Rumo ao desconhecido ...

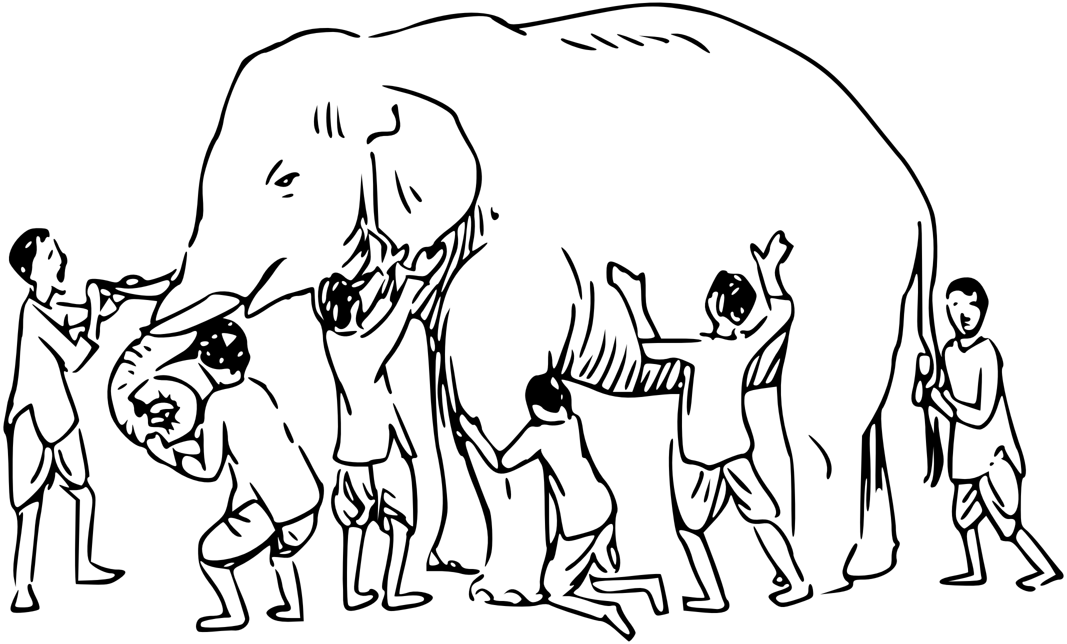
Rumo ao desconhecido ...

Embora para modelos mais simples seja possível derivar a distribuição posterior diretamente, sem o uso de simulações.

Usamos o MCMC pois não podemos derivar diretamente a posterior. A estimativa da posterior usando MCMC é como a exploração da superfície de um planeta desconhecido.

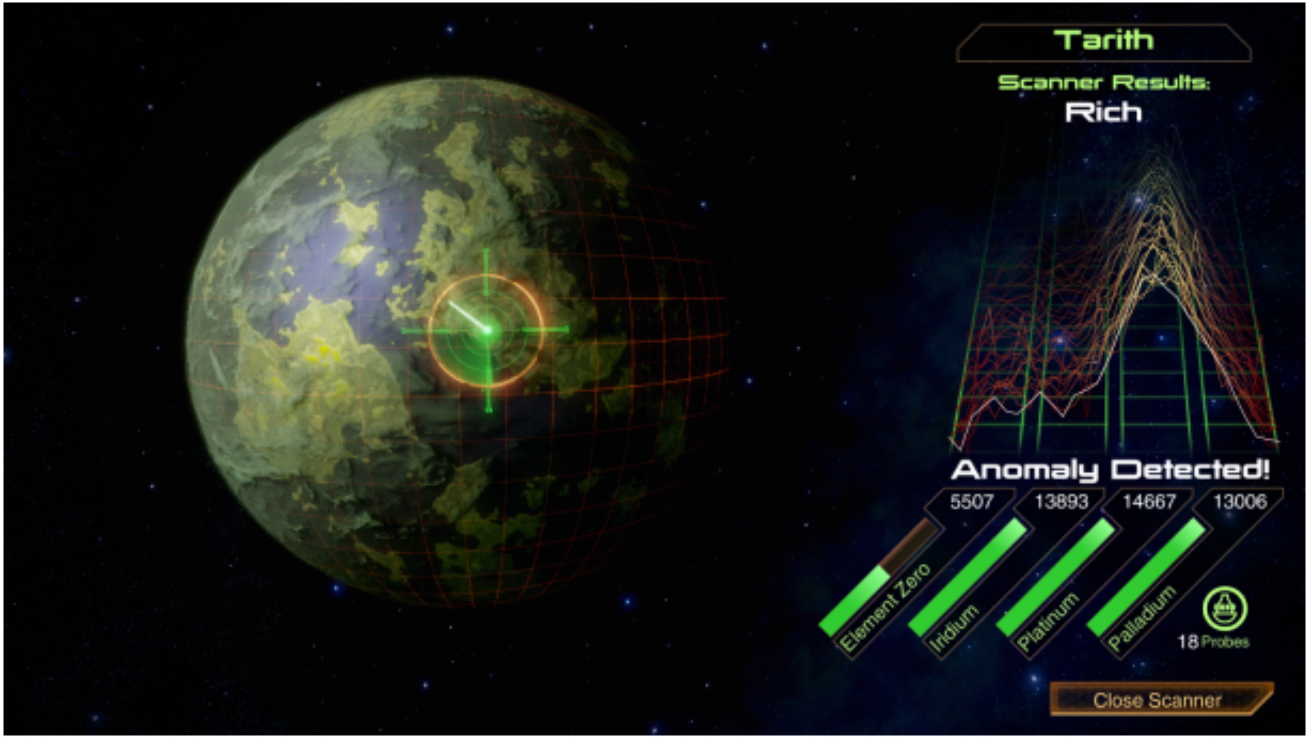
Rumo ao desconhecido ...

