Universidade Federal do Rio Grande do Norte
Instituto Metrópole Digital
IMD0601 - Bioestatística

Monte Carlo Markov Chain

Prof. Dr. Tetsu Sakamoto
Instituto Metrópole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br

Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

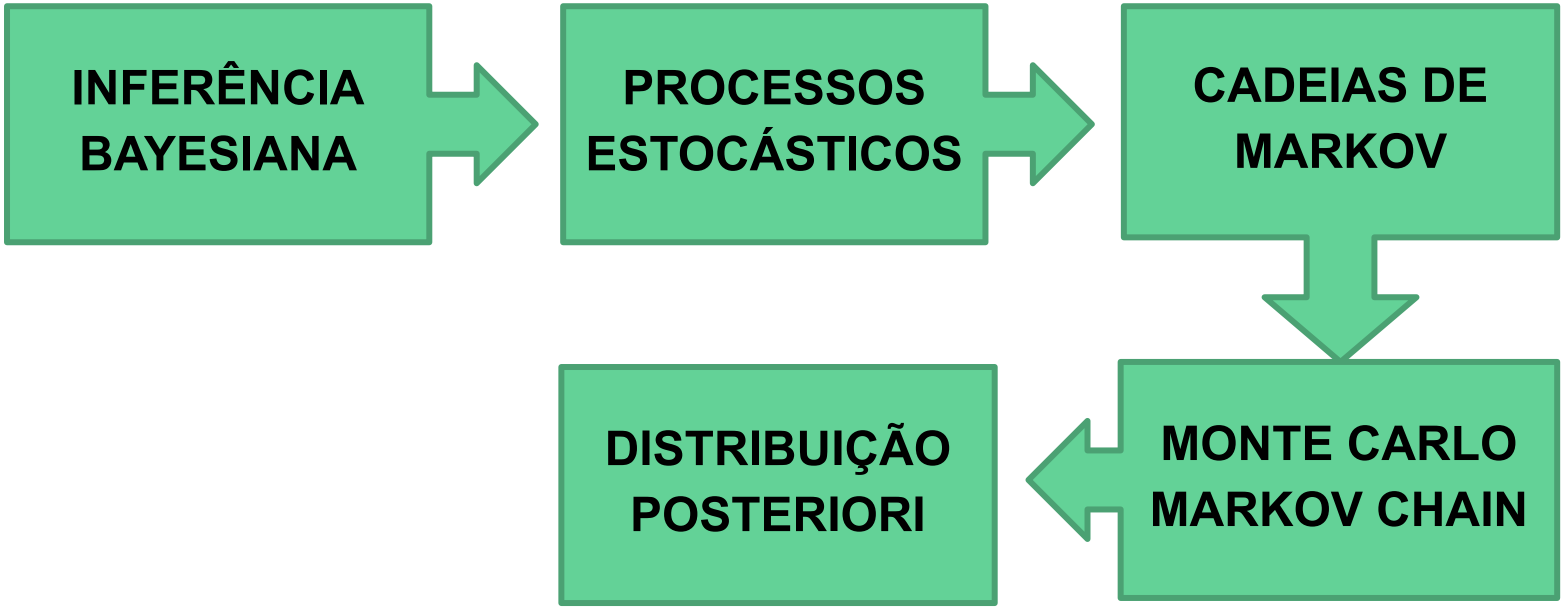
```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601

> git pull
```

Road map


INFERÊNCIA BAYESIANA
PROCESSOS ESTOCÁSTICOS
CADEIAS DE MARKOV
MONTE CARLO MARKOV CHAIN
DISTRIBUIÇÃO POSTERIORI

Revisão

Cadeias de Markov

Distribuição estacionária → Distribuição posteriori

Cadeia de Markov deve ser:

- Finita → espaço amostral;
- Ergódica → Aperiódico e Irredutível;

Outras aplicações da cadeia de Markov:

- Padrões no DNA;
- Gerador de sequências aleatórias;

Objetivo

- Metodologia bayesiana + Cadeias de Markov;
- Simular e estudar as distribuições posteriori;
- MCMC;

Simulações Monte Carlo

Monte Carlo Casino (Mônaco)

"Qualquer técnica que obtém soluções de problemas utilizando números aleatórios."

