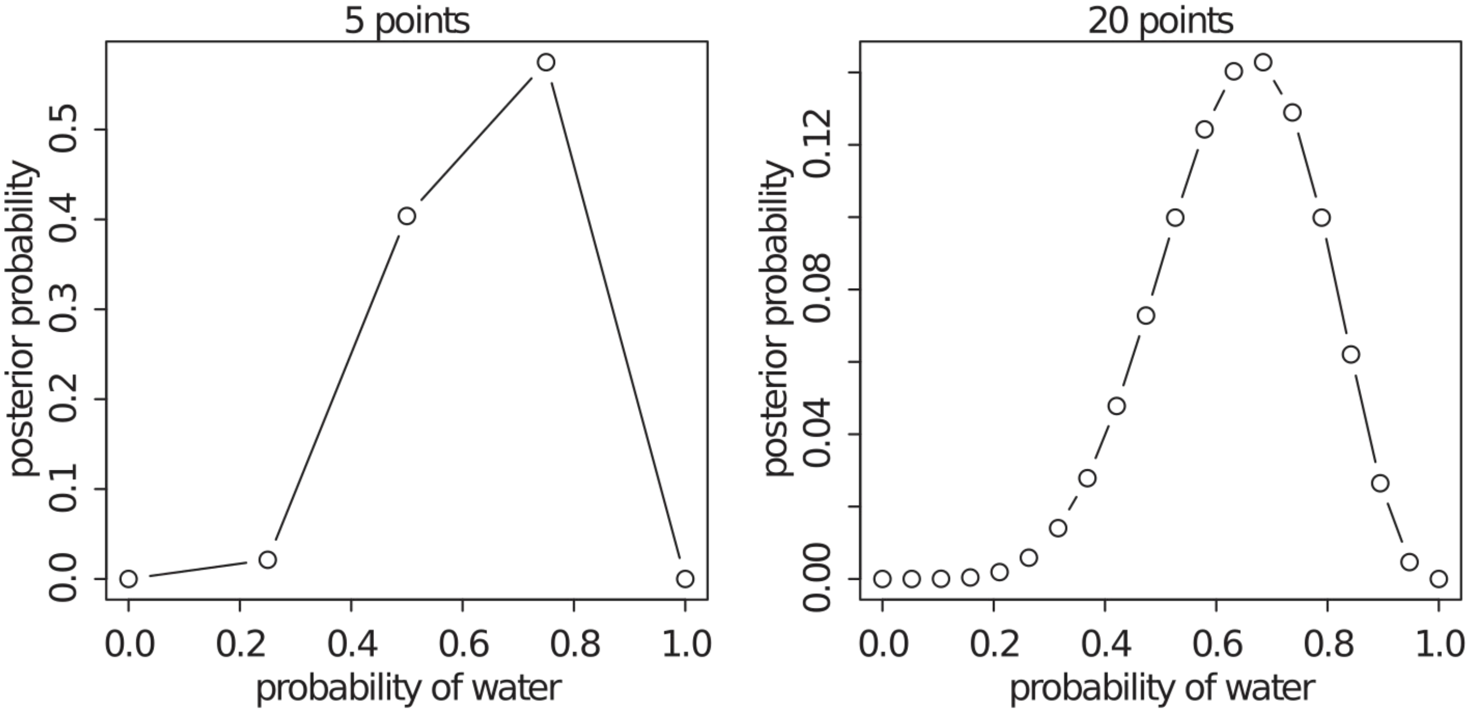
Introdução ao MCMC

Com mais equações, admito

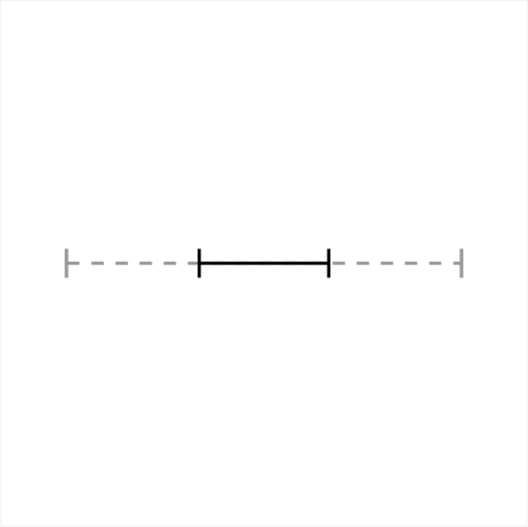
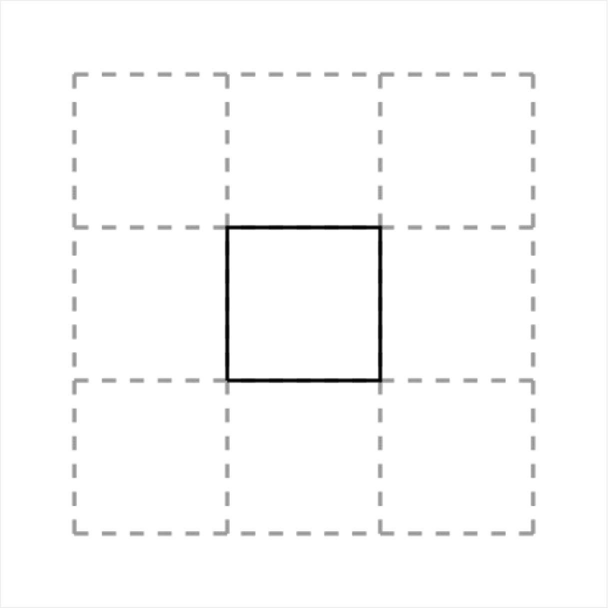
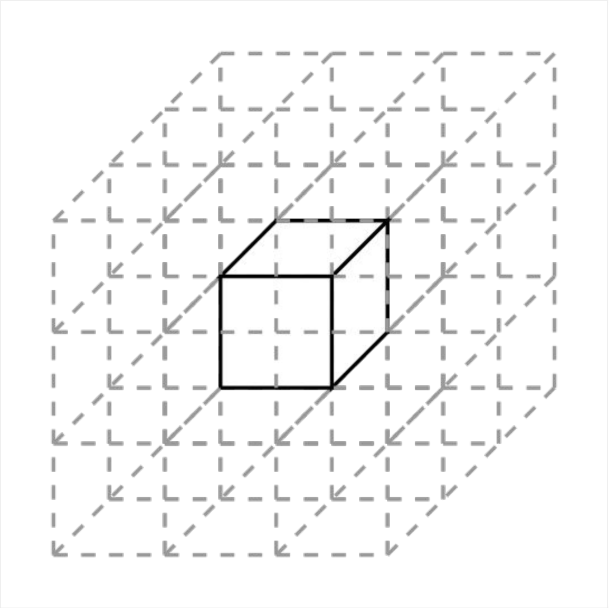
Como calcular as posteriores?

- Nos casos que vimos até agora, é simples analisar e tirar conclusões a partir da distribuição dos parâmetros
- Isso nem sempre é verdade
