Introdução a linguagem Stan (*rstan*), um software para modelos bayesianos.

Universidade Federal do Oeste do Pará (UFOPA)
Campus de Monte Alegre – Engenharia de Aquicultura

Universidade Federal de Lavras (UFLA)
PPG – Estatística e Experimentação Agropecuária

AGRADECIMENTO E COLABORADORES:

UNIVERSIDADE FEDERAL DE LAVRAS

NÚCLEO DE ESTUDOS EM REGRESSÃO NÃO LINEAR APLICADA

UFOPA

Professor: Carlos Antônio Zarzar
E-mail: carloszarzar_@hotmail.com
carlos.zarzar@ufopa.edu.br

Data: 09/03/2022

➢ Sumário

- Estatística Bayesiana;
- Modelos Hierárquicos;
- **Método computacional de integração MCMC;**
- Algoritmo de amostragem;
  - Gibbs;
  - Metropolis-Hasting;
  - Hamiltoniano Monte Carlo;
  - NUTS;

Métodos computacionais de integrações

- Método de Monte Carlo (MC);
- Método de Monte Carlo via Cadeia de Markov (MCMC);
- Aplicação à Estatística Bayesiana.

- Teorema de Bayes:

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta, x)d\theta}$$

carloszarzar_@hotmail.com

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**

- Teorema de Bayes:

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta, x)d\theta}$$

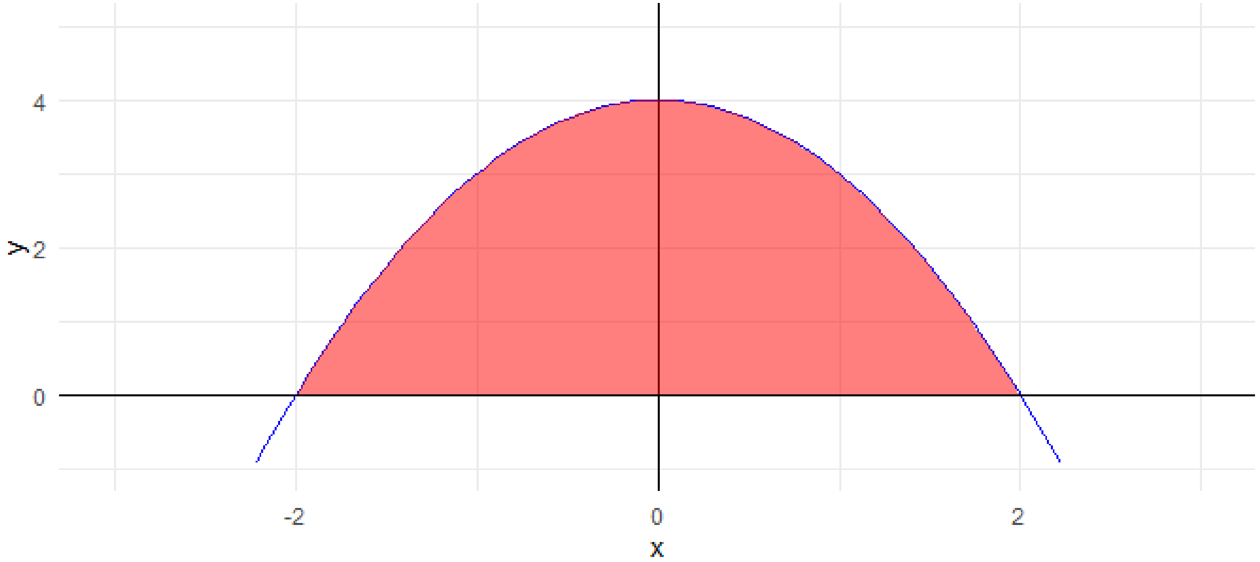
Geração de variáveis aleatórias independentes com mesma distribuição