Técnicas usadas para explorar o desconhecido?



Rumo ao desconhecido ...

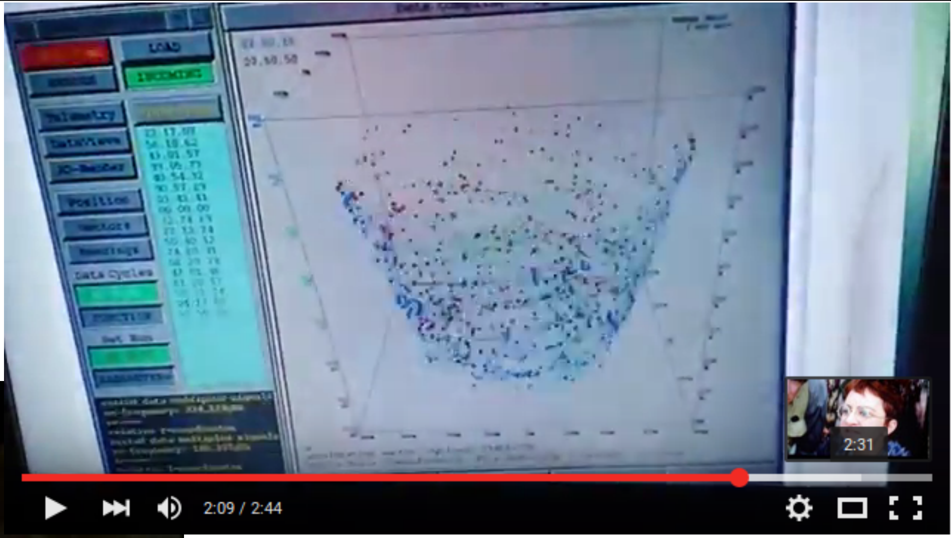
Técnicas usadas para explorar o desconhecido?



Mass Effect II

Rumo ao desconhecido ...

Técnicas usadas para explorar o desconhecido?

