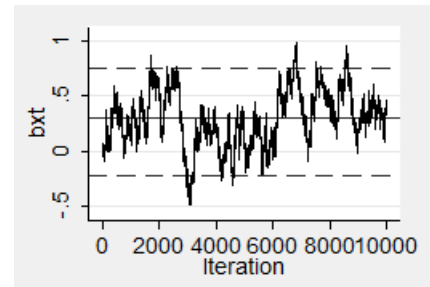
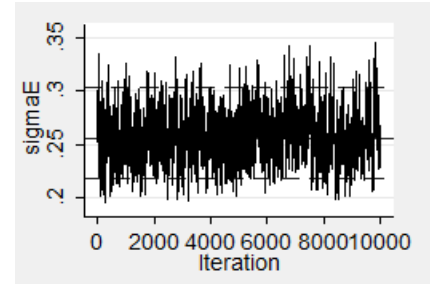
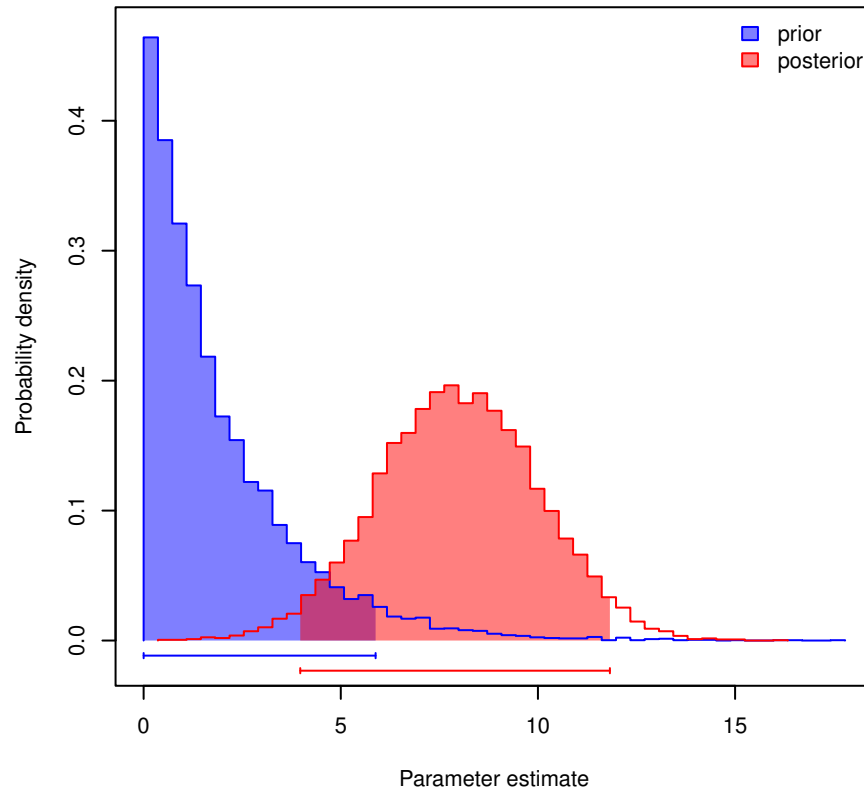


Análise da posterior

Se o MCMC rodou conforme esperado temos uma cadeia de valores, nossa distribuição posterior.



Análise da posterior

Infelizmente nossa análise não está completa. A teoria do MCMC garante que o método atinge a distribuição posterior correta no limite $n \rightarrow \infty$.

No entanto, rodamos o MCMC por alguns milhões de gerações, o que está BEM longe de $n \rightarrow \infty$.

Por isso precisamos verificar a **convergência** das cadeias.

Análise da posterior

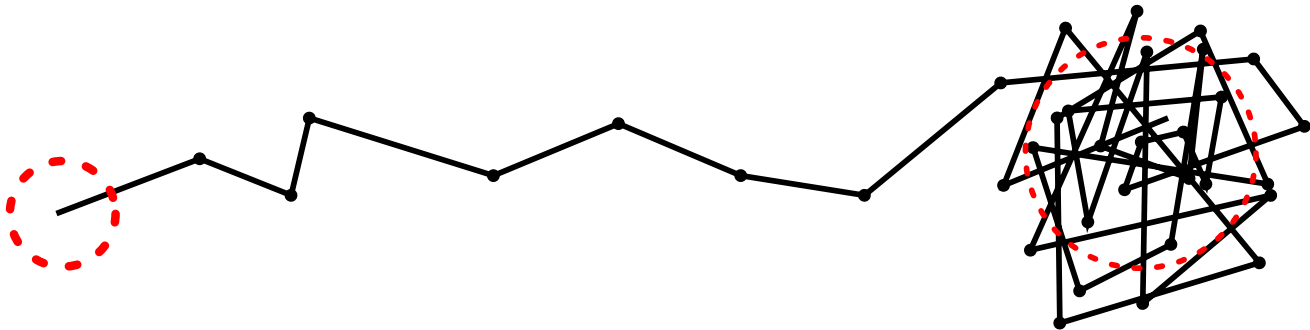
Vamos abordar as análises utilizadas para avaliar os resultados do **MCMC**.

