta, \theta') \\ &+ \left(1 - \int \alpha(\theta, u) q(\theta, u) du\right) \delta_{\theta}(\theta') \pi(\theta).\end{aligned}$$

- By definition of the kernel, we have

$$\begin{aligned}\pi(\theta) K(\theta,\theta') &= \pi(\theta)\alpha(\theta,\theta') q(\theta,\theta') \\ &\quad + \left(1 - \int \alpha(\theta,u) q(\theta,u) du\right) \delta_{\theta}(\theta') \pi(\theta).\end{aligned}$$

- Then

$$\begin{aligned}\pi(\theta)\alpha(\theta,\theta') q(\theta,\theta') &= \pi(\theta)\min\left(1, \frac{\pi(\theta') q(\theta',\theta)}{\pi(\theta) q(\theta,\theta')}\right) q(\theta,\theta') \\ &= \min\left(\pi(\theta) q(\theta,\theta'), \pi(\theta') q(\theta',\theta)\right) \\ &= \pi(\theta')\min\left(1, \frac{\pi(\theta) q(\theta,\theta')}{\pi(\theta') q(\theta',\theta)}\right) q(\theta',\theta) \\ &= \pi(\theta')\alpha(\theta',\theta) q(\theta',\theta).\end{aligned}$$

- We have obviously

$$
\begin{aligned}
& \left(1-\int \alpha\left(\theta, u\right) q\left(\theta, u\right) du\right) \delta_{\theta}\left(\theta'\right) \pi\left(\theta\right) \\
= & \left(1-\int \alpha\left(\theta', u\right) q\left(\theta', u\right) du\right) \delta_{\theta'}\left(\theta\right) \pi\left(\theta'\right).
\end{aligned}
$$

- We have obviously

$$\begin{aligned}
& \left(1-\int \alpha\left(\theta,u\right)q\left(\theta,u\right)du\right)\delta_{\theta}\left(\theta'\right)\pi\left(\theta\right) \\
= \quad & \left(1-\int \alpha\left(\theta',u\right)q\left(\theta',u\right)du\right)\delta_{\theta'}\left(\theta\right)\pi\left(\theta'\right).
\end{aligned}$$

- It follows that

$$\pi\left(\theta\right)K\left(\theta,\theta'\right)=\pi\left(\theta'\right)K\left(\theta',\theta\right).$$

- We have obviously

$$\begin{aligned} & \left(1-\int \alpha\left(\theta, u\right) q\left(\theta, u\right) du\right) \delta_{\theta}\left(\theta'\right) \pi\left(\theta\right) \\ = & \left(1-\int \alpha\left(\theta', u\right) q\left(\theta', u\right) du\right) \delta_{\theta'}\left(\theta\right) \pi\left(\theta'\right). \end{aligned}$$

- It follows that

$$\pi\left(\theta\right) K\left(\theta, \theta'\right) = \pi\left(\theta'\right) K\left(\theta', \theta\right).$$

- Hence, π is the invariant distribution of the transition kernel K.

Irreducibility and Aperiodicity

- To ensure irreducibility, a sufficient but not necessary condition is that

$$\pi\left(\theta'\right) > 0 \Rightarrow q\left(\theta,\theta'\right) > 0.$$

Irreducibility and Aperiodicity

- To ensure irreducibility, a sufficient but not necessary condition is that

$$\pi\left(\theta'\right) > 0 \Rightarrow q\left(\theta, \theta'\right) > 0.$$

- Aperiodicity is automatically ensured as there is always a strictly positive probability to reject the candidate.

Irreducibility and Aperiodicity

- To ensure irreducibility, a sufficient but not necessary condition is that

$$\pi\left(\theta'\right) > 0 \Rightarrow q\left(\theta, \theta'\right) > 0.$$

- Aperiodicity is automatically ensured as there is always a strictly positive probability to reject the candidate.
- Theoretically, the MH algorithm converges under very weak assumptions to the target distribution π. In practice, this convergence can be so slow that the algorithm is useless.

- If you are using independent proposals then you would like to have $q(\theta) \approx \pi(\theta)$.

- If you are using independent proposals then you would like to have $q(\theta) \approx \pi(\theta)$.
- In practice, similarly to Rejection sampling or Importance Sampling, you need to ensure that

$$\frac{\pi(\theta)}{q(\theta)} \leq C$$

to obtain good performance.

- If you are using independent proposals then you would like to have $q(\theta) \approx \pi(\theta)$.
- In practice, similarly to Rejection sampling or Importance Sampling, you need to ensure that

$$\frac{\pi(\theta)}{q(\theta)} \leq C$$

  to obtain good performance.
- If you don't ensure this condition, the algorithm might give you the impression it works well... but it does NOT.

Examples

- **Example**: Consider the case where

$$\pi(\theta) \propto \exp\left(-\frac{\theta^2}{2}\right).$$

Examples

- **Example**: Consider the case where

$$\pi(\theta) \propto \exp\left(-\frac{\theta^2}{2}\right).$$

- We implement the MH algorithm for

$$q_1(\theta) \propto \exp\left(-\frac{\theta^2}{2(0.2)^2}\right)$$

so $\pi(\theta)/q_1(\theta) \rightarrow \infty$ as $\theta \rightarrow \infty$ and for

$$q_2(\theta) \propto \exp\left(-\frac{\theta^2}{2(5)^2}\right)$$

so $\pi(\theta)/q_2(\theta) \leq C < \infty$ for all θ.