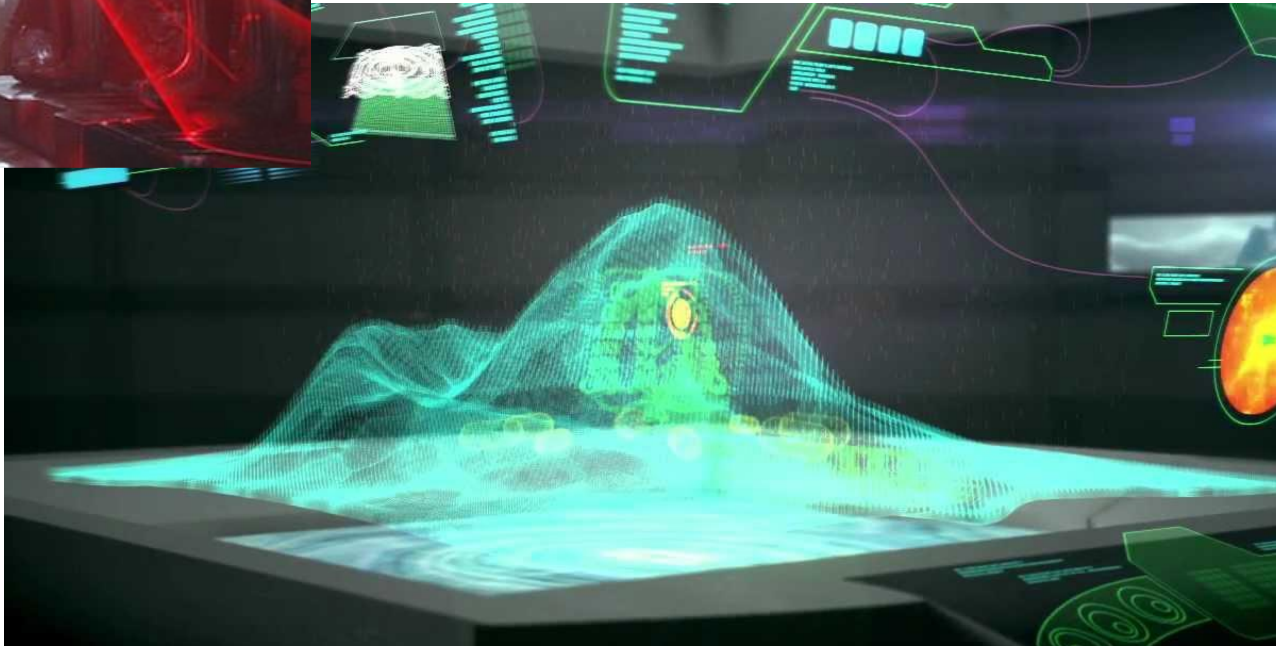


"Dorothy IV"
- *Twister* (1996)

Rumo ao desconhecido ...

Técnicas usadas para explorar o desconhecido?

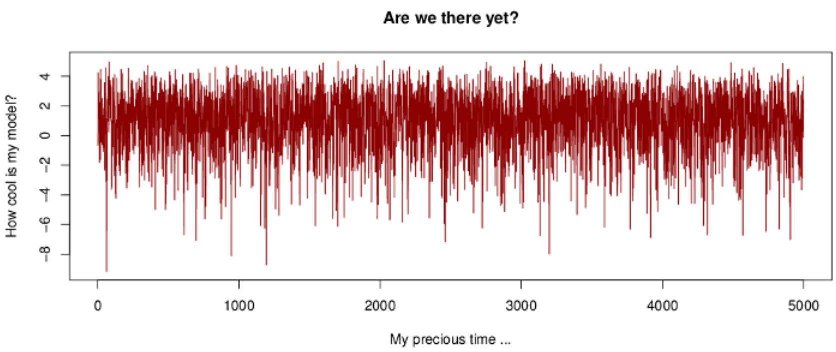
3D holo drones
- *Prometheus* (2012)



Regras de exploração

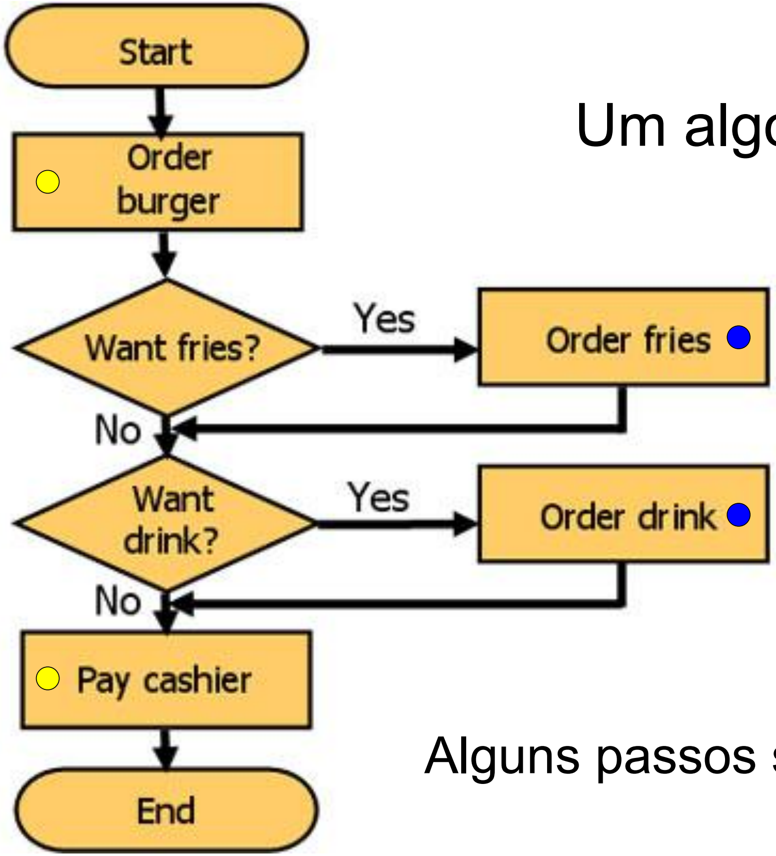
Características do **MCMC**

Bayesian monster
Type: Construct / Outsider
Alignment: L/N
Quote: "I shall walk..."



Seu mundo é visto pelos olhos da função de likelihood.
Existem lendas sobre esse mundo que podem lhe guiar.
Sua memória é curta, somente lembra o último passo.
Ele sempre escala ao alto, mas as vezes desce alguns passos.

Regras de exploração



Um algoritmo é como um fluxograma

O **MCMC** segue um algoritmo.
Os passos seguem uma ordem definida.

Alguns passos são sempre efetuados --

Outros dependem de certas condições --

Visão geral do algoritmo

Metropolis algorithm

1) Selecione um ponto de partida, θ^0.

2) Repita por *n* gerações:

2.1) Faça uma proposta de um novo valor para θ^*.