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



$$\int_0^h f(x)dx = \int_{-2}^{2} -x^2 + 4\, dx$$

Geração de variáveis aleatórias independentes com mesma distribuição

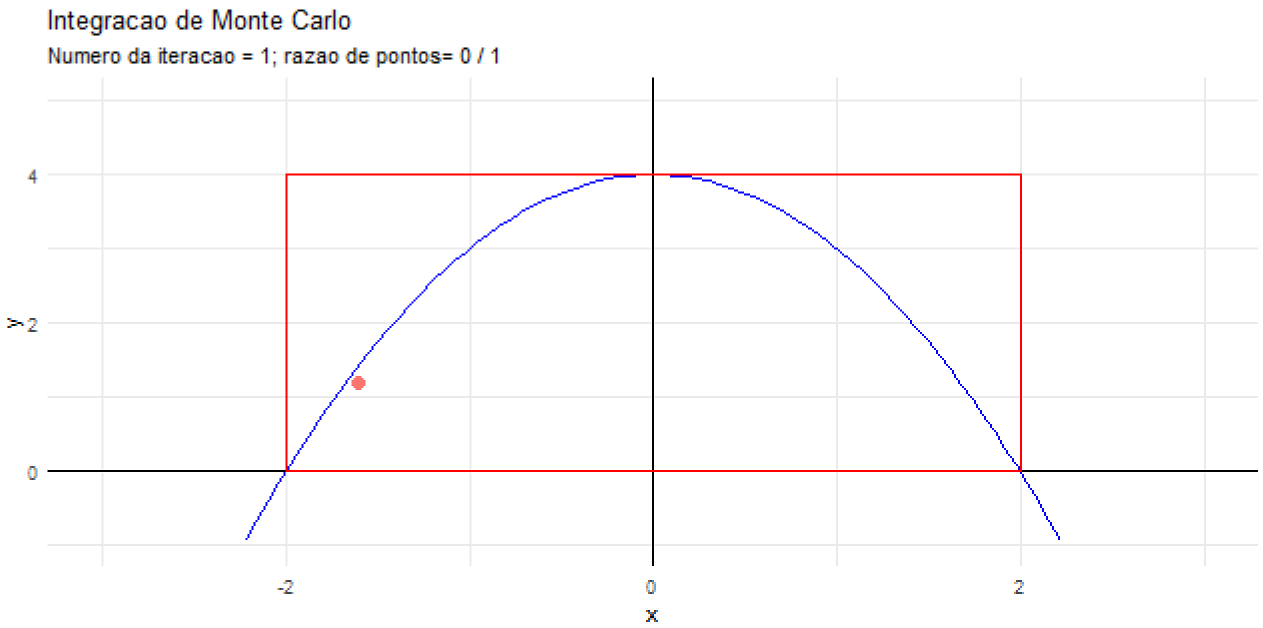
Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**

$$\int_0^h f(x)dx = \int_{-2}^{2} -x^2 + 4\,dx$$



Geração de variáveis aleatórias independentes com mesma distribuição

carloszarzar_@hotmail.com

- Número de pontos **fora** da curva = 172;
- Número de pontos **abaixo** da curva = 328;
- Número de pontos **Totais** = 500;