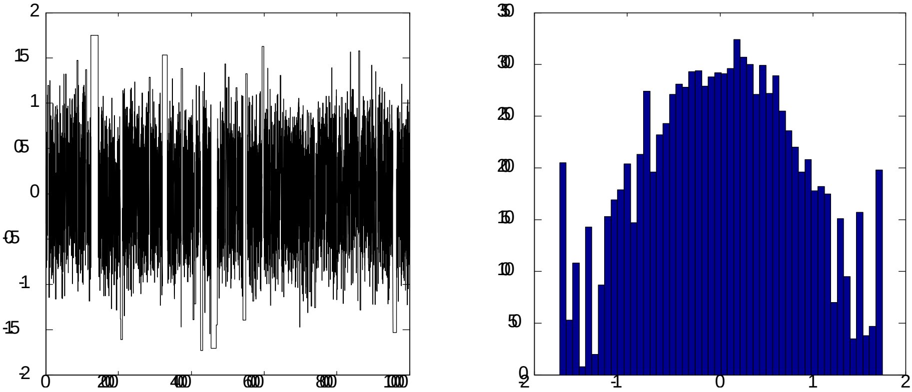

Figure: MCMC output for q_1, we estimate $\mathbb{E}(\theta) = 0.0206$ and $\mathbb{V}(\theta) = 0.83$.

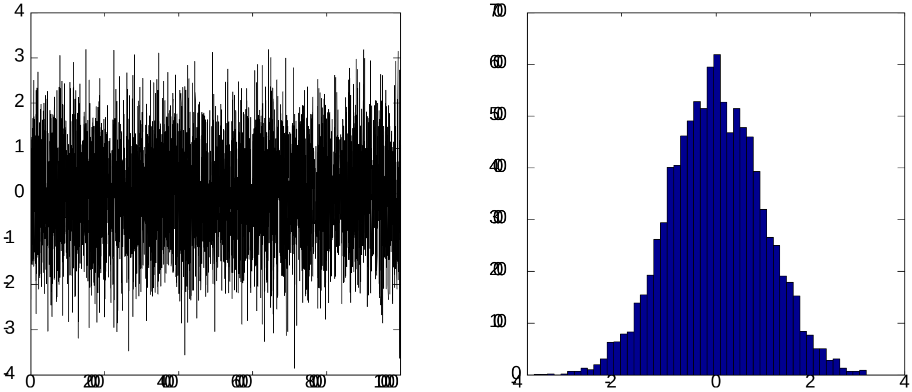

Figure: MCMC output for q_2, we estimate $\mathbb{E}(\theta) = -0.004$ and $\mathbb{V}(\theta) = 1.00$.

- Consider now a random walk move. In this case, there is no clear guideline how to select the proposal distribution.

- Consider now a random walk move. In this case, there is no clear guideline how to select the proposal distribution.
- When the variance of the random walk increments (if it exists) is very small then the acceptance rate can be expected to be around 0.5-0.7.

- Consider now a random walk move. In this case, there is no clear guideline how to select the proposal distribution.
- When the variance of the random walk increments (if it exists) is very small then the acceptance rate can be expected to be around 0.5-0.7.
- You would like to scale the random walk moves such that it is possible to move reasonably fast in regions of positive probability masses under π.

- **Example**: Consider the case where

$$\pi\left(\theta\right) \propto \exp\left(-\frac{\theta^2}{2}\right).$$

- **Example**: Consider the case where

$$\pi\left(\theta\right) \propto \exp\left(-\frac{\theta^2}{2}\right).$$

- We implement the MH algorithm for

$$\begin{aligned}
q_1\left(\theta, \theta'\right) &\propto \exp\left(-\frac{\left(\theta' - \theta\right)^2}{2\left(0.2\right)^2}\right), \\
q_2\left(\theta, \theta'\right) &\propto \exp\left(-\frac{\left(\theta' - \theta\right)^2}{2\left(5\right)^2}\right), \\
q_3\left(\theta, \theta'\right) &\propto \exp\left(-\frac{\left(\theta' - \theta\right)^2}{2\left(0.02\right)^2}\right).
\end{aligned}$$

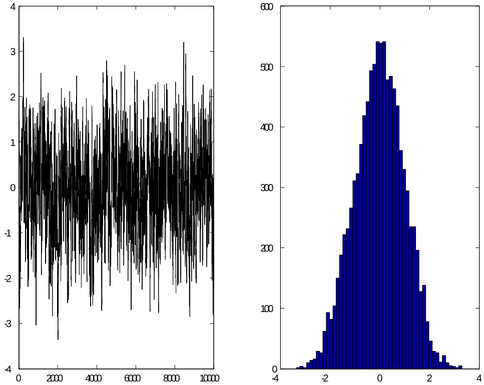

Figure: MCMC output for q_1, we estimate $\mathbb{E}(\theta) = -0.02$ and $\mathbb{V}(\theta) = 0.99$

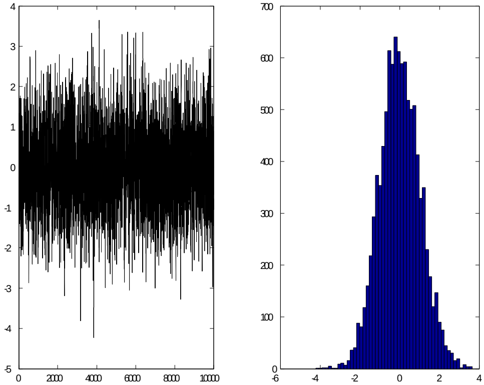

Figure: MCMC output for q_2, we estimate $\mathbb{E}(\theta) = 0.00$ and $\mathbb{V}(\theta) = 1.02$.