Aumento explosivo com o advento dos computadores;

Métodos de simulação de distribuição

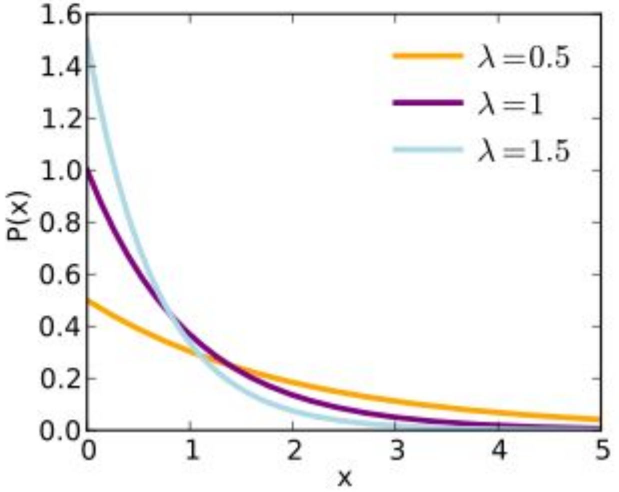
- Função Distribuição Acumulada Inversa;
- Amostragem por rejeição;

Função Distribuição Acumulada Inversa;

- Método mais simples de simular distribuição de probabilidade;
- Utiliza geradores de números randômicos ➔ gera sequências de números aleatórios de 0 a 9;
  - Gera qualquer número real entre 0 e 1 (até uma certa precisão);
  - Se muitos números aleatórios são gerados ➔ distribuição uniforme ➔ corresponde a função **runif()**;
- **FDA inversa** ➔ transforma um conjunto de números que segue uma distribuição uniforme(0, 1) em uma distribuição desejada.

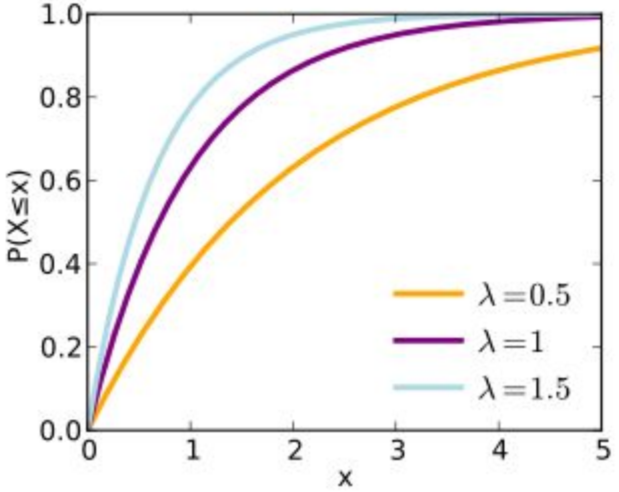
Função Distribuição Acumulada

Distribuição Exponencial



Função densidade de probabilidade

$$P(x) = \lambda \, e^{-\lambda x}$$



Função distribuição acumulada

$$F(x) = 1 - e^{-\lambda x}$$

Função Distribuição Acumulada Inversa

Distribuição exponencial:

$$F(x) = 1 - e^{-\lambda x}$$

$$u = 1 - e^{-\lambda x}$$

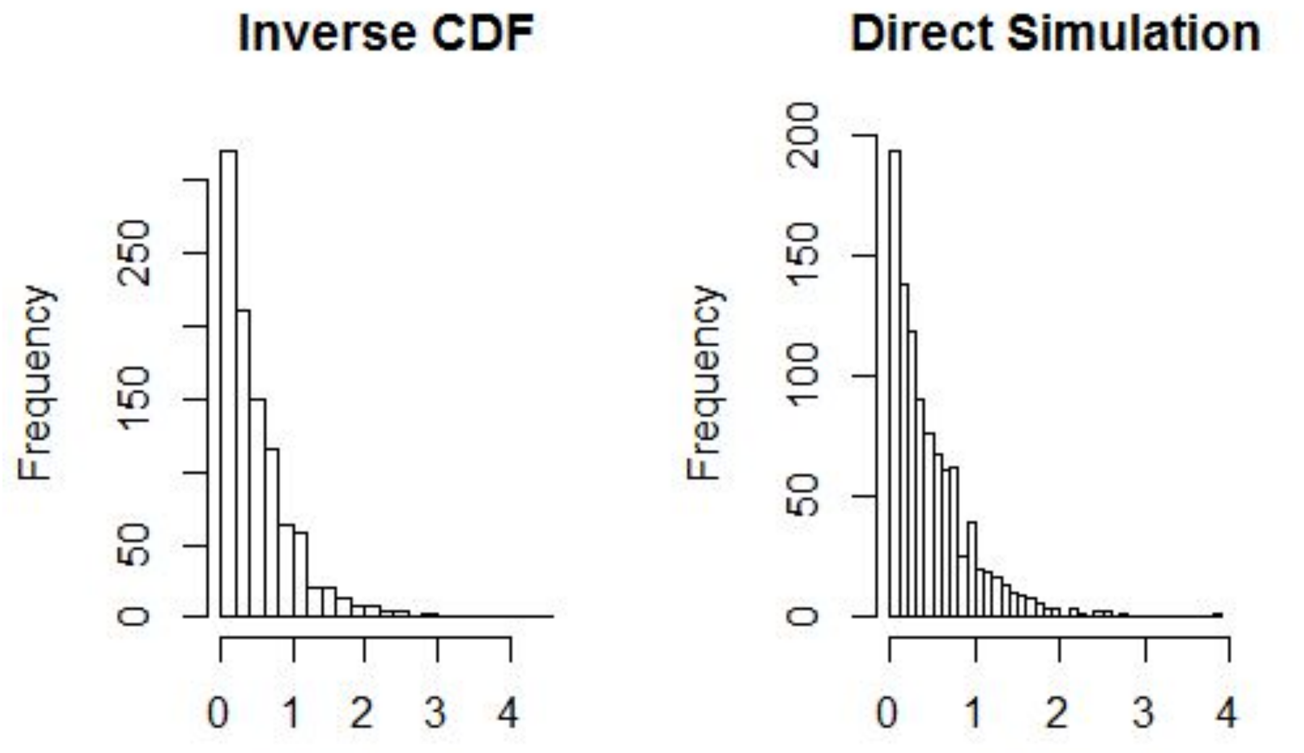
$$1 - u = e^{-\lambda x}$$

$$log(1 - u) = -\lambda x$$

$$x = -\frac{1}{\lambda} log(1 - u)$$

Função Distribuição Acumulada Inversa

Distribuição Exponencial



Amostragem por Rejeição

Método Monte Carlo de simulação de distribuição;

Útil quando o FDA inversa é difícil de ser obtida;

Aproxima os dados que segue uma distribuição para a distribuição desejada;

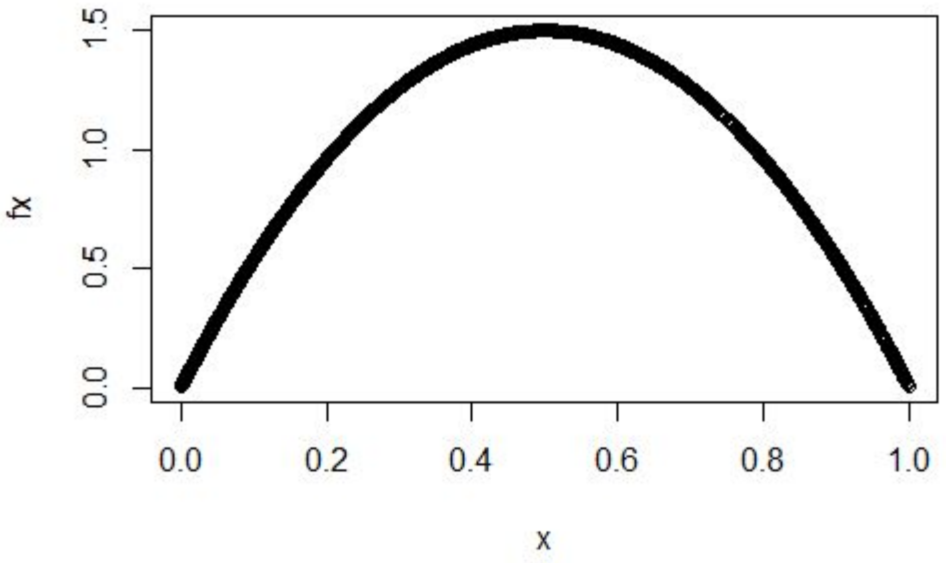
- Distribuição envelope
- Conjunto de regras algorítmicas (aceita ou rejeita a amostra)