$$\text{Área } f(x) = \frac{\text{Número}}{\text{Número}}$$

4

4

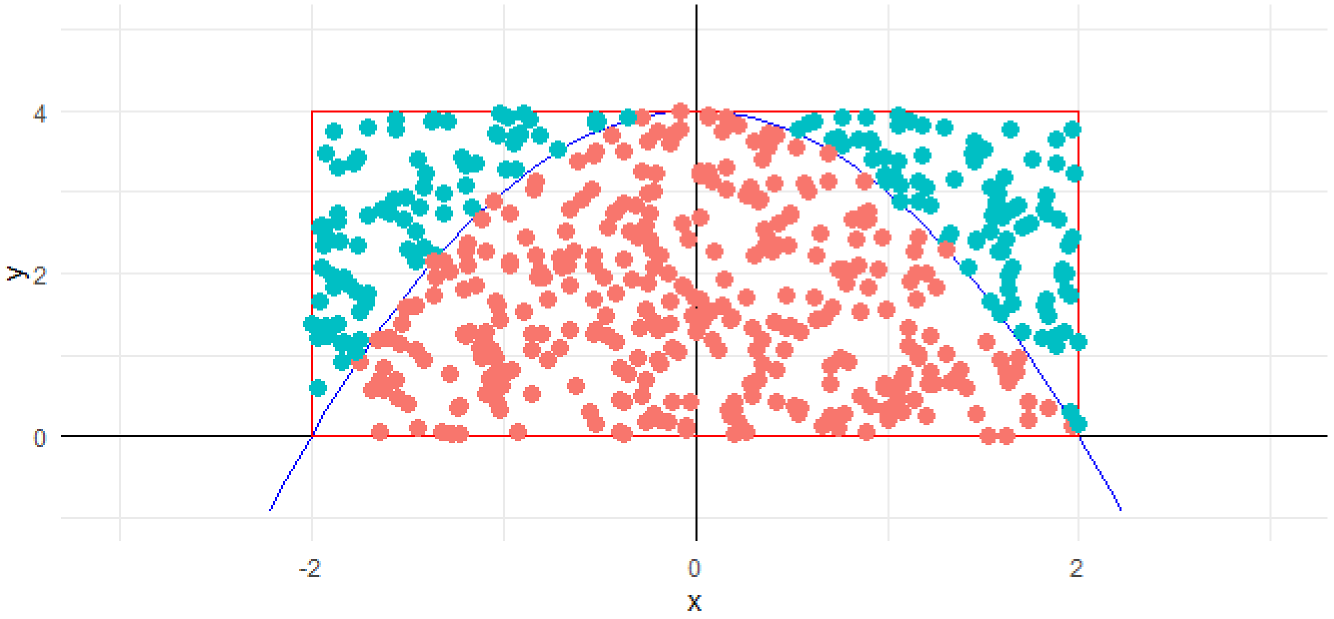
$$\int_0^h f(x)dx = \int_{-2}^{2} -x^2 + 4\, dx$$

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



Geração de variáveis aleatórias independentes com mesma distribuição

- Número de pontos **fora** da curva = 172;
- Número de pontos **abaixo** da curva = 328;
- Número de pontos **Totais** = 500;

$$\text{Área } f(x) = \frac{\text{Número}}{\text{Número}} \cdot 4 \cdot 4$$

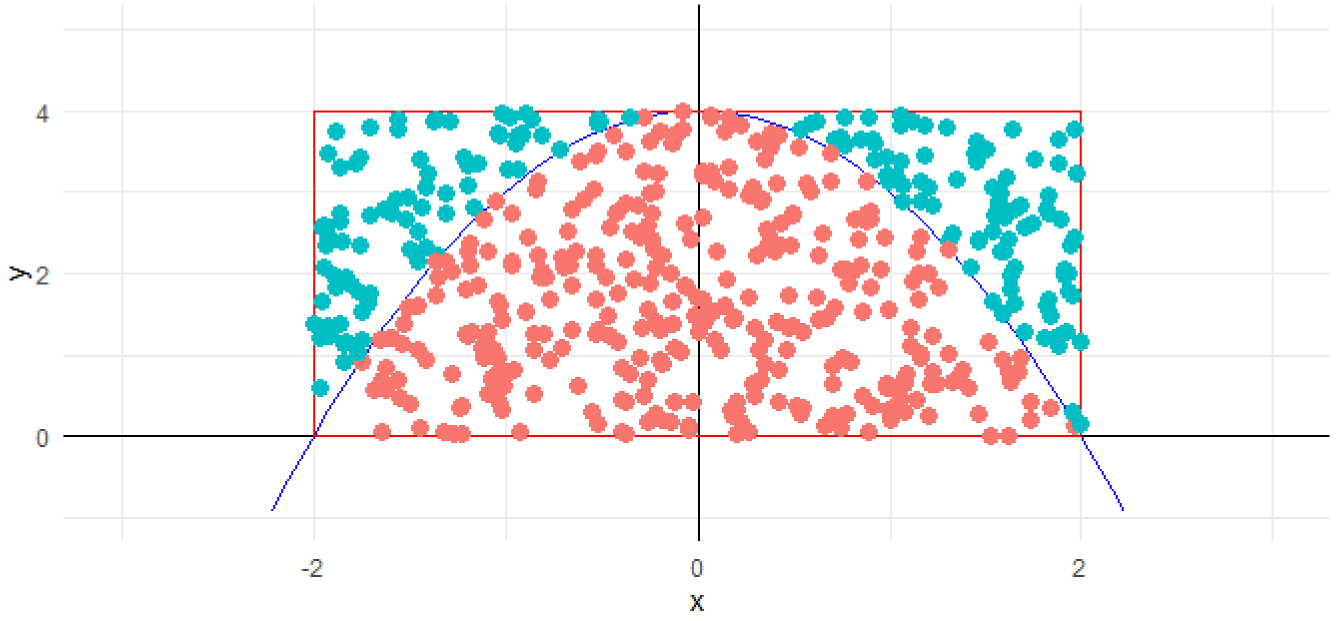
$$\text{Área } f(x) = \frac{328}{500} * (4x4) = 10{,}496$$