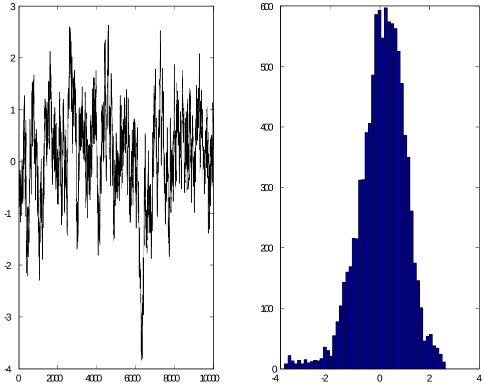

Figure: MCMC output for q_3, we estimate $\mathbb{E}(\theta) = 0.10$ and $\mathbb{V}(\theta) = 0.92$.

- Heavy tails increments can prevent you from getting trappedb in modes.

- Heavy tails increments can prevent you from getting trappedb in modes.
- It is tempting to adapt the variance of the increments given the simulation output... Unfortunately this breaks the Markov property and biases results if one is not careful.

Mixture of Proposals

- In practice, random walk proposals can be used to explore locally the space whereas independent walk proposals can be used to jump into the space.

Mixture of Proposals

- In practice, random walk proposals can be used to explore locally the space whereas independent walk proposals can be used to jump into the space.
- So a good strategy can be to use a proposal distribution of the form

$$q\left(\theta, \theta'\right) = \lambda q_1\left(\theta'\right) + (1-\lambda)\, q_2\left(\theta, \theta'\right)$$

where $0 < \lambda < 1$.

Mixture of Proposals

- In practice, random walk proposals can be used to explore locally the space whereas independent walk proposals can be used to jump into the space.
- So a good strategy can be to use a proposal distribution of the form

$$q\left(\theta,\theta'\right)=\lambda q_1\left(\theta'\right)+\left(1-\lambda\right)q_2\left(\theta,\theta'\right)$$

  where $0<\lambda<1$.
- This algorithm is definitely valid as it is just a particular case of the MH algorithm.

Mixture of Kernels

- An alternative achieving the same purpose is to use a transition kernel

$$K\left(\theta, \theta'\right) = \lambda K_1\left(\theta, \theta'\right) + (1 - \lambda) K_2\left(\theta, \theta'\right)$$

where K_1 (resp. K_2) is an MH algorithm of proposal q_1 (resp. q_2).

Mixture of Kernels

- An alternative achieving the same purpose is to use a transition kernel

$$K(\theta,\theta') = \lambda K_1(\theta,\theta') + (1-\lambda) K_2(\theta,\theta')$$

where K_1 (resp. K_2) is an MH algorithm of proposal q_1 (resp. q_2).
- This algorithm is different from using $q(\theta,\theta') = \lambda q_1(\theta') + (1-\lambda) q_2(\theta,\theta')$. It is computationally cheaper and still valid as

$$\begin{aligned}
& \int \pi(\theta) K(\theta,\theta') \, d\theta \\
= \; & \lambda \int \pi(\theta) K_1(\theta,\theta') \, d\theta + (1-\lambda) \int \pi(\theta) K_2(\theta,\theta') \, d\theta \\
= \; & \lambda \pi(\theta') + (1-\lambda) \pi(\theta') \\
= \; & \pi(\theta')
\end{aligned}$$

Langevin-type algorithm

- We usually wants to sample candidates in regions of high probability masses.

Langevin-type algorithm

- We usually wants to sample candidates in regions of high probability masses.
- We can use

$$\theta' = \theta + \frac{\sigma^2}{2} \nabla \log \pi(\theta) + \sigma V \text{ where } V \sim \mathcal{N}(0,1)$$

where σ^2 is selected such that the acceptance ratio is approximately 0.57.

Langevin-type algorithm

- We usually wants to sample candidates in regions of high probability masses.
- We can use

$$\theta' = \theta + \frac{\sigma^2}{2}\nabla \log \pi(\theta) + \sigma V \text{ where } V \sim \mathcal{N}(0,1)$$

where σ^2 is selected such that the acceptance ratio is approximately 0.57.

- The motivation is that, we know that in continuous-time

$$d\theta_t = \frac{1}{2}\nabla \log \pi(\theta) + \sigma dW_t$$

admits π has an invariant distribution.

More general strategies

- To build $q(\theta, \theta')$, you can use complex deterministic strategies.

More general strategies

- To build $q\left(\theta,\theta'\right)$, you can use complex deterministic strategies.
- Assume you are in θ and you want to propose

$$\theta' \sim \mathcal{N}\left(\varphi\left(\theta\right),\sigma^2\right).$$

More general strategies

- To build $q(\theta, \theta')$, you can use complex deterministic strategies.
- Assume you are in θ and you want to propose

$$\theta' \sim \mathcal{N}\left(\varphi(\theta), \sigma^2\right).$$

- You do not need to have an explicit form for the mapping φ! As long as φ is a *deterministic* mapping, then it is fine. For example $\varphi(\theta)$ could be the local maximum of π closest to θ that has been determined using a gradient algorithm.

More general strategies

- To build $q\left(\theta, \theta'\right)$, you can use complex deterministic strategies.
- Assume you are in θ and you want to propose

$$\theta' \sim \mathcal{N}\left(\varphi\left(\theta\right), \sigma^2\right).$$

- You do not need to have an explicit form for the mapping φ! As long as φ is a *deterministic* mapping, then it is fine. For example $\varphi\left(\theta\right)$ could be the local maximum of π closest to θ that has been determined using a gradient algorithm.
- To compute the acceptance probability of the candidate θ', you will need to compute $\varphi\left(\theta'\right)$ and then you can compute the MH acceptance ratio.