2.2) Calcule a *odds ratio*:

$$r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}$$

2.3) Defina:

$$\theta^t = \begin{cases} \theta^* \text{ com probabilidade } min(r,1) \\ \theta^{t-1} \text{ do contrario.} \end{cases}$$

2.4) Salve θ^t em algum lugar.

3) $P(\theta|y) \sim \text{histograma de } \{\theta^{t=1}, \theta^{t=2}, \ldots, \theta^{t=n}\}$

MCMC passo a passo

1) Selecione um ponto de partida, θ^0.

- O ponto de partida deve ter valor de verossimilhança maior que zero.

$$P(\theta^0|y) > 0$$

Valor inicial de verossimilhança com likelihood < 0 é um motivo comum para uma cadeia de MCMC não iniciar!

Lembrando que vamos trabalhar com o log(likelihood).

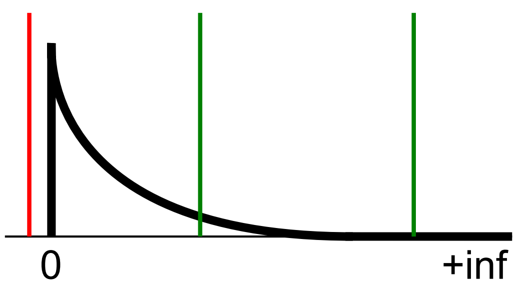
MCMC passo a passo

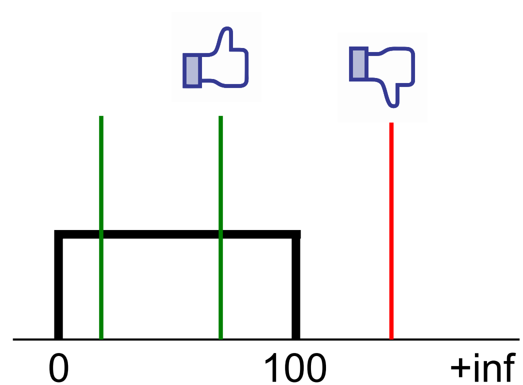
1) Selecione um ponto de partida, θ^0.

- O ponto de partida deve ter valor de verossimilhança maior que zero.

$$P(\theta^0|y) > 0$$

- O ponto de partida também deve ter valor existente de prior, ou seja, estar dentro da distribuição do prior.





MCMC passo a passo

1) Selecione um ponto de partida, θ^0.

- O ponto de partida deve ter valor de verossimilhança maior que zero.

$$P(\theta^0|y) > 0$$

- O ponto de partida também deve ter valor existente de prior, ou seja, estar dentro da distribuição do prior.

- Recomendações:
    1) Rodar cadeias com diferentes θ^0.
    2) Escolher θ^0 com base em uma estimativa rápida, como o ML.
    3) Gerar θ^0 da distribuição do prior.

Visão geral do algoritmo

Metropolis algorithm

1) Selecione um ponto de partida, θ^0.

2) Repita por *n* gerações:

2.1) Faça uma proposta de um novo valor para θ^*.

2.2) Calcule a *odds ratio*:

$$r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}$$

2.3) Defina:

$$\theta^t = \begin{cases} \theta^* \text{ com probabilidade } min(r,1) \\ \theta^{t-1} \text{ do contrario.} \end{cases}$$

2.4) Salve θ^t em algum lugar.

3) $P(\theta|y) \sim \text{histograma de } \{\theta^{t=1}, \theta^{t=2}, \ldots, \theta^{t=n}\}$

MCMC passo a passo

O ponto de partida (θ^0) é o início de nossa cadeia.

Como o algoritmo caminha pela superfície de likelihood?

$\theta^0 \longrightarrow \theta^*$

$$10 + unif(-1, +1) = 9.8$$

$\theta^0 \longrightarrow \theta^*$

MCMC passo a passo

O ponto de partida (θ^0) é o início de nossa cadeia.

Como o algoritmo caminha pela superfície de likelihood?

$$\theta^0 \longrightarrow \theta^*$$

Distribuição de proposta
(*proposal distribution*)

$$10 + unif(-1,+1) = 9.8$$

$$\theta^0 \longrightarrow \theta^*$$

MCMC passo a passo

A distribuição de proposta gera uma quantidade que é adicionada ao valor atual do parâmetro θ^t , formando o θ^*, ou *valor proposta*.