- Quando isso acontece, apelamos para métodos computacionais

Método força bruta



Método força bruta em alta dimensionalidade não funciona

Objetivo geral de um método de amostragem

- Ao invés de calcular a posterior em todas as possibilidades de um parâmetro, vamos tentar amostrar só as que têm probabilidade posterior alta

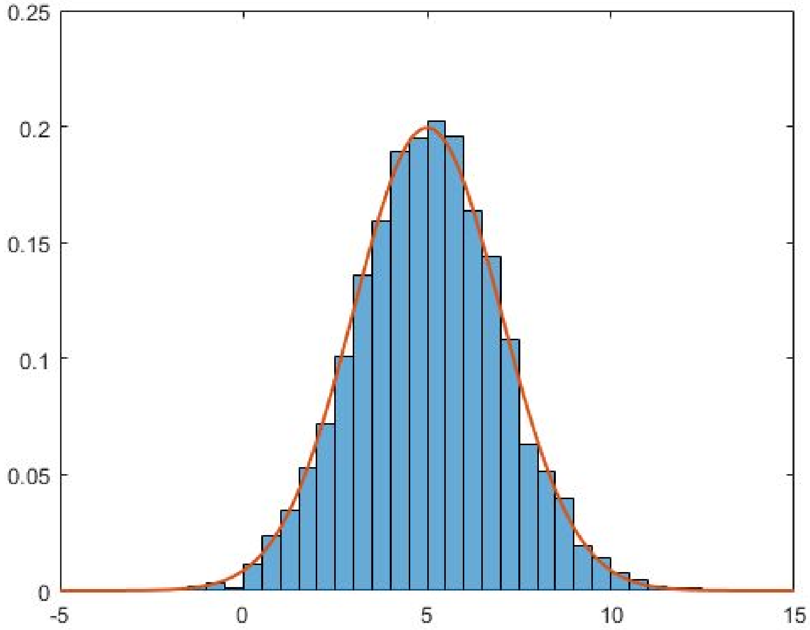
Objetivo geral de um método de amostragem