Extensions

- The standard MH algorithm uses the acceptance probability

$$\alpha\left(\theta,\theta'\right)=\min\left(1,\frac{\pi\left(\theta'\right)q\left(\theta',\theta\right)}{\pi\left(\theta\right)q\left(\theta,\theta'\right)}\right)$$

Extensions

- The standard MH algorithm uses the acceptance probability

$$\alpha\left(\theta,\theta'\right)=\min\left(1,\frac{\pi\left(\theta'\right)q\left(\theta',\theta\right)}{\pi\left(\theta\right)q\left(\theta,\theta'\right)}\right)$$

- This is not necessary and one can also use any function

$$\alpha\left(\theta,\theta'\right)=\frac{\delta\left(\theta,\theta'\right)}{\pi\left(\theta\right)q\left(\theta,\theta'\right)}$$

which is such that

$$\delta\left(\theta,\theta'\right)=\delta\left(\theta',\theta\right)\text{ and }0\leq\alpha\left(\theta,\theta'\right)\leq1$$

Extensions

- The standard MH algorithm uses the acceptance probability

$$\alpha\left(\theta,\theta'\right)=\min\left(1,\frac{\pi\left(\theta'\right)q\left(\theta',\theta\right)}{\pi\left(\theta\right)q\left(\theta,\theta'\right)}\right)$$

- This is not necessary and one can also use any function

$$\alpha\left(\theta,\theta'\right)=\frac{\delta\left(\theta,\theta'\right)}{\pi\left(\theta\right)q\left(\theta,\theta'\right)}$$

which is such that

$$\delta\left(\theta,\theta'\right)=\delta\left(\theta',\theta\right)\text{ and }0\leq\alpha\left(\theta,\theta'\right)\leq1$$

- Example (Baker, 1965):

$$\alpha\left(\theta,\theta'\right)=\frac{\pi\left(\theta'\right)q\left(\theta',\theta\right)}{\pi\left(\theta'\right)q\left(\theta',\theta\right)+\pi\left(\theta\right)q\left(\theta,\theta'\right)}.$$

- Indeed one can check that

$$K\left(\theta,\theta'\right)=\alpha\left(\theta,\theta'\right)q\left(\theta,\theta'\right)+\left(1-\int\alpha\left(\theta,u\right)q\left(\theta,u\right)du\right)\delta_{\theta}\left(\theta'\right)$$

is π-reversible.

- Indeed one can check that

$$K\left(\theta,\theta'\right)=\alpha\left(\theta,\theta'\right)q\left(\theta,\theta'\right)+\left(1-\int\alpha\left(\theta,u\right)q\left(\theta,u\right)du\right)\delta_{\theta}\left(\theta'\right)$$

is π-reversible.

- We have

$$\begin{aligned}\pi\left(\theta\right)\alpha\left(\theta,\theta'\right)q\left(\theta,\theta'\right) &= \pi\left(\theta\right)\frac{\delta\left(\theta,\theta'\right)}{\pi\left(\theta\right)q\left(\theta,\theta'\right)}q\left(\theta,\theta'\right)\\ &= \delta\left(\theta,\theta'\right)\\ &= \delta\left(\theta',\theta\right)\\ &= \pi\left(\theta'\right)\alpha\left(\theta',\theta\right)q\left(\theta',\theta\right).\end{aligned}$$

- Indeed one can check that

$$K\left(\theta,\theta'\right)=\alpha\left(\theta,\theta'\right)q\left(\theta,\theta'\right)+\left(1-\int\alpha\left(\theta,u\right)q\left(\theta,u\right)du\right)\delta_{\theta}\left(\theta'\right)$$

  is π-reversible.
- We have

$$\begin{aligned}\pi\left(\theta\right)\alpha\left(\theta,\theta'\right)q\left(\theta,\theta'\right) &= \pi\left(\theta\right)\frac{\delta\left(\theta,\theta'\right)}{\pi\left(\theta\right)q\left(\theta,\theta'\right)}q\left(\theta,\theta'\right)\\ &= \delta\left(\theta,\theta'\right)\\ &= \delta\left(\theta',\theta\right)\\ &= \pi\left(\theta'\right)\alpha\left(\theta',\theta\right)q\left(\theta',\theta\right).\end{aligned}$$

- The MH acceptance is favoured as it increases the acceptance probability.

Limitations of the MH algorithm

- The MH algorithm is a simple and very general algorithm to sample from a target distribution $\pi(\theta)$.

Limitations of the MH algorithm

- The MH algorithm is a simple and very general algorithm to sample from a target distribution $\pi(\theta)$.
- In practice, the choice of the proposal distribution is absolutely crucial on the performance of the algorithm.

Limitations of the MH algorithm

- The MH algorithm is a simple and very general algorithm to sample from a target distribution $\pi(\theta)$.
- In practice, the choice of the proposal distribution is absolutely crucial on the performance of the algorithm.
- In high dimensional problems, a simple MH algorithm will be useless. It will be necessary to use a combination of MH kernels.... However for the time being you might not have realized the power of the mixture and composition of kernels.

Metropolis one-at-a time

- Consider the target distribution $\pi\left(\theta_1, \theta_2\right)$.