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



Geração de variáveis aleatórias independentes com mesma distribuição

- Número de pontos **fora** da curva = 172;
- Número de pontos **abaixo** da curva = 328;
- Número de pontos **Totais** = 500;

$$\text{Área } f(x) = \frac{\text{Número}}{\text{Número}}$$

4

4

$$\text{Área } f(x) = \frac{328}{500} * (4x4) = 10{,}496$$

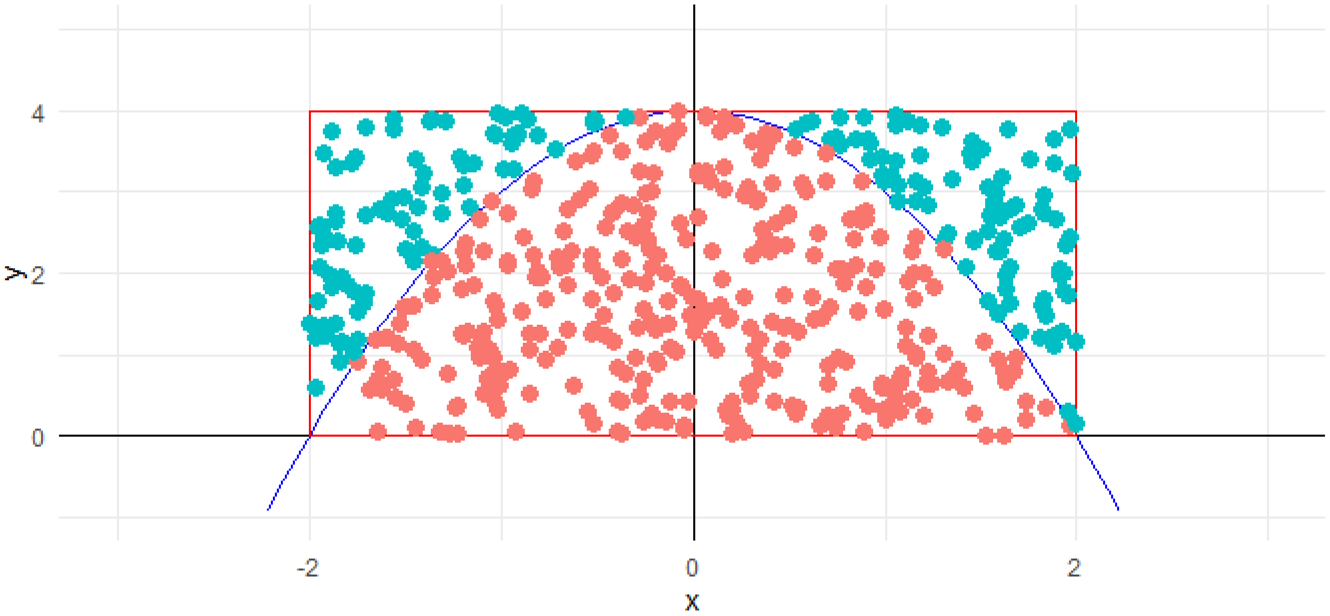
$$\int_0^h f(x)dx = \int_{-2}^{2} -x^2 + 4\ dx = 10{,}667$$