Amostragem por Rejeição

Exemplo: simular uma distribuição beta com alfa=2 e beta=2

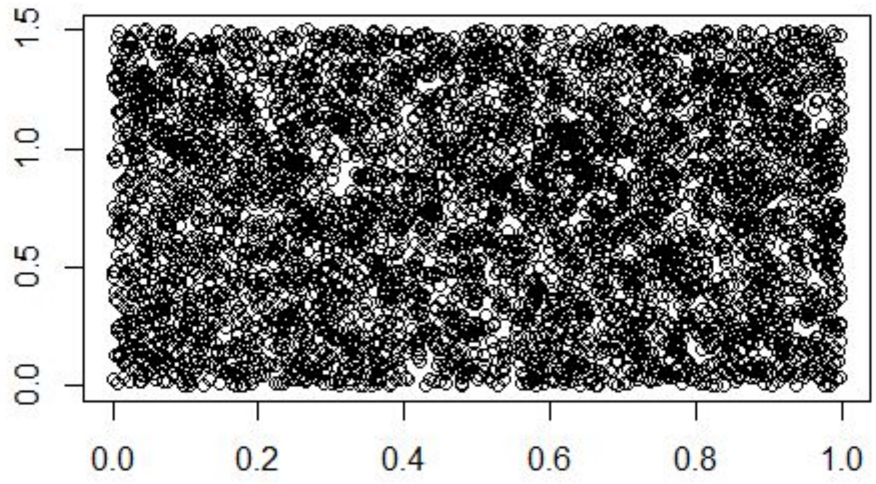
$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$$f(x) = 6x(1-x)$$