Metropolis one-at-a time

- Consider the target distribution $\pi(\theta_1, \theta_2)$.
- We use two MH kernels to sample from this distribution,

Metropolis one-at-a time

- Consider the target distribution $\pi(\theta_1, \theta_2)$.
- We use two MH kernels to sample from this distribution,
  - the kernel K_1 updates θ_1 and keeps θ_2 fixed whereas

Metropolis one-at-a time

- Consider the target distribution $\pi(\theta_1, \theta_2)$.
- We use two MH kernels to sample from this distribution,
  - the kernel K_1 updates θ_1 and keeps θ_2 fixed whereas
  - the kernel K_2 updates θ_2 and keeps θ_1 fixed.

Metropolis one-at-a time

- Consider the target distribution $\pi(\theta_1, \theta_2)$.
- We use two MH kernels to sample from this distribution,
  - the kernel K_1 updates θ_1 and keeps θ_2 fixed whereas
  - the kernel K_2 updates θ_2 and keeps θ_1 fixed.
- We then combine these kernels through mixture or composition.

- The proposal $\overline{q}_1(\theta, \theta')$ associated to $K_1(\theta, \theta')$ is given by

$$\overline{q}_1(\theta, \theta') = \overline{q}_1((\theta_1, \theta_2), (\theta'_1, \theta'_2)) = q_1((\theta_1, \theta_2), \theta'_1) \, \delta_{\theta_2}(\theta'_2).$$

- The proposal $\overline{q}_1\left(\theta,\theta'\right)$ associated to $K_1\left(\theta,\theta'\right)$ is given by

$$\overline{q}_1\left(\theta,\theta'\right)=\overline{q}_1\left(\left(\theta_1,\theta_2\right),\left(\theta_1',\theta_2'\right)\right)=q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)\delta_{\theta_2}\left(\theta_2'\right).$$

- The acceptance probability is given by $\alpha_1\left(\theta,\theta'\right)=\min\left(1,r_1\left(\theta,\theta'\right)\right)$ where

$$\begin{aligned}
r_1\left(\theta,\theta'\right) &= \frac{\pi\left(\theta'\right)\overline{q}_1\left(\theta',\theta\right)}{\pi\left(\theta\right)\overline{q}_1\left(\theta,\theta'\right)}=\frac{\pi\left(\theta_1',\theta_2'\right)q_1\left(\left(\theta_1',\theta_2'\right),\theta_1\right)\delta_{\theta_2'}\left(\theta_2\right)}{\pi\left(\theta_1,\theta_2\right)q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)\delta_{\theta_2}\left(\theta_2'\right)} \\
&= \frac{\pi\left(\theta_1',\theta_2\right)q_1\left(\left(\theta_1',\theta_2\right),\theta_1\right)}{\pi\left(\theta_1,\theta_2\right)q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)} \\
&= \frac{\pi\left(\theta_1'\middle|\theta_2\right)q_1\left(\left(\theta_1',\theta_2\right),\theta_1\right)}{\pi\left(\theta_1\middle|\theta_2\right)q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)}.
\end{aligned}$$

- The proposal $\overline{q}_1(\theta,\theta')$ associated to $K_1(\theta,\theta')$ is given by

$$\overline{q}_1(\theta,\theta') = \overline{q}_1((\theta_1,\theta_2),(\theta_1',\theta_2')) = q_1((\theta_1,\theta_2),\theta_1')\,\delta_{\theta_2}(\theta_2').$$

- The acceptance probability is given by $\alpha_1(\theta,\theta') = \min(1, r_1(\theta,\theta'))$ where

$$\begin{aligned} r_1(\theta,\theta') &= \frac{\pi(\theta')\,\overline{q}_1(\theta',\theta)}{\pi(\theta)\,\overline{q}_1(\theta,\theta')} = \frac{\pi(\theta_1',\theta_2')\,q_1((\theta_1',\theta_2'),\theta_1)\,\delta_{\theta_2'}(\theta_2)}{\pi(\theta_1,\theta_2)\,q_1((\theta_1,\theta_2),\theta_1')\,\delta_{\theta_2}(\theta_2')} \\ &= \frac{\pi(\theta_1',\theta_2)\,q_1((\theta_1',\theta_2),\theta_1)}{\pi(\theta_1,\theta_2)\,q_1((\theta_1,\theta_2),\theta_1')} \\ &= \frac{\pi(\theta_1'|\,\theta_2)\,q_1((\theta_1',\theta_2),\theta_1)}{\pi(\theta_1|\,\theta_2)\,q_1((\theta_1,\theta_2),\theta_1')}. \end{aligned}$$

- This move is also equivalent to an MH step of invariant distribution $\pi(\theta_1|\,\theta_2)$.

- The proposal $\overline{q}_2\left(\theta,\theta'\right)$ associated to $K_2\left(\theta,\theta'\right)$ is given by

$$\overline{q}_2\left(\theta,\theta'\right)=\overline{q}_2\left(\left(\theta_1,\theta_2\right),\left(\theta_1',\theta_2'\right)\right)=\delta_{\theta_1}\left(\theta_1'\right)q_2\left(\left(\theta_1,\theta_2\right),\theta_2'\right).$$

- The proposal $\overline{q}_2\left(\theta,\theta'\right)$ associated to $K_2\left(\theta,\theta'\right)$ is given by

$$\overline{q}_2\left(\theta,\theta'\right)=\overline{q}_2\left(\left(\theta_1,\theta_2\right),\left(\theta'_1,\theta'_2\right)\right)=\delta_{\theta_1}\left(\theta'_1\right)q_2\left(\left(\theta_1,\theta_2\right),\theta'_2\right).$$