- *Burn-in* ou queima da cadeia (?).
- *Thinning* ou rarefação da cadeia.
- Análise de convergência.
- *Posterior predictive simulations* ou Simulações preditivas da posterior.

Retirada do *burn-in*

A cadeia do MCMC pode ser influenciada mais pelo valor de início do que pela distribuição posterior alvo durante as primeiras gerações.

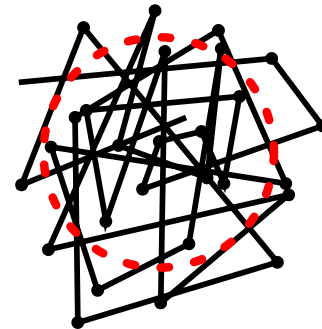
Nesse caso, a frequência com que valores são amostrados na fase de *burn-in* é diferente do que a frequência alvo.



Retirada do *burn-in*



target distribution
depois da retirada
do burn-in.



Retirada do *burn-in*

Quanto é necessário retirar de *burn-in*?



Retirada do *burn-in*

Quanto é necessário retirar de *burn-in*?

Entre 30% e 50% das
gerações da cadeia.



Retirada do *burn-in*

Quanto é necessário retirar de *burn-in*?

Propriedade de Markov
garante que o estado
inicial seja 'esquecido'.



Prática

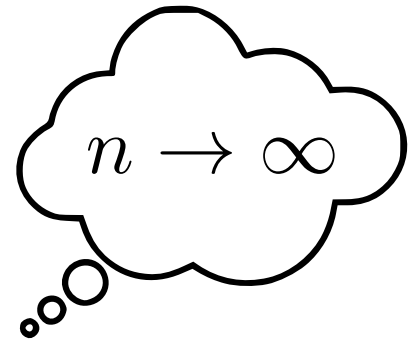


Teoria

Retirada do *burn-in*

Quanto é necessário retirar de *burn-in*?

Propriedade de Markov
garante que o estado
inicial seja 'esquecido'.



Prática



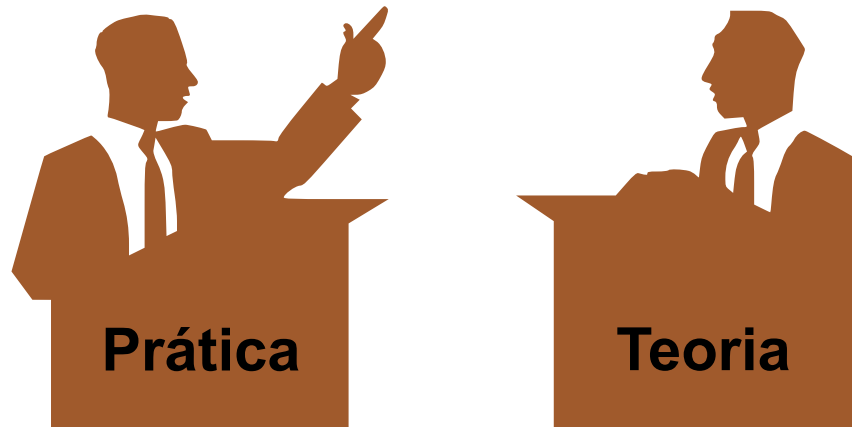
Teoria

Retirada do *burn-in*

Quanto é necessário retirar de *burn-in*?

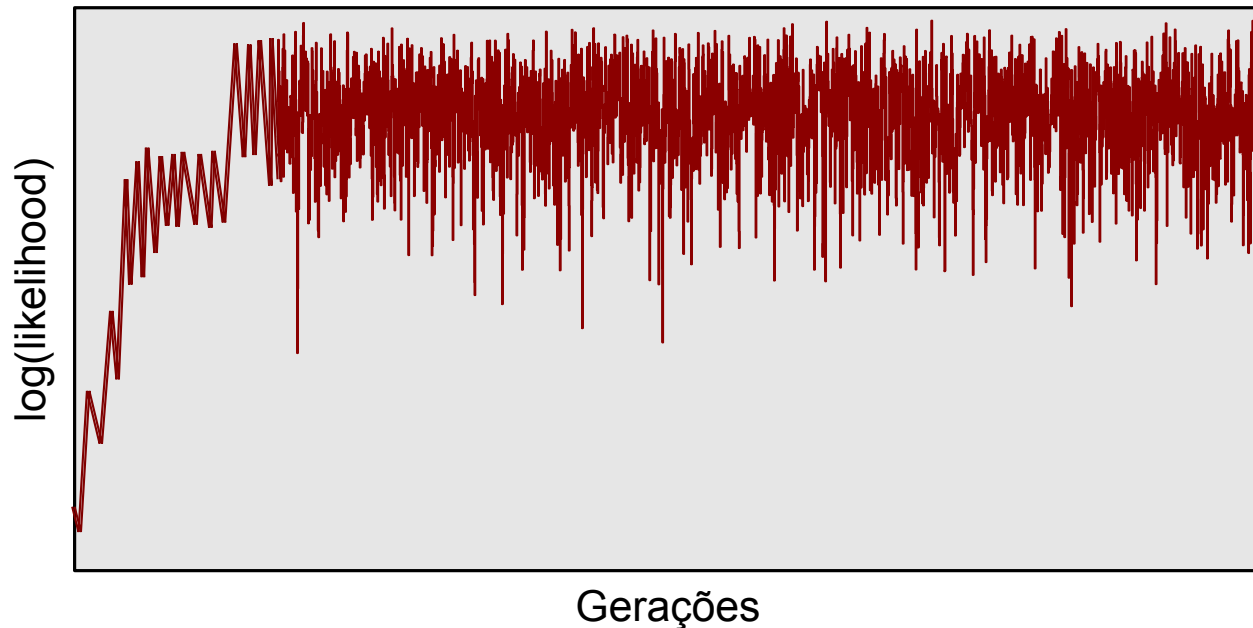
A grande maioria das análises retira uma porção da cadeia como *burn-in*. A proporção varia.