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



Geração de variáveis aleatórias independentes com mesma distribuição

carloszarzar_@hotmail.com

- Número de pontos **fora** da curva = 172;
- Número de pontos **abaixo** da curva = 328;
- Número de pontos **Totais** = 500;

$$\text{Área } f(x) = \frac{\text{Número}}{\text{Número}}$$

4

4

➢ Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis $i.i.d.$

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu$$

Convergindo quase certo

$$n \to \infty$$

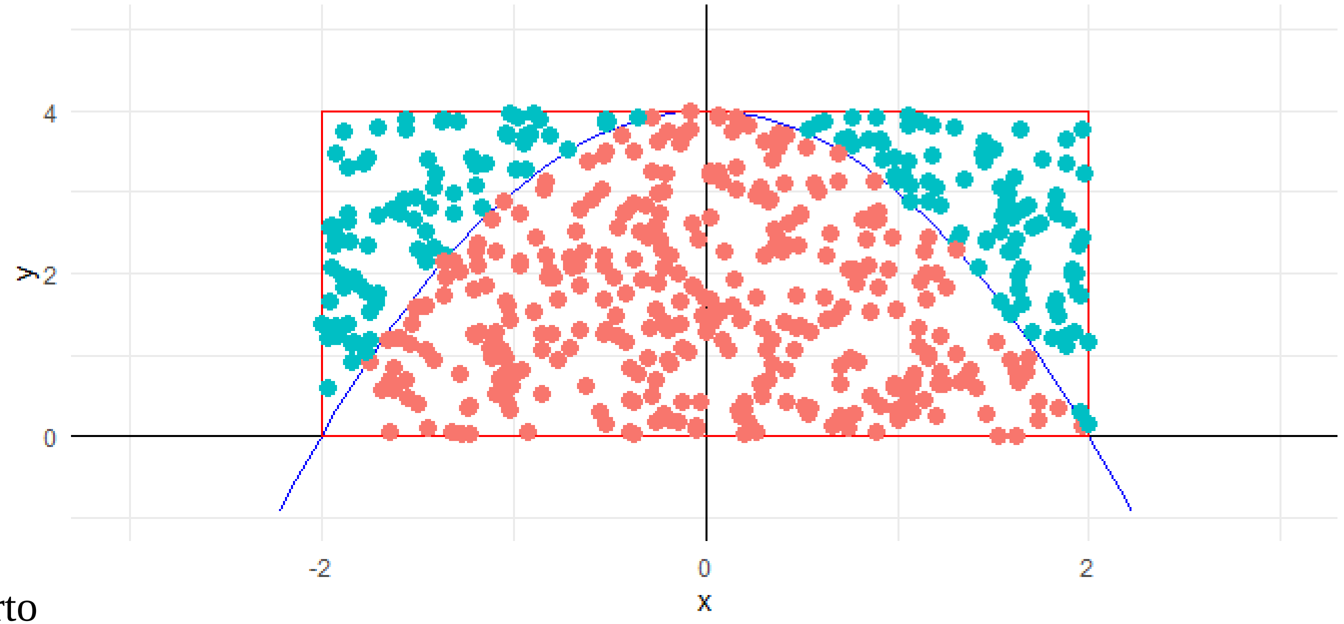
$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