- The acceptance probability is given by $\alpha_2\left(\theta,\theta'\right)=\min\left(1,r_2\left(\theta,\theta'\right)\right)$ where

$$\begin{aligned}
r\left(\theta,\theta'\right) &= \frac{\pi\left(\theta'\right)\overline{q}_2\left(\theta',\theta\right)}{\pi\left(\theta\right)\overline{q}_2\left(\theta,\theta'\right)}=\frac{\pi\left(\theta'_1,\theta'_2\right)\delta_{\theta'_1}\left(\theta_1\right)q_2\left(\left(\theta'_1,\theta'_2\right),\theta_2\right)}{\pi\left(\theta_1,\theta_2\right)\delta_{\theta_1}\left(\theta'_1\right)q_2\left(\left(\theta_1,\theta_2\right),\theta'_2\right)} \\
&= \frac{\pi\left(\theta_1,\theta'_2\right)q_2\left(\left(\theta_1,\theta'_2\right),\theta_2\right)}{\pi\left(\theta_1,\theta_2\right)q_2\left(\left(\theta_1,\theta_2\right),\theta'_2\right)} \\
&= \frac{\pi\left(\theta'_2\middle|\,\theta_1\right)q_2\left(\left(\theta_1,\theta'_2\right),\theta_2\right)}{\pi\left(\theta_2\middle|\,\theta_1\right)q_2\left(\left(\theta_1,\theta_2\right),\theta'_2\right)}.
\end{aligned}$$

- The proposal $\overline{q}_2(\theta, \theta')$ associated to $K_2(\theta, \theta')$ is given by

$$\overline{q}_2(\theta, \theta') = \overline{q}_2((\theta_1, \theta_2), (\theta_1', \theta_2')) = \delta_{\theta_1}(\theta_1') \, q_2((\theta_1, \theta_2), \theta_2').$$

- The acceptance probability is given by $\alpha_2(\theta, \theta') = \min(1, r_2(\theta, \theta'))$ where

$$\begin{aligned} r(\theta, \theta') &= \frac{\pi(\theta') \overline{q}_2(\theta', \theta)}{\pi(\theta) \overline{q}_2(\theta, \theta')} = \frac{\pi(\theta_1', \theta_2') \delta_{\theta_1'}(\theta_1) q_2((\theta_1', \theta_2'), \theta_2)}{\pi(\theta_1, \theta_2) \delta_{\theta_1}(\theta_1') q_2((\theta_1, \theta_2), \theta_2')} \\ &= \frac{\pi(\theta_1, \theta_2') q_2((\theta_1, \theta_2'), \theta_2)}{\pi(\theta_1, \theta_2) q_2((\theta_1, \theta_2), \theta_2')} \\ &= \frac{\pi(\theta_2' | \theta_1) q_2((\theta_1, \theta_2'), \theta_2)}{\pi(\theta_2 | \theta_1) q_2((\theta_1, \theta_2), \theta_2')}. \end{aligned}$$

- This move is also equivalent to an MH step of invariant distribution $\pi(\theta_2 | \theta_1)$.

- Assume we use a composition of these kernels, then the resulting algorithm proceeds as follows at iteration i.
  MH step to update component 1

- Assume we use a composition of these kernels, then the resulting algorithm proceeds as follows at iteration i.
  MH step to update component 1
- Sample $\theta_1^* \sim q_1\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \cdot\right)$ and compute

$$\begin{aligned} & \alpha_1\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \left(\theta_1^*, \theta_2^{(i-1)}\right)\right) \\ = & \min\left(1, \frac{\pi\left(\theta_1^* \middle| \theta_2^{(i-1)}\right) q_1\left(\left(\theta_1^*, \theta_2^{(i-1)}\right), \theta_1^{(i-1)}\right)}{\pi\left(\theta_1^{(i-1)} \middle| \theta_2^{(i-1)}\right) q_1\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \theta_1^*\right)}\right) \end{aligned}$$

- Assume we use a composition of these kernels, then the resulting algorithm proceeds as follows at iteration i.
  MH step to update component 1
- Sample $\theta_1^* \sim q_1\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \cdot\right)$ and compute

$$\begin{aligned}
&\alpha_1\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \left(\theta_1^*, \theta_2^{(i-1)}\right)\right) \\
= \quad &\min\left(1, \frac{\pi\left(\theta_1^* \middle| \theta_2^{(i-1)}\right) q_1\left(\left(\theta_1^*, \theta_2^{(i-1)}\right), \theta_1^{(i-1)}\right)}{\pi\left(\theta_1^{(i-1)} \middle| \theta_2^{(i-1)}\right) q_1\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \theta_1^*\right)}\right)
\end{aligned}$$

- With probability $\alpha_1\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \left(\theta_1^*, \theta_2^{(i-1)}\right)\right)$, set $\theta_1^{(i)} = \theta_1^*$ and otherwise $\theta_1^{(i)} = \theta_1^{(i-1)}$.

MH step to update component 2

- Sample $\theta_2^* \sim q_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \cdot\right)$ and compute

$$
\begin{aligned}
& \alpha_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \left(\theta_1^{(i)}, \theta_2^{*}\right)\right) \\
= \; & \min\left(1, \frac{\pi\left(\theta_2^* \middle| \theta_1^{(i)}\right) q_2\left(\left(\theta_1^{(i)}, \theta_2^*\right), \theta_2^{(i-1)}\right)}{\pi\left(\theta_2^{(i-1)} \middle| \theta_1^{(i)}\right) q_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \theta_2^*\right)}\right)
\end{aligned}
$$

MH step to update component 2

- Sample $\theta_2^* \sim q_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \cdot\right)$ and compute

$$\begin{aligned}&\alpha_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \left(\theta_1^{(i)}, \theta_2^*\right)\right)\\ = \quad &\min\left(1, \frac{\pi\left(\theta_2^* \middle| \theta_1^{(i)}\right) q_2\left(\left(\theta_1^{(i)}, \theta_2^*\right), \theta_2^{(i-1)}\right)}{\pi\left(\theta_2^{(i-1)} \middle| \theta_1^{(i)}\right) q_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \theta_2^*\right)}\right)\end{aligned}$$

- With probability $\alpha_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \left(\theta_1^{(i)}, \theta_1^*\right)\right)$, set $\theta_2^{(i)} = \theta_2^*$ otherwise $\theta_2^{(i)} = \theta_2^{(i-1)}$.