Não retirar a fase de *burn-in* quando a análise tem o ponto inicial com base em uma estimativa prévia, como o MLE, faz sentido.



Retirada do *burn-in*

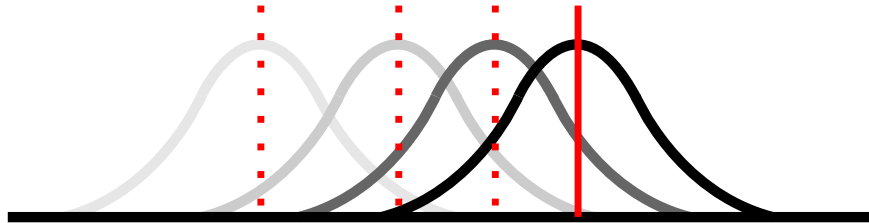
Uma prática comum é verificar a proporção do burn-in usando o gráfico do $\log(\text{likelihood})$ pelo número de gerações.



Thinning da cadeia

O *thinning* ou rarefação da cadeia é aplicado com o objetivo de diminuir o efeito da autocorrelação entre as gerações do MCMC.

Lembre-se que os passos do MCMC seguem uma distribuição de proposta:

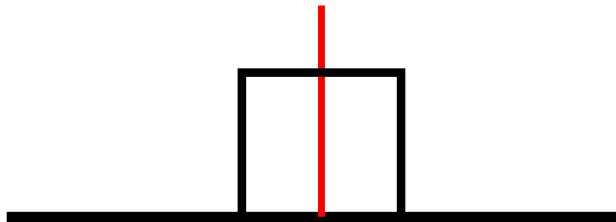


Analizando a série temporal do MCMC podemos observar que as gerações geralmente não são independentes.

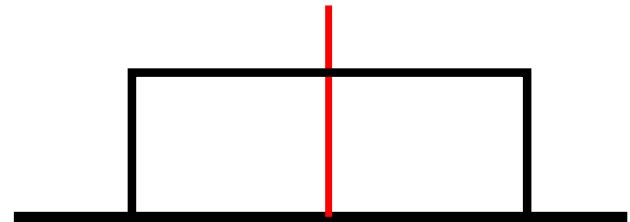
Thinning da cadeia

A autocorrelação entre as gerações é influenciada pela largura dos passos da cadeia.

Quando a distribuição de proposta é estreita (passos curtos) a autocorrelação é maior.