$$\theta^0 \longrightarrow \theta^*$$

Distribuição de proposta
(*proposal distribution*)

$$10 + unif(-1, +1) = 9.8$$

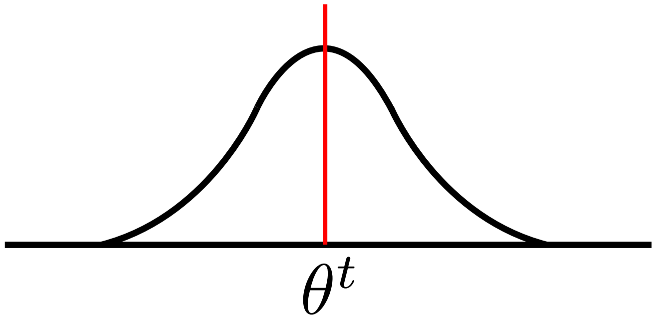
$$\theta^0 \longrightarrow \theta^*$$

MCMC passo a passo

A distribuição de proposta gera uma quantidade que é adicionada ao valor atual do parâmetro θ^t, formando o θ^*, ou *valor proposta*.

$$\theta^t + unif(min, max) = \theta^*$$
$$\theta^t + norm(\mu = \theta^t, \sigma^2) = \theta^*$$

entre outras ...



MCMC passo a passo

A distribuição de proposta gera uma quantidade que é adicionada ao valor atual do parâmetro θ^t , formando o θ^*, ou *valor proposta*.

Importante notar que o novo valor de parâmetro deve pertencer ao conjunto de valores possíveis para o parâmetro.

Por exemplo, $\sigma^2 = [0, \infty)$

Nesses casos temos que usar estratégias para propor novos valores dentro dos valores possíveis.

Visão geral do algoritmo

Metropolis algorithm

1) Selecione um ponto de partida, θ^0.

2) Repita por *n* gerações:

2.1) Faça uma proposta de um novo valor para θ^*.

2.2) Calcule a *odds ratio*:

$$r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}$$

2.3) Defina:

$$\theta^t = \begin{cases} \theta^* \text{ com probabilidade } min(r, 1) \\ \theta^{t-1} \text{ do contrario.} \end{cases}$$

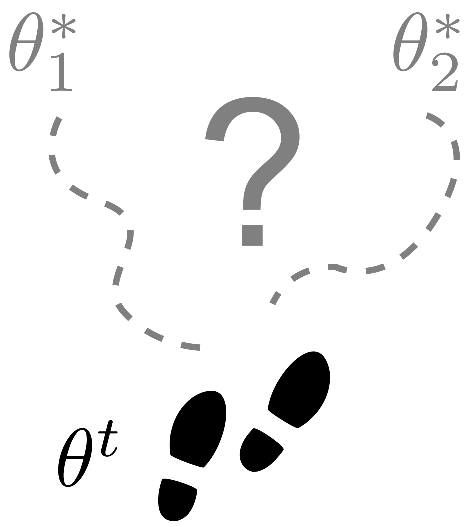
2.4) Salve θ^t em algum lugar.

3) $P(\theta|y) \sim \text{histograma de } \{\theta^{t=1}, \theta^{t=2}, \ldots, \theta^{t=n}\}$

MCMC passo a passo

Agora temos o valor de parâmetro θ^t e a proposta θ^*.

Precisamos decidir se aceitamos θ^* e caminhamos um passo na cadeia ou se rejeitamos θ^* e ficamos no mesmo lugar.

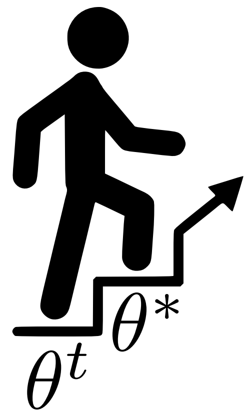


Se o valor proposta tiver melhor fit que o valor atual, então aceitamos θ^*.

Se o valor proposta tiver pior fit que o valor atual, então rejeitamos θ^* com uma certa probabilidade.

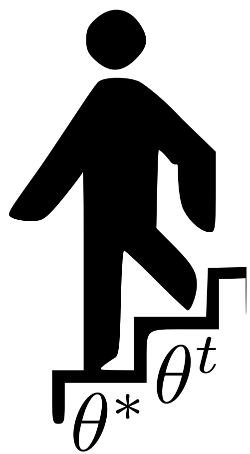
MCMC passo a passo

Se o valor proposta tiver melhor fit que o valor atual, então aceitamos θ^*.



Sempre aceito!

*Aceito com chance igual a **r** (odds ratio)*

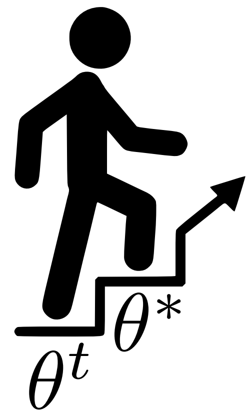


Se o valor proposta tiver pior fit que o valor atual, então rejeitamos θ^* com uma certa probabilidade.

MCMC passo a passo

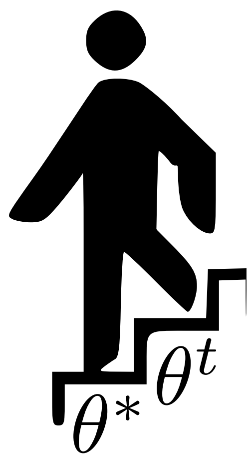
$$r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}$$

$$r \geq 1$$



Sempre aceito!