- Assume we use a even mixture of these kernels, then the resulting algorithm proceeds as follows at iteration i.

- Assume we use a even mixture of these kernels, then the resulting algorithm proceeds as follows at iteration i.
- Sample the index of the component to update $J \sim U\{1, 2\}$.

- Assume we use a even mixture of these kernels, then the resulting algorithm proceeds as follows at iteration i.
- Sample the index of the component to update $J \sim U\{1, 2\}$.
- Set $\theta_{-J}^{(i)} = \theta_{-J}^{(i-1)}$.

- Assume we use a even mixture of these kernels, then the resulting algorithm proceeds as follows at iteration i.
- Sample the index of the component to update $J \sim U\{1, 2\}$.
- Set $\theta_{-J}^{(i)} = \theta_{-J}^{(i-1)}$.
- Sample $\theta_J^* \sim q_J\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \cdot\right)$ and compute

$$
\begin{aligned}
& \alpha_J\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \left(\theta_J^*, \theta_{-J}^{(i)}\right)\right) \\
= \quad & \min\left(1, \frac{\pi\left(\theta_J^* \middle| \theta_{-J}^{(i)}\right) q_J\left(\left(\theta_J^*, \theta_{-J}^{(i)}\right), \theta_J^{(i-1)}\right)}{\pi\left(\theta_J^{(i-1)} \middle| \theta_{-J}^{(i)}\right) q_K\left(\left(\theta_J^{(i-1)}, \theta_{-J}^{(i)}\right), \theta_J^*\right)}\right).
\end{aligned}
$$

- Assume we use a even mixture of these kernels, then the resulting algorithm proceeds as follows at iteration i.
- Sample the index of the component to update $J \sim U\{1,2\}$.
- Set $\theta_{-J}^{(i)} = \theta_{-J}^{(i-1)}$.
- Sample $\theta_J^* \sim q_J\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \cdot\right)$ and compute

$$\begin{aligned} & \alpha_J\left(\left(\theta_1^{(i-1)}, \theta_2^{(i-1)}\right), \left(\theta_J^*, \theta_{-J}^{(i)}\right)\right) \\ = \ & \min\left(1, \frac{\pi\left(\theta_J^* \middle| \theta_{-J}^{(i)}\right) q_J\left(\left(\theta_J^*, \theta_{-J}^{(i)}\right), \theta_J^{(i-1)}\right)}{\pi\left(\theta_J^{(i-1)} \middle| \theta_{-J}^{(i)}\right) q_K\left(\left(\theta_J^{(i-1)}, \theta_{-J}^{(i)}\right), \theta_J^*\right)}\right). \end{aligned}$$

- With probability $\alpha_J\left(\left(\theta_J^{(i-1)}, \theta_J^{(i-1)}\right), \left(\theta_J^*, \theta_{-J}^{(i)}\right)\right)$, set $\theta_J^{(i)} = \theta_J^*$ otherwise $\theta_J^{(i)} = \theta_J^{(i-1)}$.

- It is clear that in such cases both K_1 and K_2 are NOT irreducible and aperiodic as each of them only update one component.

- It is clear that in such cases both K_1 and K_2 are NOT irreducible and aperiodic as each of them only update one component.
- However, the composition and mixture of these kernels can be irreducible and aperiodic because then all the components are updated.

- Consider now the case where

$$q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)=\pi\left(\theta_1'\middle|\,\theta_2\right).$$

then

$$r_1\left(\theta,\theta'\right)=\frac{\pi\left(\theta_1'\middle|\,\theta_2\right)q_1\left(\left(\theta_1',\theta_2\right),\theta_1\right)}{\pi\left(\theta_1\middle|\,\theta_2\right)q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)}=\frac{\pi\left(\theta_1'\middle|\,\theta_2\right)\pi\left(\theta_1\middle|\,\theta_2\right)}{\pi\left(\theta_1\middle|\,\theta_2\right)\pi\left(\theta_1'\middle|\,\theta_2\right)}=1$$

- Consider now the case where

$$q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)=\pi\left(\left.\theta_1'\right|\theta_2\right).$$

then

$$r_1\left(\theta,\theta'\right)=\frac{\pi\left(\left.\theta_1'\right|\theta_2\right)q_1\left(\left(\theta_1',\theta_2\right),\theta_1\right)}{\pi\left(\left.\theta_1\right|\theta_2\right)q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)}=\frac{\pi\left(\left.\theta_1'\right|\theta_2\right)\pi\left(\left.\theta_1\right|\theta_2\right)}{\pi\left(\left.\theta_1\right|\theta_2\right)\pi\left(\left.\theta_1'\right|\theta_2\right)}=1$$

- Similarly if $q_2\left(\left(\theta_1,\theta_2\right),\theta_2'\right)=\pi\left(\left.\theta_2'\right|\theta_1\right)$ then $r_2\left(\theta,\theta'\right)=1$.