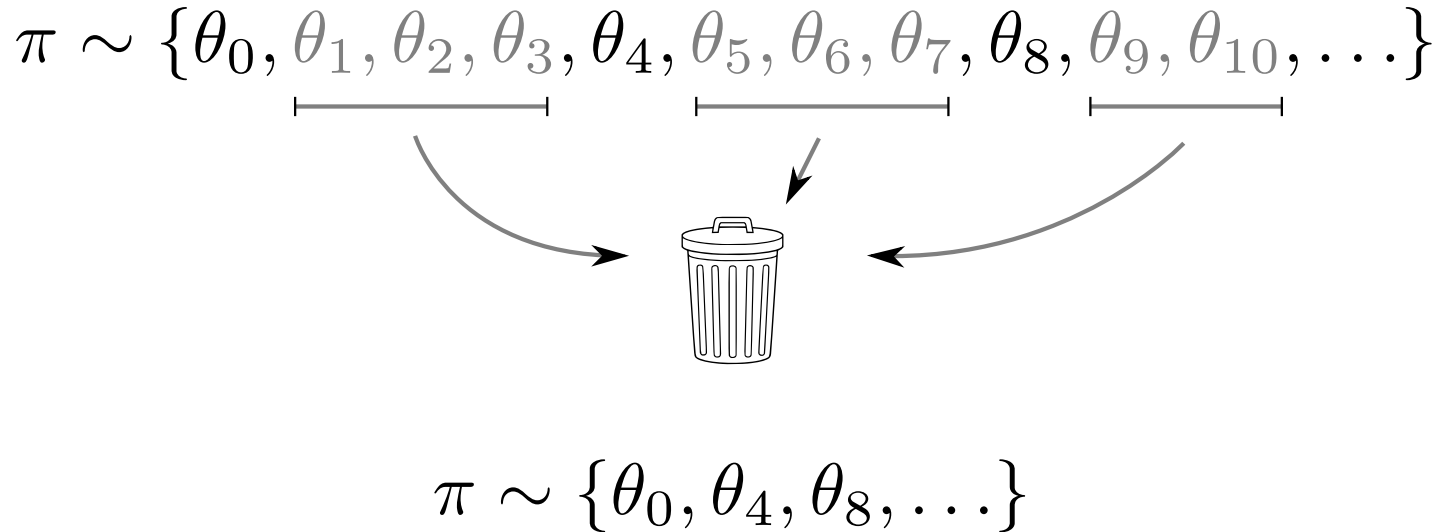
maior autocorrelação



menor autocorrelação

Thinning da cadeia

Para aplicar o *thinning*, somente cada k geração é mantida enquanto todo o resto é descartado.



Thinning da cadeia

Methods in Ecology and Evolution

FORUM

On thinning of chains in MCMC

William A. Link and Mitchell J. Eaton

Article first published online: 17 JUN 2011

DOI: 10.1111/j.2041-210X.2011.00131.x

© 2011 The Authors. Methods in Ecology and Evolution © 2011 British Ecological Society

Issue



Methods in Ecology and Evolution

Volume 3, Issue 1, pages
112–115, February 2012



Thinning da cadeia

Embora a aplicação de *thinning* seja comum, a principal justificativa é a economia de memória computacional.

Melhor que aplicar *thinning* seria melhorar a estratégia de proposta para diminuir a autocorrelação sem "jogar resultados fora".

Resultados mostram que a precisão de estimativas baseadas em cadeias sem rarefação é mais alta.

Link and Eaton, *Methods in Ecology and Evolution*, 2012(3)112-115

Análise de convergência

A cadeia de MCMC tem por objetivo fazer amostras da distribuição posterior do modelo por meio de simulações.

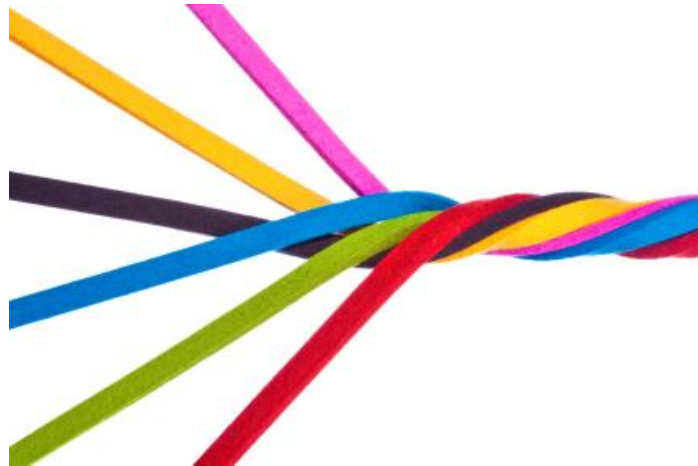
Quando o MCMC atinge este objetivo, dizemos que a cadeia atingiu a **convergência**.