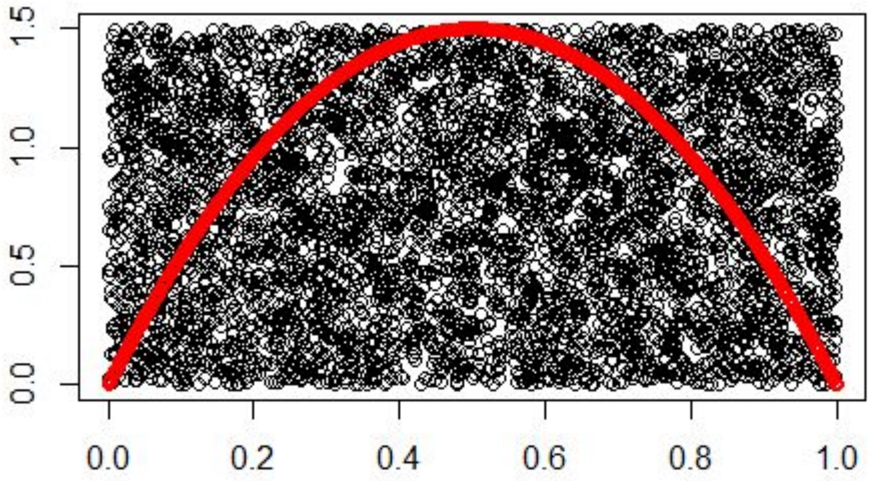
Amostragem por Rejeição

Etapa 1: Gerar dados que seguem uma distribuição



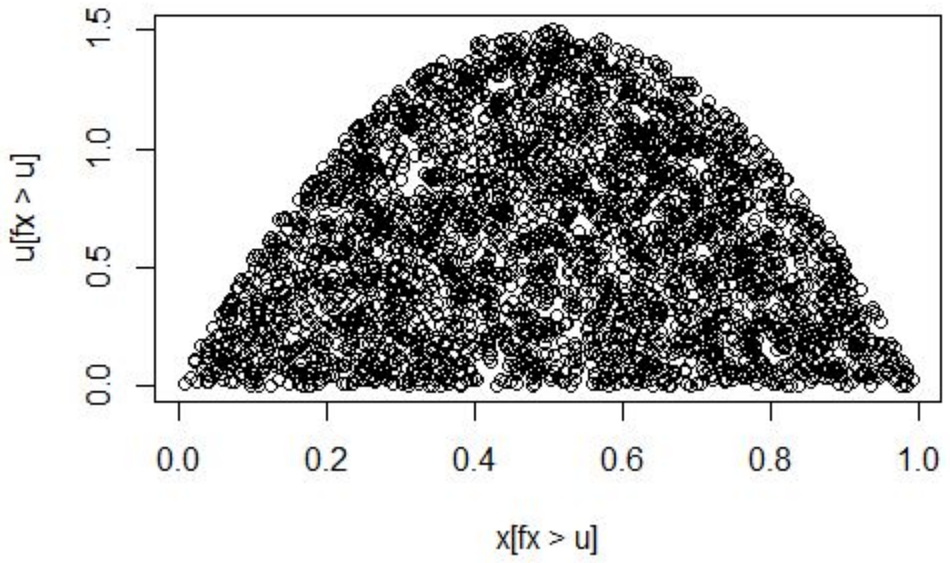
Amostragem por Rejeição

Etapa 2: Criar regras de rejeição



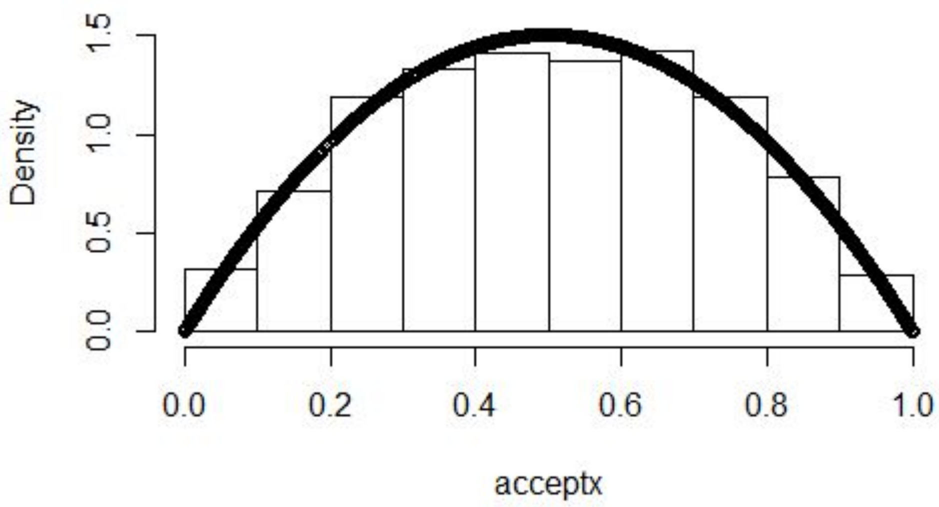
Amostragem por Rejeição

Etapa 3: Escolher os pontos que seguem a regra



Amostragem por Rejeição

Etapa 4: Conferir a distribuição dos dados



Algoritmo de Metropolis

Um método **MCMC**

Distribuição proposta ➔ Uma distribuição utilizada para iniciar o algoritmo. Como se fosse a matriz de transição;

Distribuição alvo ➔ Distribuição posteriori. Matriz estacionária de uma cadeia de Markov;

Ideia principal ➔ Encontrar o topo da distribuição

- Simular valores da distribuição proposta;
- Utilizar critérios que define se aceitamos ou rejeitamos os valores simulados;
    - Baseado na proporção da densidade da distribuição posteriori;