- Consider now the case where

$$q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)=\pi\left(\theta_1'\middle|\,\theta_2\right).$$

then

$$r_1\left(\theta,\theta'\right)=\frac{\pi\left(\theta_1'\middle|\,\theta_2\right)q_1\left(\left(\theta_1',\theta_2\right),\theta_1\right)}{\pi\left(\theta_1|\,\theta_2\right)q_1\left(\left(\theta_1,\theta_2\right),\theta_1'\right)}=\frac{\pi\left(\theta_1'\middle|\,\theta_2\right)\pi\left(\theta_1|\,\theta_2\right)}{\pi\left(\theta_1|\,\theta_2\right)\pi\left(\theta_1'\middle|\,\theta_2\right)}=1$$

- Similarly if $q_2\left(\left(\theta_1,\theta_2\right),\theta_2'\right)=\pi\left(\theta_2'\middle|\,\theta_1\right)$ then $r_2\left(\theta,\theta'\right)=1$.
- If you take for proposal distributions in the MH kernels the full conditional distributions then you have the Gibbs sampler!

- Generally speaking, to sample from $\pi(\theta)$ where $\theta = (\theta_1, ..., \theta_p)$, we can use the following algorithm at iteration i.

- Generally speaking, to sample from $\pi(\theta)$ where $\theta = (\theta_1, ..., \theta_p)$, we can use the following algorithm at iteration i.
- Iteration i; $i \geq 1$:

- Generally speaking, to sample from $\pi(\theta)$ where $\theta = (\theta_1, ..., \theta_p)$, we can use the following algorithm at iteration i.
- Iteration i; $i \geq 1$:
  - For $k = 1 : p$

- Generally speaking, to sample from $\pi(\theta)$ where $\theta = (\theta_1, ..., \theta_p)$, we can use the following algorithm at iteration i.
- Iteration i; $i \geq 1$:
  - For $k = 1 : p$
    - Sample $\theta_k^{(i)}$ using an MH step of proposal distribution $q_k\left(\left(\theta_{-k}^{(i)}, \theta_k^{(i-1)}\right), \theta_k'\right)$ and target $\pi\left(\theta_k | \theta_{-k}^{(i)}\right)$ where $\theta_{-k}^{(i)} = \left(\theta_1^{(i)}, ..., \theta_{k-1}^{(i)}, \theta_{k+1}^{(i-1)}, ..., \theta_p^{(i-1)}\right)$.

- If we have $q_k\left(\theta_{1:p}, \theta_k'\right) = \pi\left(\theta_k' \middle| \theta_{-k}\right)$ then we are back to the Gibbs sampler.

- If we have $q_k\left(\theta_{1:p},\theta_k'\right)=\pi\left(\theta_k'\middle|\,\theta_{-k}\right)$ then we are back to the Gibbs sampler.
- We can update some parameters according to $\pi\left(\theta_k'\middle|\,\theta_{-k}\right)$ (and the move is automatically accepted) and others according to different proposals.

- If we have $q_k\left(\theta_{1:p},\theta'_k\right)=\pi\left(\theta'_k\middle|\theta_{-k}\right)$ then we are back to the Gibbs sampler.
- We can update some parameters according to $\pi\left(\theta'_k\middle|\theta_{-k}\right)$ (and the move is automatically accepted) and others according to different proposals.
- **Example**: Assume we have $\pi\left(\theta_1,\theta_2\right)$ where it is easy to sample from $\pi\left(\theta_1\middle|\theta_2\right)$ and then use an MH step of invariant distribution $\pi\left(\theta_2\middle|\theta_1\right)$.

At iteration i.

- Sample $\theta_1^{(i)} \sim \pi\left(\theta_1 \middle| \theta_2^{(i-1)}\right)$.

At iteration i.

- Sample $\theta_1^{(i)} \sim \pi\left(\theta_1 \middle| \theta_2^{(i-1)}\right)$.
- Sample $\theta_2^{(i)}$ using one MH step of proposal distribution $q_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \theta_2\right)$ and target $\pi\left(\theta_2 \middle| \theta_1^{(i)}\right)$.

At iteration i.

- Sample $\theta_1^{(i)} \sim \pi\left(\theta_1 \middle| \theta_2^{(i-1)}\right)$.
- Sample $\theta_2^{(i)}$ using one MH step of proposal distribution $q_2\left(\left(\theta_1^{(i)}, \theta_2^{(i-1)}\right), \theta_2\right)$ and target $\pi\left(\theta_2 \middle| \theta_1^{(i)}\right)$.
- There is NO NEED to run the MH algorithm multiple steps to ensure that $\theta_2^{(i)} \sim \pi\left(\theta_2 \middle| \theta_2^{(i-1)}\right)$.

- In practice, we divide the parameter space $\theta = (\theta_1, ..., \theta_p)$.

- In practice, we divide the parameter space $\theta = (\theta_1, ..., \theta_p)$.
- We update each parameter θ_k according to an MH step of propsal distribution $q_k\left(\theta_{1:p}, \theta'_k\right) = q_k\left(\left(\theta_{-k}, \theta_k\right), \theta'_k\right)$ and invariant distribution $\pi\left(\theta_k \middle| \theta_{-k}\right)$.

- In practice, we divide the parameter space $\theta = (\theta_1, ..., \theta_p)$.
- We update each parameter θ_k according to an MH step of propsal distribution $q_k \left(\theta_{1:p}, \theta'_k\right) = q_k \left(\left(\theta_{-k}, \theta_k\right), \theta'_k\right)$ and invariant distribution $\pi \left(\theta_k \middle| \theta_{-k}\right)$.
- You are now equipped to fit advanced statistical models...