- Ao invés de calcular a posterior em todas as possibilidades de um parâmetro, vamos tentar amostrar só as que têm probabilidade posterior alta
- Ao invés de ter uma expressão fechada para a posterior, vamos ter um monte de valores, com proporção igual às probabilidades da posterior

Amostra da posterior

$$\{\theta_1, \theta_2, \theta_3, ...\} \sim P(\theta|y)$$

$$E[\theta] = \frac{1}{N}\sum_{i=1}^{N}\theta_i$$



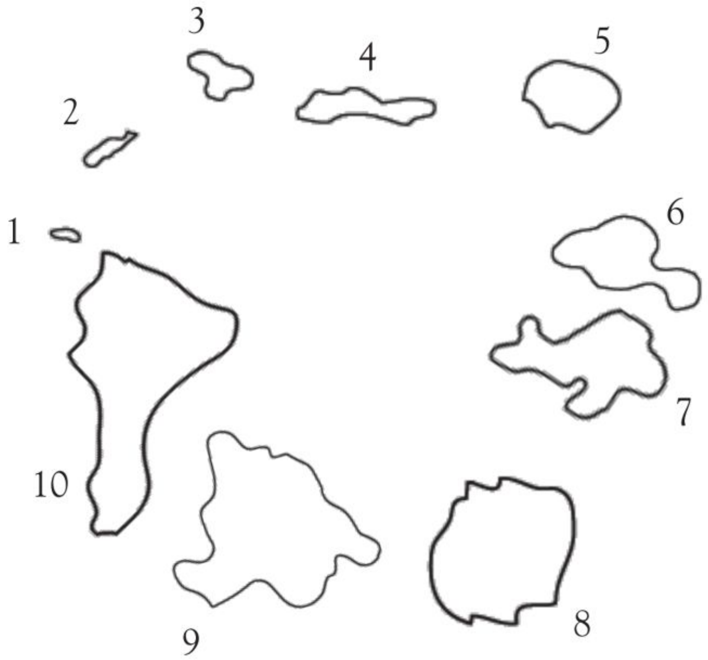
Como fazer uma amostra?

- Distribuições simples podem ser amostradas usando um gerador de números aleatórios e um pouco de matemática.
- Normal, binomial, Poisson, ... todas são bem simples e podem ser usadas como blocos para montar distribuições mais complicadas. (todas as funções r*algumacoisa* do R)

Como fazer uma amostra?

- Distribuições simples podem ser amostradas usando um gerador de números aleatórios e um pouco de matemática.
- Normal, binomial, Poisson, ... todas são bem simples e podem ser usadas como blocos para montar distribuições mais complicadas. (todas as funções r*algumacoisa* do R)
- **Quando não tem jeito fácil, usamos MCMC.**

Reino do Bom Rei Markov



- Cada ilha tem população proporcional ao seu número
- O Bom Rei quer visitar todas as ilhas, mas passar em cada uma uma quantidade de tempo que seja proporcional à população
- O Rei tem preguiça de montar um cronograma, então ele desenvolveu um conjunto de regras para suas viagens