*Aceito com chance igual a **r** (odds ratio)*

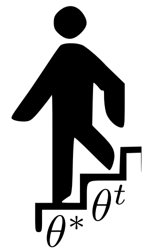


$$0 < r < 1$$
$$P(aceite) \sim r$$

MCMC passo a passo

$$r = \frac{P(\theta^* | y)}{P(\theta^{t-1} | y)}$$

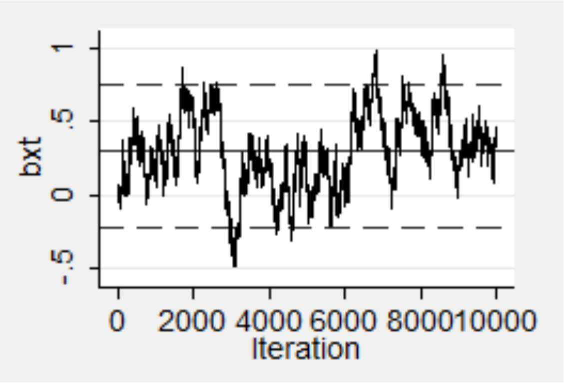
*Aceito com chance igual a **r** (odds ratio)*



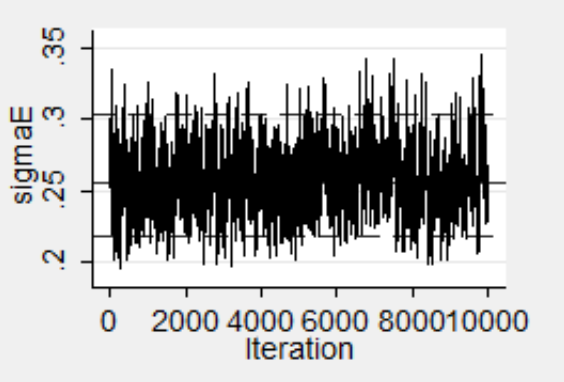
$$0 < r < 1$$
$$P(aceite) \sim r$$

Aceitar um passo para "baixo" é uma característica fundamental do MCMC. Isso possibilita que o algoritmo faça amostragens da distribuição posterior mesmo quando existem multiplos ótimos locais.

MCMC passo a passo



poor mixing



good mixing

Se nenhuma proposta for aceita a cadeia pode ficar no mesmo lugar por várias gerações. Pois os valores de proposta são muito piores que o valor atual.

Isso é chamado de *poor mixing* da cadeia.

Visão geral do algoritmo

Metropolis algorithm

1) Selecione um ponto de partida, θ^0.

2) Repita por *n* gerações:

2.1) Faça uma proposta de um novo valor para θ^*.

2.2) Calcule a *odds ratio*:

$$r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}$$

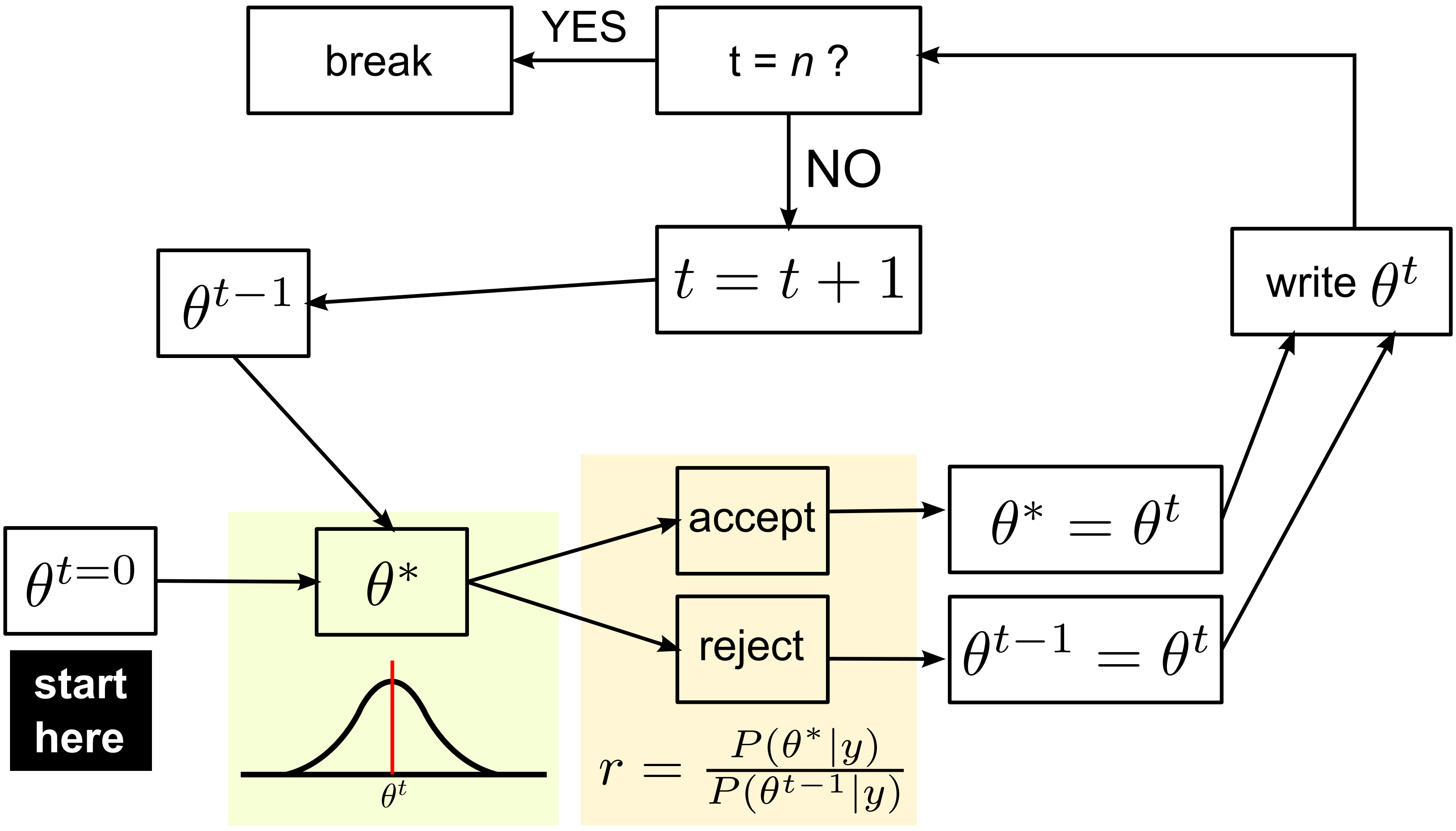
2.3) Defina:

$$\theta^t = \begin{cases} \theta^* \text{ com probabilidade } min(r,1) \\ \theta^{t-1} \text{ do contrario.} \end{cases}$$

2.4) Salve θ^t em algum lugar.

3) $P(\theta|y) \sim \text{histograma de } \{\theta^{t=1}, \theta^{t=2}, \ldots, \theta^{t=n}\}$

MCMC passo a passo



Visão geral do algoritmo

Metropolis algorithm

1) Selecione um ponto de partida, θ^0.

2) Repita por *n* gerações:

  2.1) Faça uma proposta de um novo valor para θ^*.

  2.2) Calcule a *odds ratio*:

$$r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}$$

  2.3) Defina:

$$\theta^t = \begin{cases} \theta^* \text{ com probabilidade } min(r, 1) \\ \theta^{t-1} \text{ do contrario.} \end{cases}$$

  2.4) Salve θ^t em algum lugar.

3) $P(\theta|y) \sim \text{histograma de } \{\theta^{t=1}, \theta^{t=2}, \ldots, \theta^{t=n}\}$

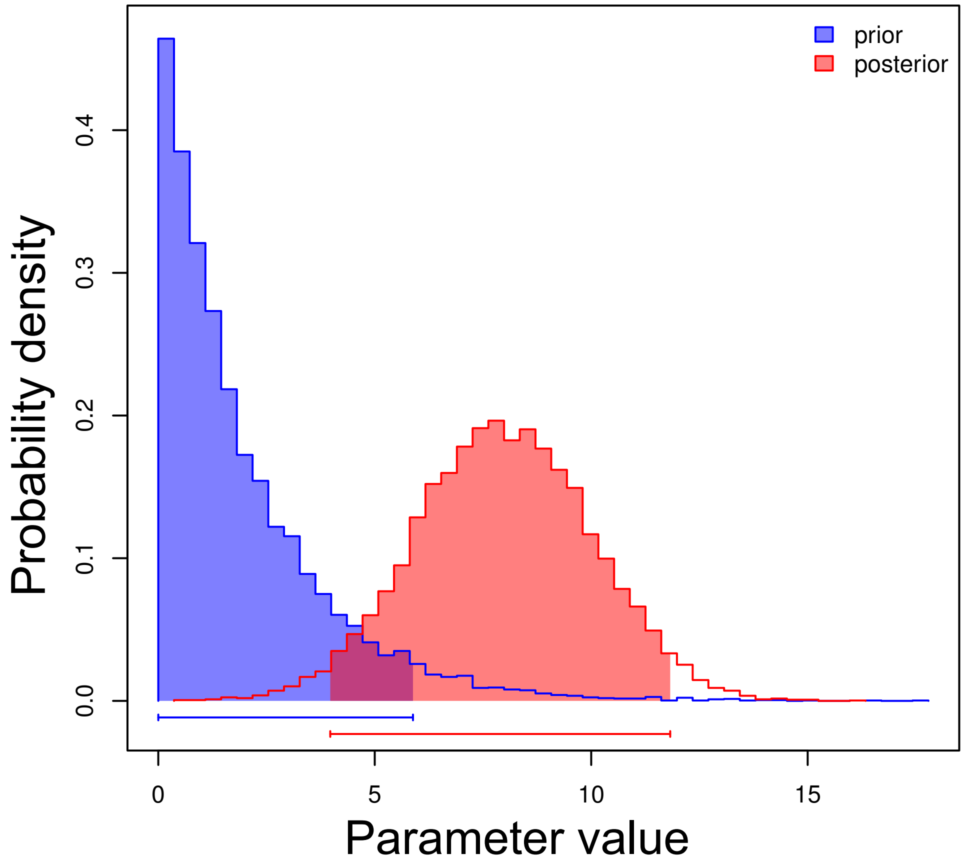
MCMC passo a passo