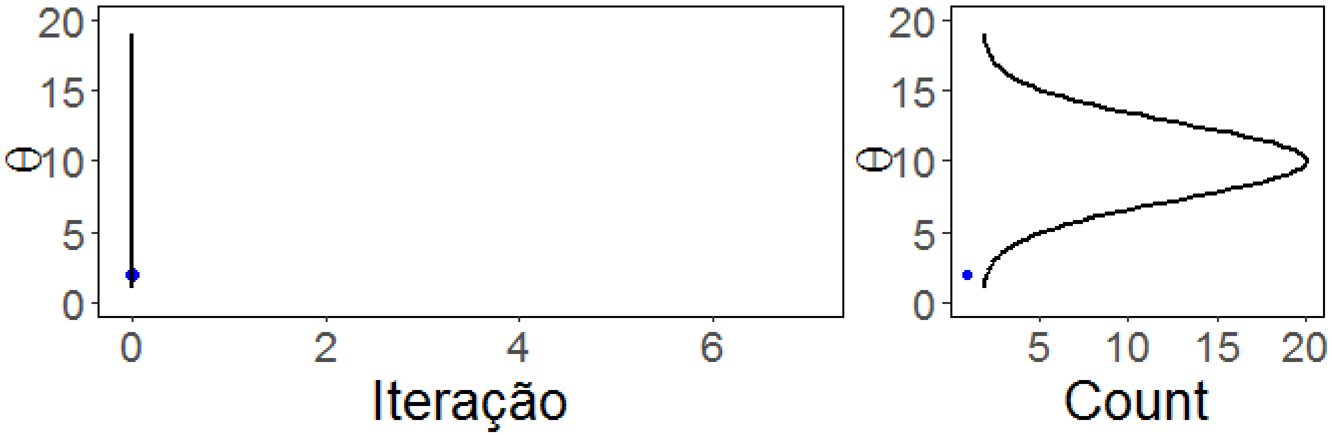
Geração de variáveis aleatórias independentes com mesma distribuição

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



➢ Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis *i.i.d.*

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu$$ Convergindo quase certo $n \to \infty$

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

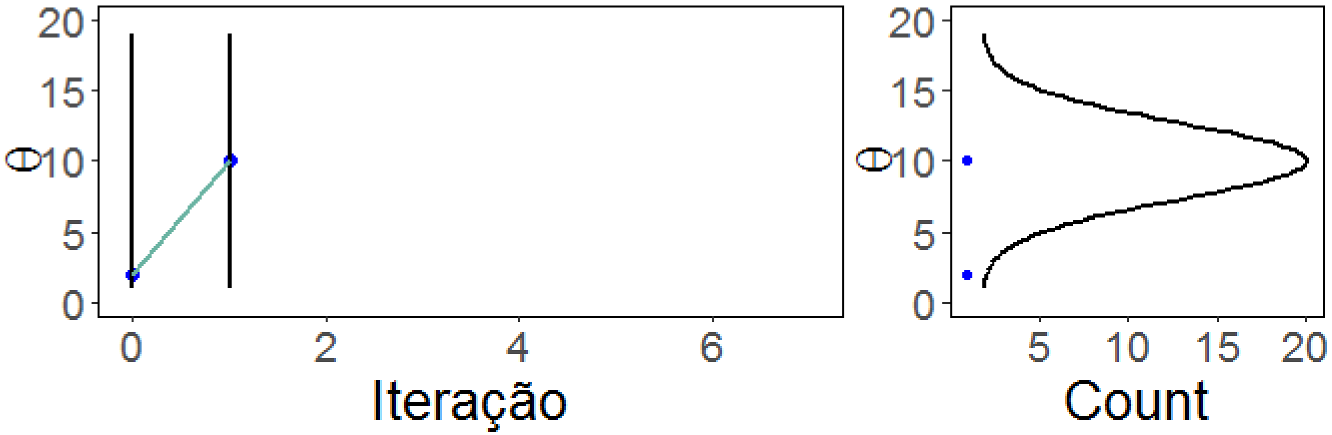
Geração de variáveis aleatórias independentes com mesma distribuição