$$MCMC \rightarrow P(\theta|y)$$

Análise de convergência

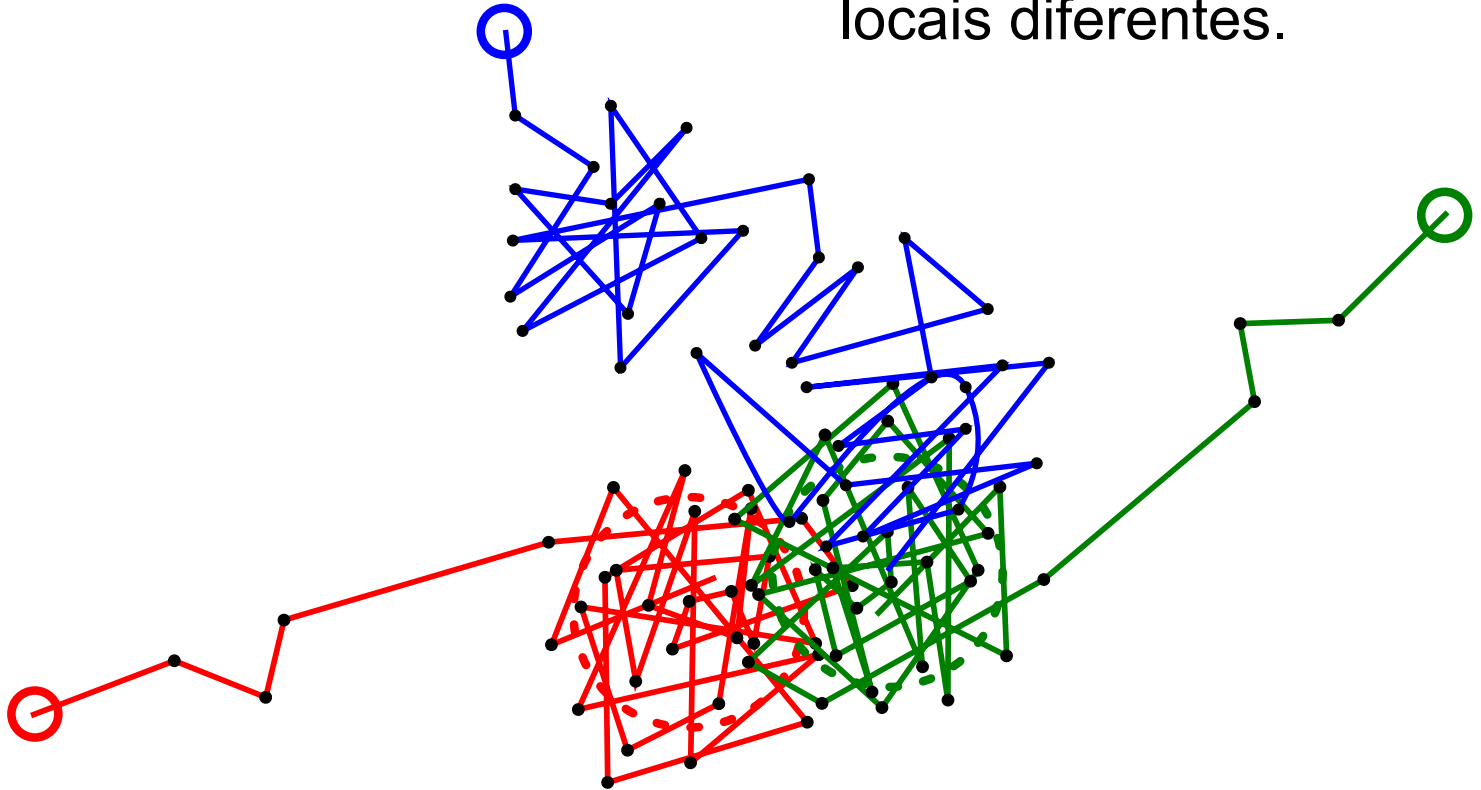
Analisar a convergência é importante para certificar que as cadeias não estão amostrando diferentes ótimos locais na superfície de verossimilhança.

A idéia central da análise de convergência é que duas ou mais cadeias chegaram na mesma região vindo de pontos de partida diferentes.