Algoritmo de Metropolis

Etapas do algoritmo:

1. Gerar novos valores da distribuição proposta, denominado theta*;
2. Calcular a proporção R entre os posteriores do novo valor (theta*) e do velho (theta)

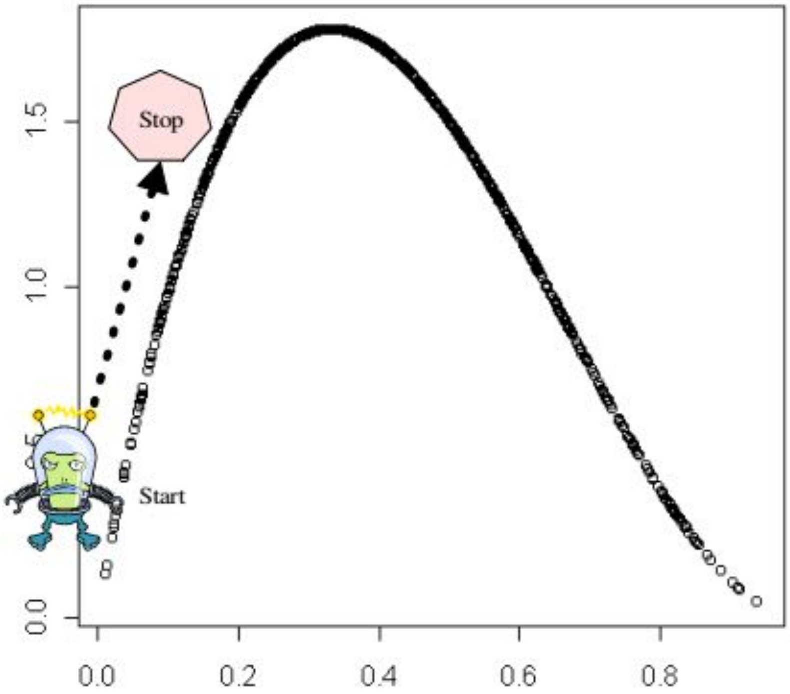
$$R = \frac{p(\theta^* | dados)}{p(\theta | dados)} \qquad R = \frac{p(dados | \theta^*) p(\theta^*)}{p(dados | \theta) p(\theta)}$$

3. Amostrar um valor u de uma distribuição uniforme [0,1]
4. Aceitar o novo valor theta* se u < min(1, R)

Algoritmo de Metropolis

Aceitar theta se u < min(1, R)

$$R = \frac{p(dados|\theta^*)p(\theta^*)}{p(dados|\theta)p(\theta)}$$



Mover para cima é sempre bom → Encontrando áreas de maior densidade;

P antigo → 0.5 P novo → 1.5

R = 3

Algoritmo de Metropolis

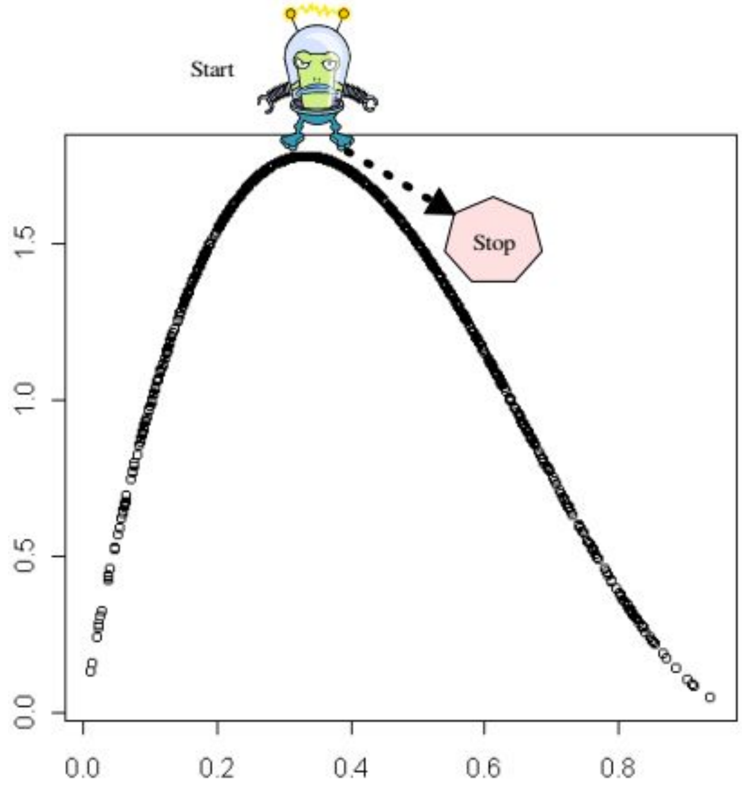
Aceitar theta se u < min(1, R)