O algoritmo do Rei Markov

1. Toda semana o Rei decide se ele vai viajar para alguma das ilhas vizinhas ou adiar a viagem. Para decidir, ele joga uma moeda. Se der cara ele considera ir para a ilha que está no sentido horário, se der coroa no sentido anti horário.
2. Decidida a direção, ele decide se vai viajar ou ficar. Para isso, ele pega um número de conchas proporcional à população da ilha para a qual ele pode ir. Depois, ele pega um número de pedras proporcional à população da ilha que ele está.
3. Se ele tiver mais conchas que pedras, ele viaja. Se não, ele joga fora um número de pedras igual ao número de conchas e coloca a pedra e as conchas que sobraram num saco. Depois, ele sorteia desse saco. Se der concha ele muda de ilha.

```
num_weeks <- 1e5
positions <- rep(0,num_weeks)
current <- 10
for ( i in 1:num_weeks ) {
     # record current position
     positions[i] <- current
     # flip coin to generate proposal
     proposal <- current + sample( c(-1,1) , size=1 )
     # now make sure he loops around the archipelago
     if ( proposal < 1 ) proposal <- 10
     if ( proposal > 10 ) proposal <- 1
     # move?
     prob_move <- proposal/current
     current <- ifelse( runif(1) < prob_move , proposal ,
     current )
}
plot(table(positions))
```

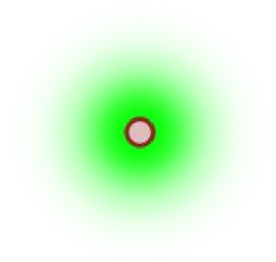
O algoritmo de Metropolis

1. Começa com qualquer valor do(s) parâmetro(s)
2. Sorteia uma mudança no parâmetro.
3. Decidido qual a mudança candidata, verificamos se a probabilidade posterior aumentou ou diminuiu.
4. Se a probabilidade posterior aumentou, aceitamos o novo parâmetro. Se diminuiu, aceitamos o novo parâmetro com probabilidade igual a razão entre a probabilidade posterior candidata e a atual.

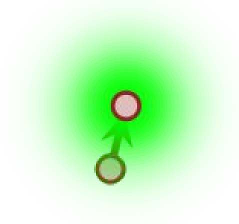
Por que essa maluquice funciona?

- A ideia é montar uma cadeia de Markov que tenha como estado estacionário a distribuição alvo
- A escolha da probabilidade de transição usando a razão de probabilidades posteriores garante que a cadeia chegue na distribuição desejada
- Se a cadeia respeitar alguns pré-requisitos, ela converge para a distribuição que queremos.
  - Ergoticidade
  - Balanço detalhado

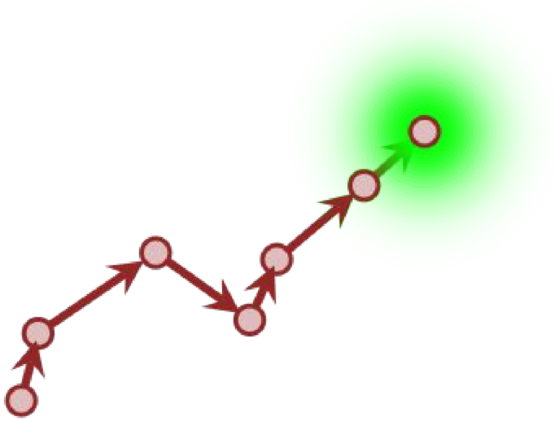
Caminhando no espaço de parâmetros



Caminhando no espaço de parâmetros



Caminhando no espaço de parâmetros



Transição de Metropolis

$$\theta_0 \to \theta_1 \to \theta_2 \to \cdots \sim P(\theta|y)$$

Transição de Metropolis

$$\theta_0 \to \theta_1 \to \theta_2 \to \cdots \sim P(\theta|y)$$

Distribuição da proposta: $\theta^* \sim q(\theta^*|\theta_{t-1})$

Transição de Metropolis

$$\theta_0 \rightarrow \theta_1 \rightarrow \theta_2 \rightarrow \cdots \sim P(\theta|y)$$

Distribuição da proposta: $\theta^* \sim q(\theta^*|\theta_{t-1})$

Caso onde a distribuição é simétrica:

$$q(\theta^*|\theta_{t-1}) = q(\theta_{t-1}|\theta^*)$$

Transição de Metropolis

$$\theta_0 \rightarrow \theta_1 \rightarrow \theta_2 \rightarrow \cdots \sim P(\theta|y)$$