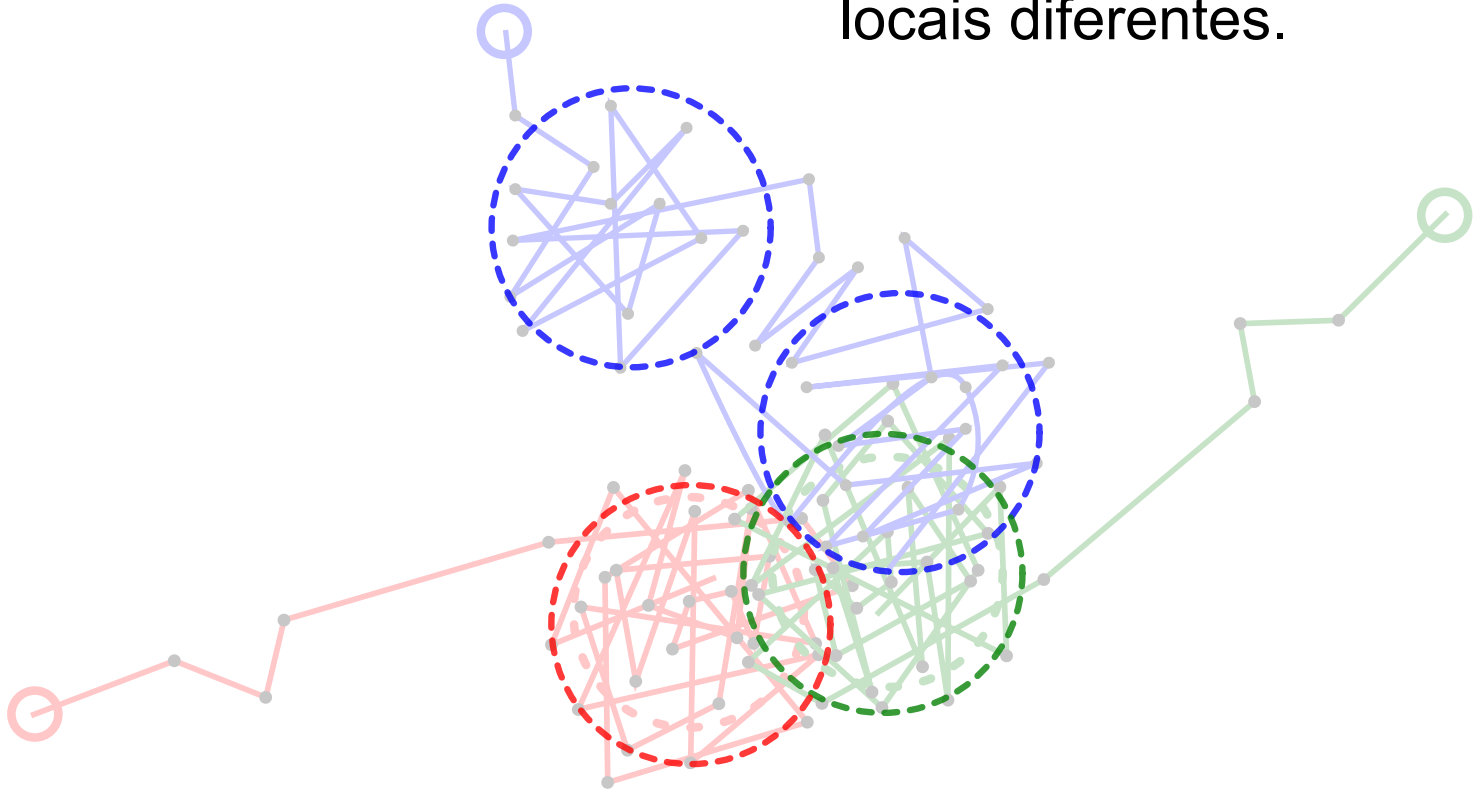
Análise de convergência

Três cadeias atingem ótimos locais diferentes.



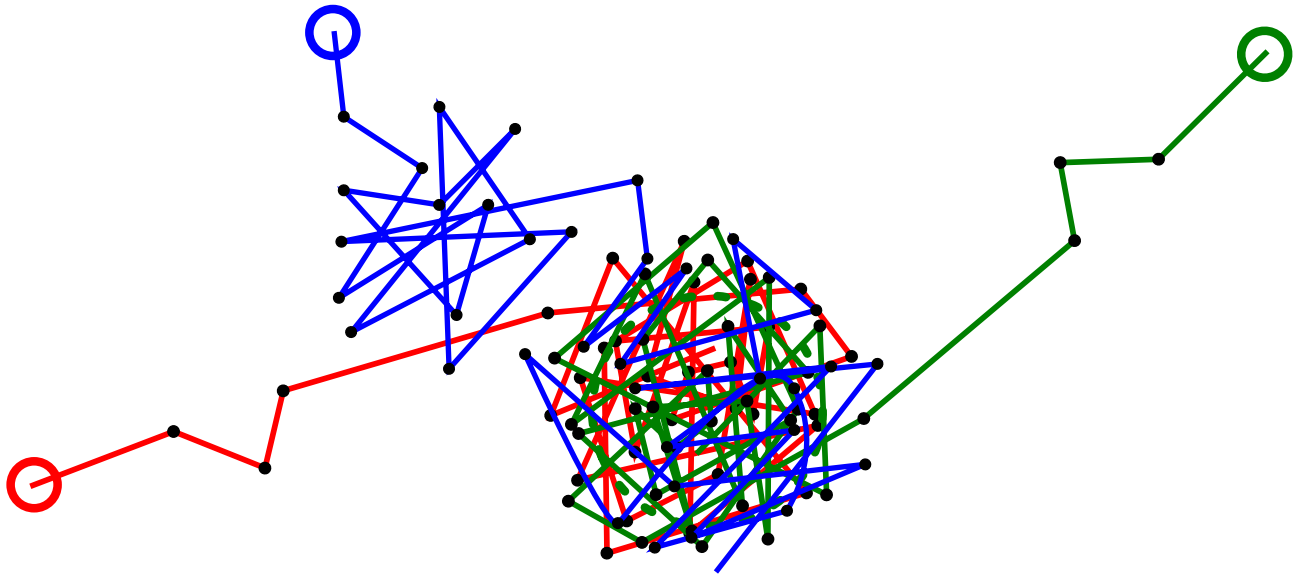
Análise de convergência

Três cadeias atingem ótimos locais diferentes.



Análise de convergência

Três cadeias atingem o mesmo
ótimo global.