$\boxed{\text{write } \theta^t}$ vai gerar $\{\theta^{t=0}, \theta^{t=1}, \ldots, \theta^{t=n}\}$

Após retirar o "burnin" (opcional), a série de valores geradas pelo MCMC são as amostras proporcionais à distribuição posterior do modelo.

Usamos essas amostras para explorar a frequência de cada valor de θ_i na posterior.

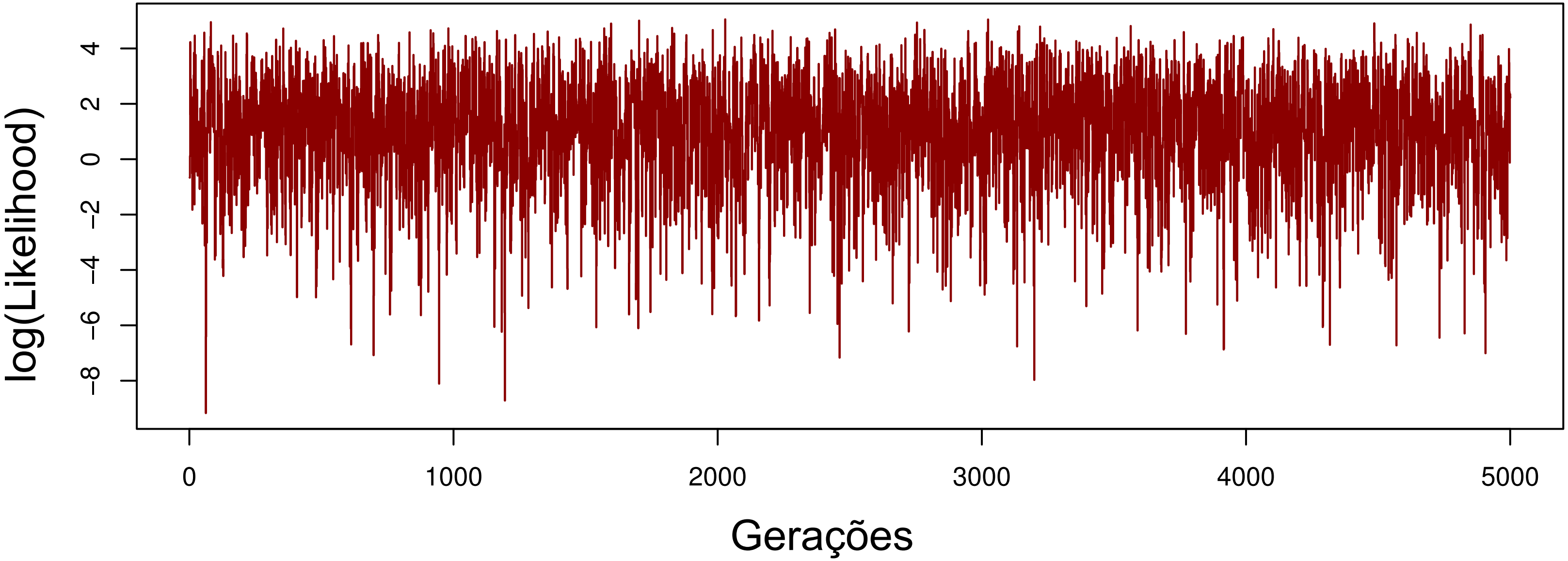
MCMC passo a passo



$$P(\theta|y)\ \&\ P(\theta|y) \sim hist(\theta^{t=0}, \theta^{t=1}, \ldots, \theta^{t=n})$$

MCMC passo a passo

O gráfico da verossimilhança em função das gerações também é bastante explorado. Aqui podemos verificar se a cadeia tem um bom *mixing*.



MCMC passo a passo

Agora vamos trabalhar esses conceitos com exemplos.

Vamos também explorar como o algoritmo de Metropolis é usado para estimar parâmetros de modelos comparativos filogenéticos.

Próximo passos: **Metropolis-Hastings algorithm**
Exemplos com PCMs
Análise da posterior estimada
(burnin, convergência e etc...)