Distribuição da proposta: $\theta^* \sim q(\theta^*|\theta_{t-1})$

Caso onde a distribuição é simétrica:

$$q(\theta^*|\theta_{t-1}) = q(\theta_{t-1}|\theta^*)$$

Probabilidade de aceitar a proposta: $P(\theta_t \leftarrow \theta^*) = min(1, \frac{P(\theta^*|y)}{P(\theta_{t-1}|y)})$

Transição de Metropolis-Hastings

$$\theta_0 \rightarrow \theta_1 \rightarrow \theta_2 \rightarrow \cdots \sim P(\theta|y)$$

Distribuição da proposta: $\theta^* \sim q(\theta^*|\theta_{t-1})$

Caso onde a distribuição não é simétrica:

$$q(\theta^*|\theta_{t-1}) \neq q(\theta_{t-1}|\theta^*)$$

Probabilidade de aceitar a proposta:

$$P(\theta_t \leftarrow \theta^*) = min(1, \frac{P(\theta^*|y)}{P(\theta_{t-1}|y)} \frac{q(\theta_{t-1}|\theta^*)}{q(\theta^*|\theta_{t-1})})$$

Condições de convergência

$$P(\theta|y)\sum_{\theta'\neq\theta} P(\theta'|\theta) = \sum_{\theta'\neq\theta} P(\theta'|y)P(\theta|\theta')$$

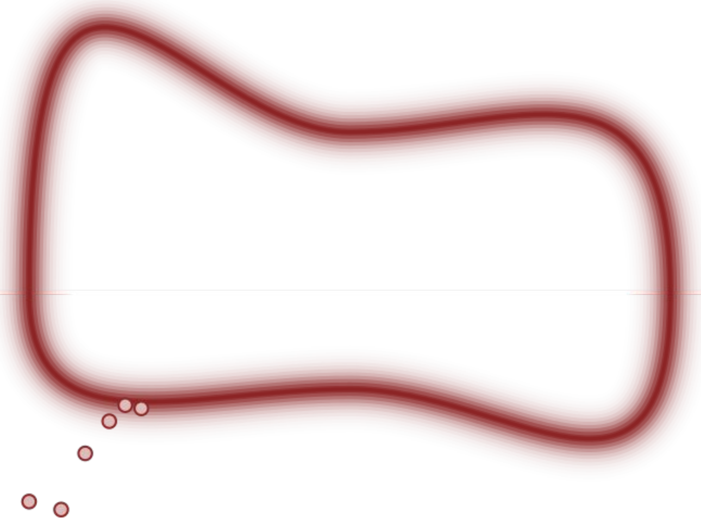
- Ergoticidade:
  - Todos os estados têm que ser possíveis de serem acessados. Não importa de onde vc começa, depois de um número suficiente de passos vc pode estar em qualquer lugar.
  - Difícil de provar.
  - Se a cadeia é ergótica, a distribuição estacionária é única.