$$R = \frac{p(dados|\theta^*)p(\theta^*)}{p(dados|\theta)p(\theta)}$$

Mover um pouco para baixo ➔ Boa chance do valor proposto ser aceitável;

P antigo ➔ 1.7 P novo ➔ 1.4

R = 0.82



Algoritmo de Metropolis

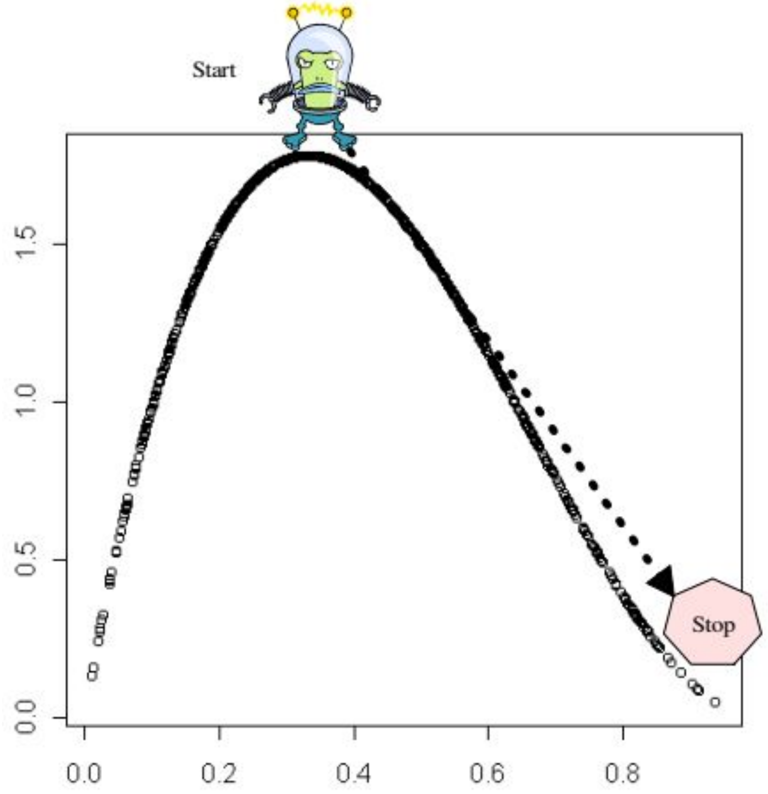
Aceitar theta se u < min(1, R)

$$R = \frac{p(dados|\theta^*)p(\theta^*)}{p(dados|\theta)p(\theta)}$$

Mover muito para baixo ➔ Valor proposto é rejeitado na maioria das vezes;

P antigo ➔ 1.7 P novo ➔ 0.3

R = 0.176



Algoritmo de Metropolis

Exemplo:

Distribuição posteriori de theta;

Theta como probabilidade de sucesso em uma distribuição binomial;

Dados: y = 3 sucessos e n = 5 jogadas;

Distribuição priori ➜ uniforme(0,1)

Distribuição posteriori $f(\theta \mid Y=3) \propto \Pr(Y=3 \mid \theta) f(\theta) = \theta^3 (1-\theta)^2 \cdot 1$

distribuição beta, alfa=1+y = 4 e beta=1+n-y = 3, média = 0.57