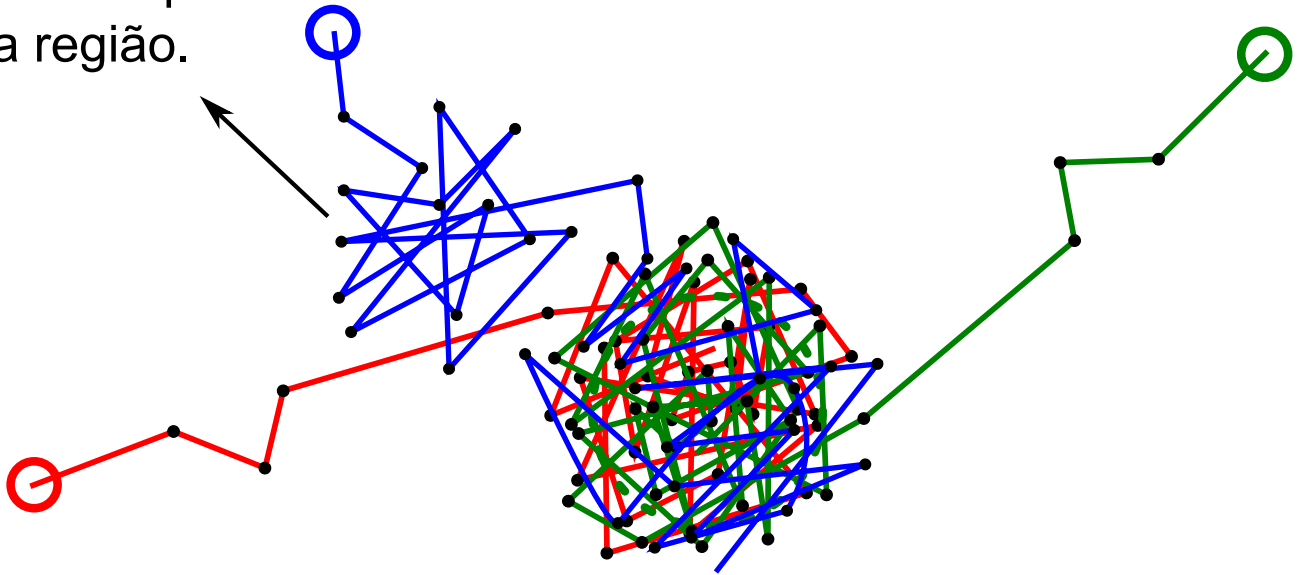


Análise de convergência

Cadeia **azul** fica 'presa' em um outro pico.

Sugere melhor exploração com valor início perto dessa região.

Três cadeias atingem o mesmo ótimo global.



Análise de convergência

Um dos melhores testes de convergência (talvez o melhor!) é o Gelman and Rubin (1992) **R** ou 'potential reduction factor'.

Essa estatística se baseia em uma média ponderada da variância dentro de cada cadeia e entre as cadeias. Quando **R** se aproxima de **1** (limite máximo) a convergência é aceita.

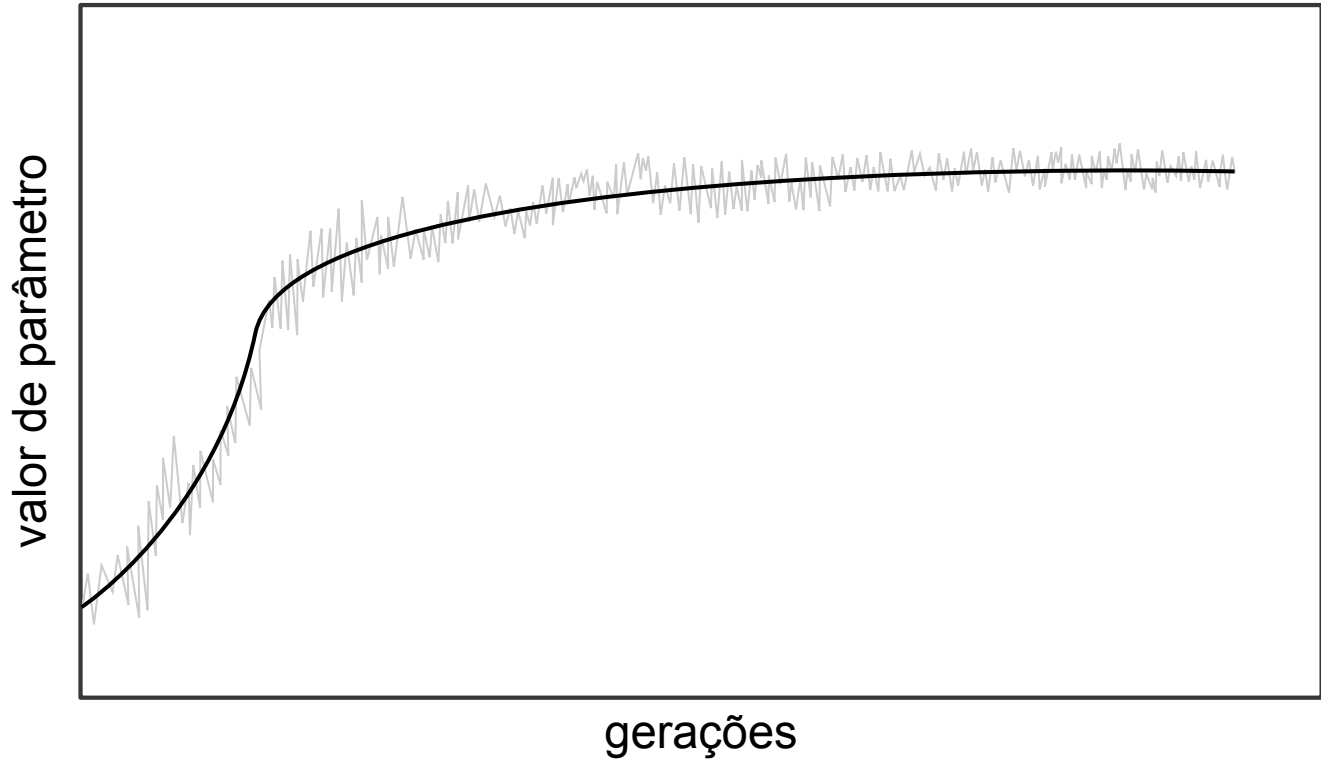
A vantagem desta estatística é que ela estima uma quantidade que estamos diretamente interessados, se as cadeias "contam a mesma história".