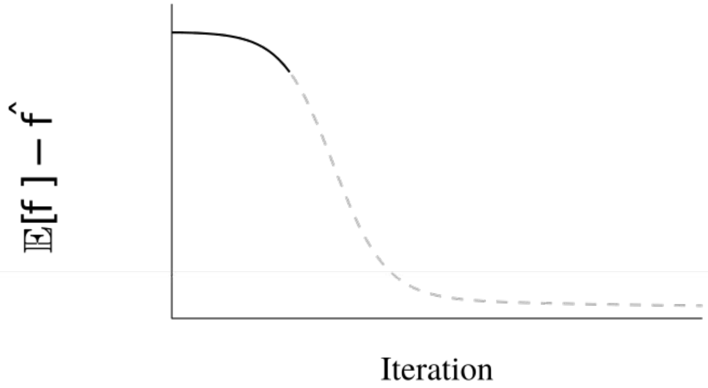
Condição forte de convergência

$$P(\theta|y)P(\theta'|\theta) = P(\theta'|y)P(\theta|\theta')$$

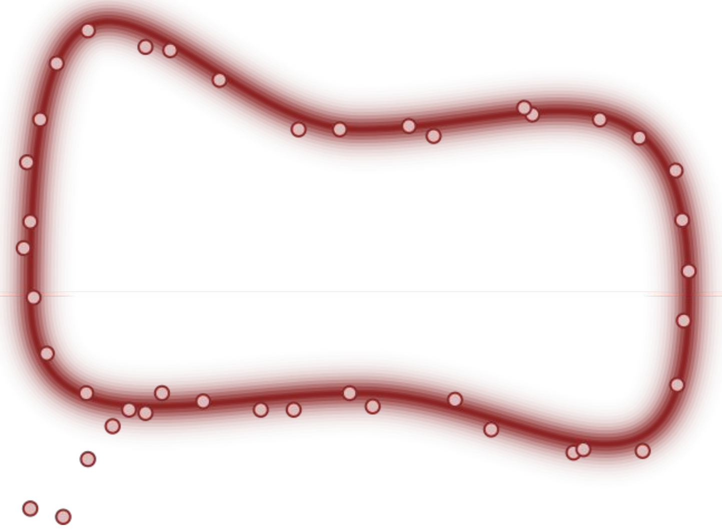
- Balanço detalhado:
  - Condição barra pesada.
  - Fácil de provar.
  - Garante unicidade da distribuição estacionária.