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



- Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis $i.i.d.$

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu$$

Convergindo quase certo

$$n \to \infty$$

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

Geração de variáveis aleatórias independentes com mesma distribuição

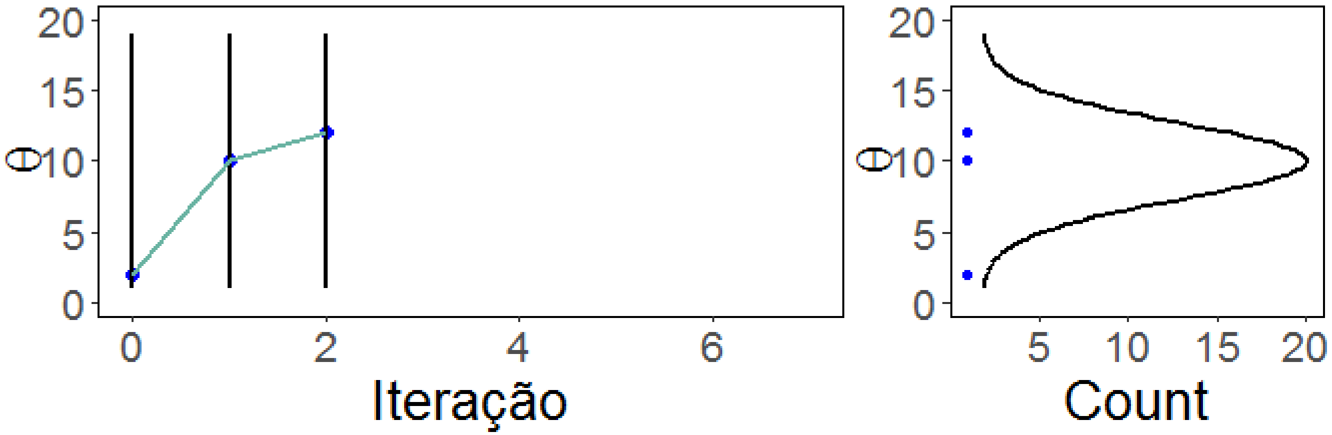
carloszarzar_@hotmail.com

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



➢ Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis $i.i.d.$

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \rightarrow \mu$$ Convergindo quase certo $n \rightarrow \infty$

$$P\left(\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

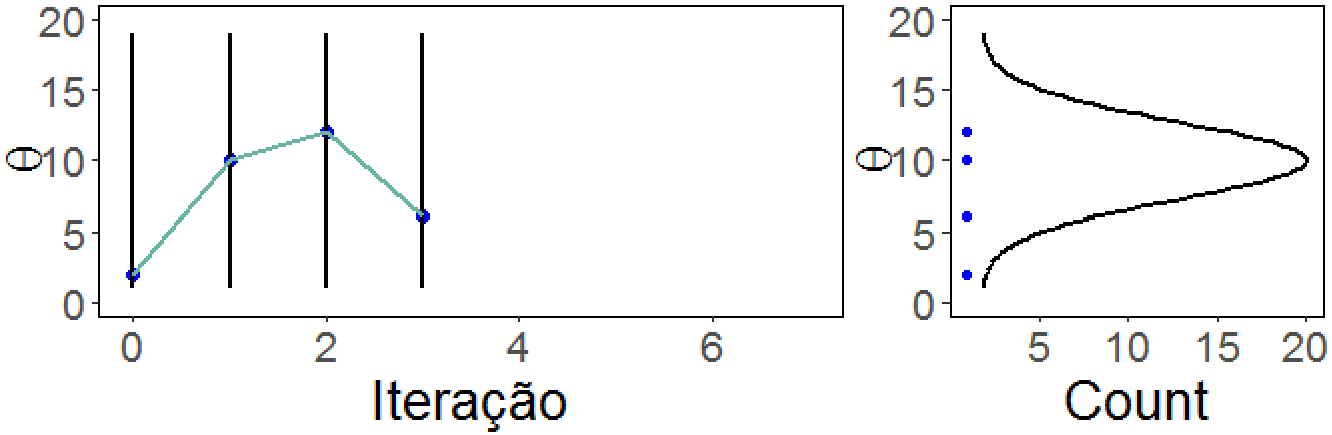
Geração de variáveis aleatórias independentes com mesma distribuição