Análise de convergência

Infelizmente as vezes não é possível rodar uma série de cadeias com pontos iniciais distintos.

Existem maneiras de acessar se a cadeia chegou em convergência com base em somente uma análise, no entanto estes métodos são menos confiáveis.

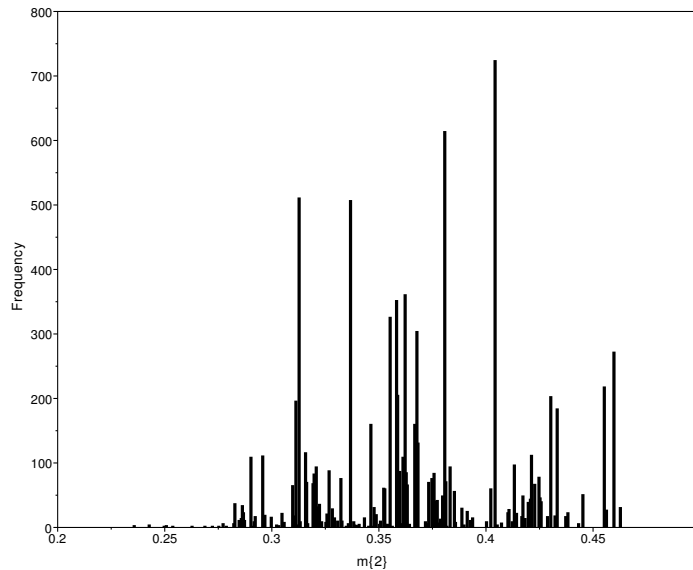
Análise de convergência



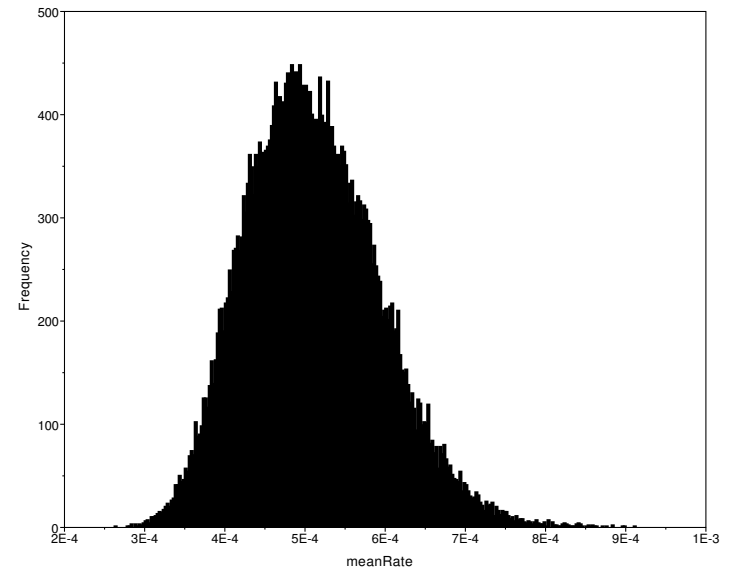
Valor do parâmetro estabiliza com o número de gerações.

Análise de convergência

Histograma da posterior dos parâmetros.



baixo *mixing*
baixo **ESS**



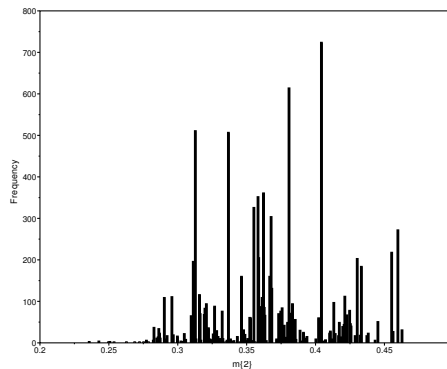
bom *mixing*
boa **ESS**

Análise de convergência

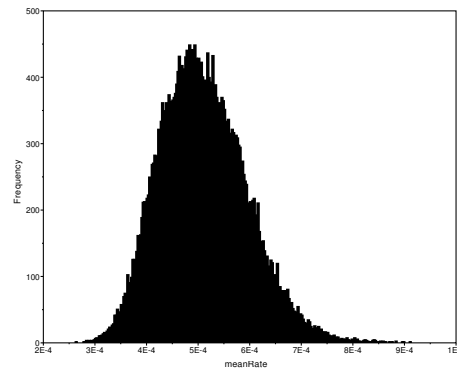
Effective Sample Size (ESS)

O 'tamanho efetivo da amostra' estima o número de amostras independentes resultantes do MCMC.

O ESS é o número "real" de amostras feitas pelo MCMC.



baixo mixing
baixo ESS



bom mixing
boa ESS

Análise de convergência

O **ESS** é o número de amostras que a estimativa do parâmetro está se baseando (assumindo convergência).

Quando a autocorrelação é grande, precisamos rodar o MCMC por um número maior de gerações para conseguir amostras independentes para estimar os parâmetros.

Não existe um número mágico para a **ESS**. Quantas amostras parece razoável para estimar seus parâmetros?

Análise de convergência

Em suma,

Esteja **ALERTA**.