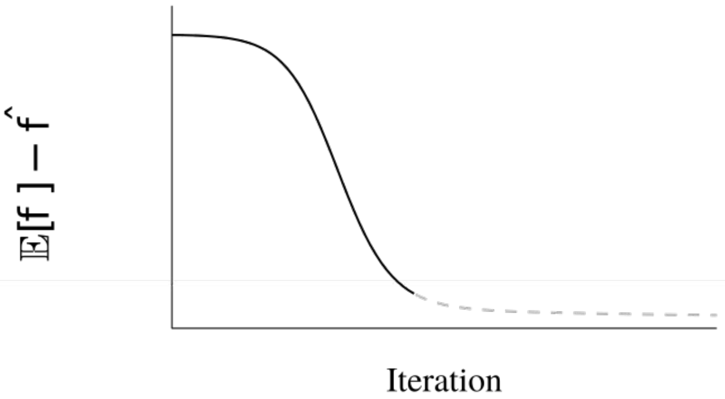
MCMC ideal



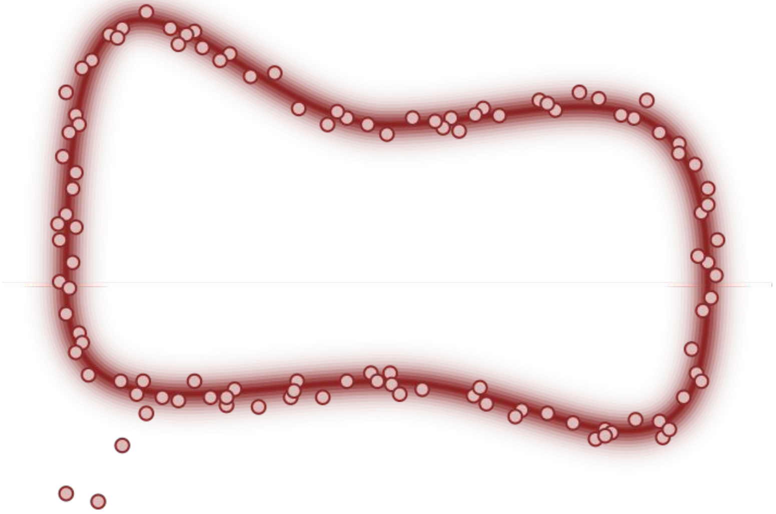


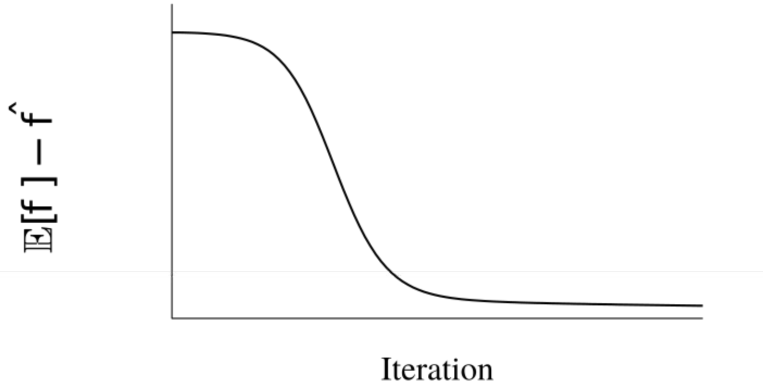
MCMC ideal

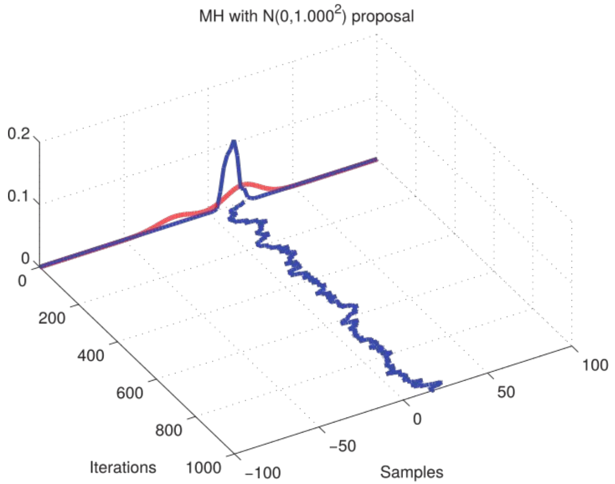




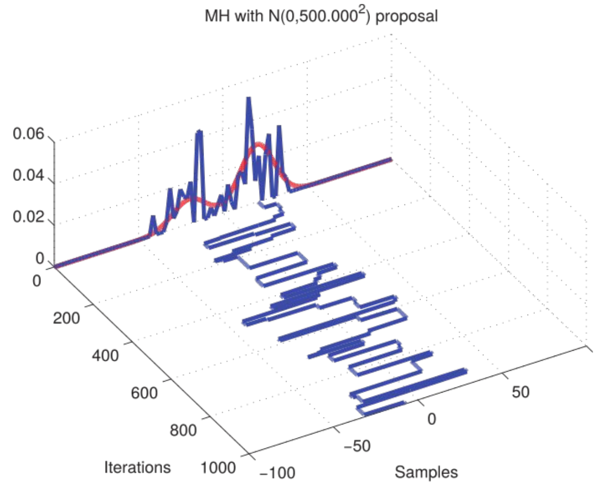
MCMC ideal



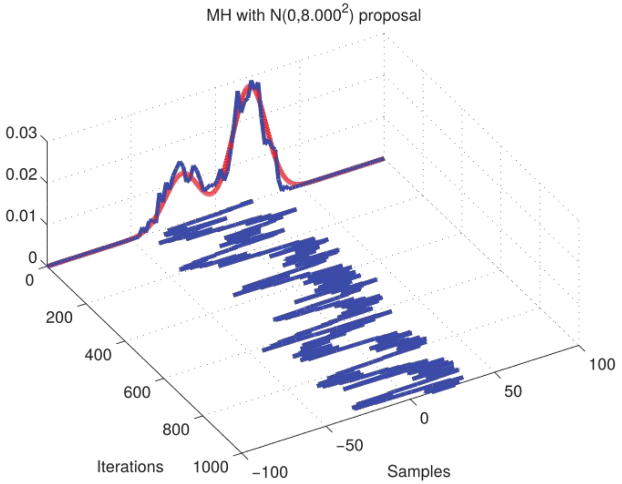



MH with N(0,1.000^2) proposal
0.2
0.1
0
0
200
400
600
800
1000
Iterations
−100
−50
0
50
100
Samples

(a)


MH with N(0,500.000^2) proposal
0.06
0.04
0.02
0
0
200
400
600
800
1000
Iterations
−100
−50
0
50
Samples

(b)


MH with N(0,8.000^2) proposal
0.03
0.02
0.01
0
0
200
400
600
800
1000
Iterations
−100
−50
0
50
100
Samples