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



Geração de variáveis aleatórias independentes com mesma distribuição

➢ Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis $i.i.d.$

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \rightarrow \mu$$

Convergindo quase certo

$$n \rightarrow \infty$$

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

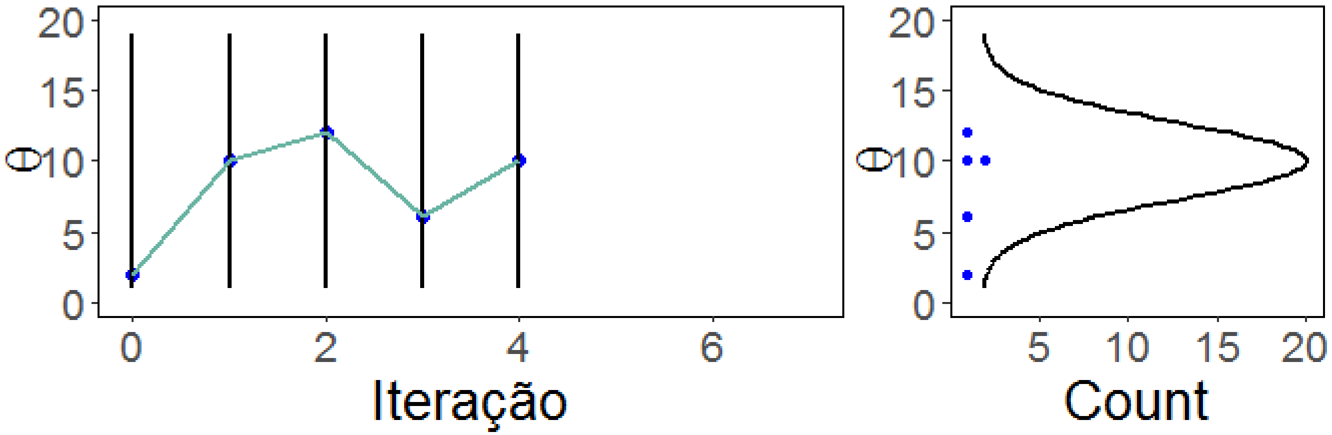
carloszarzar_@hotmail.com

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



➢ Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis $i.i.d.$

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \rightarrow \mu$$ Convergindo quase certo $n \rightarrow \infty$

$$P\left(\lim_{n\rightarrow\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

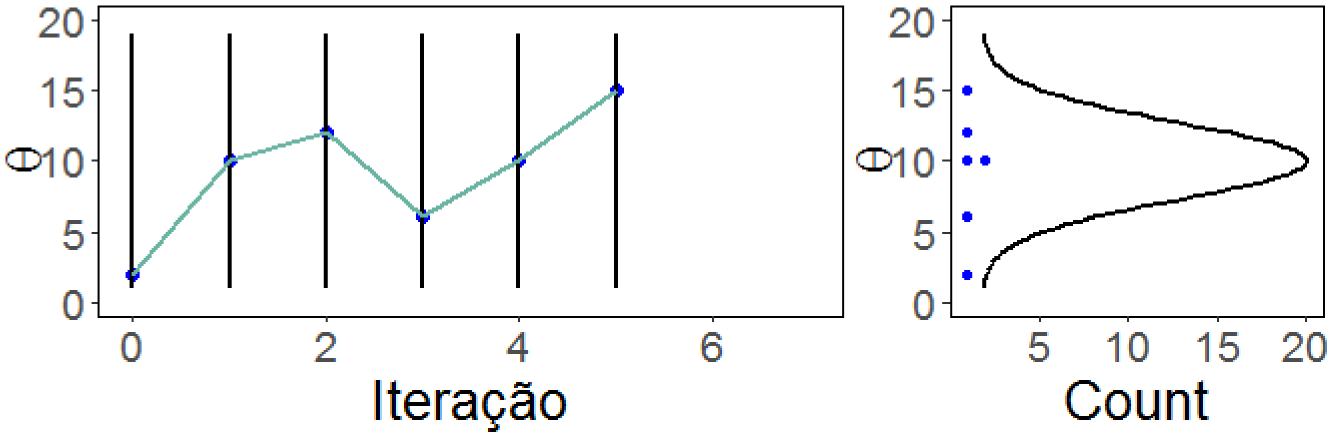
Geração de variáveis aleatórias independentes com mesma distribuição

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



➢ Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis $i.i.d.$

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu$$ Convergindo quase certo $n \to \infty$

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

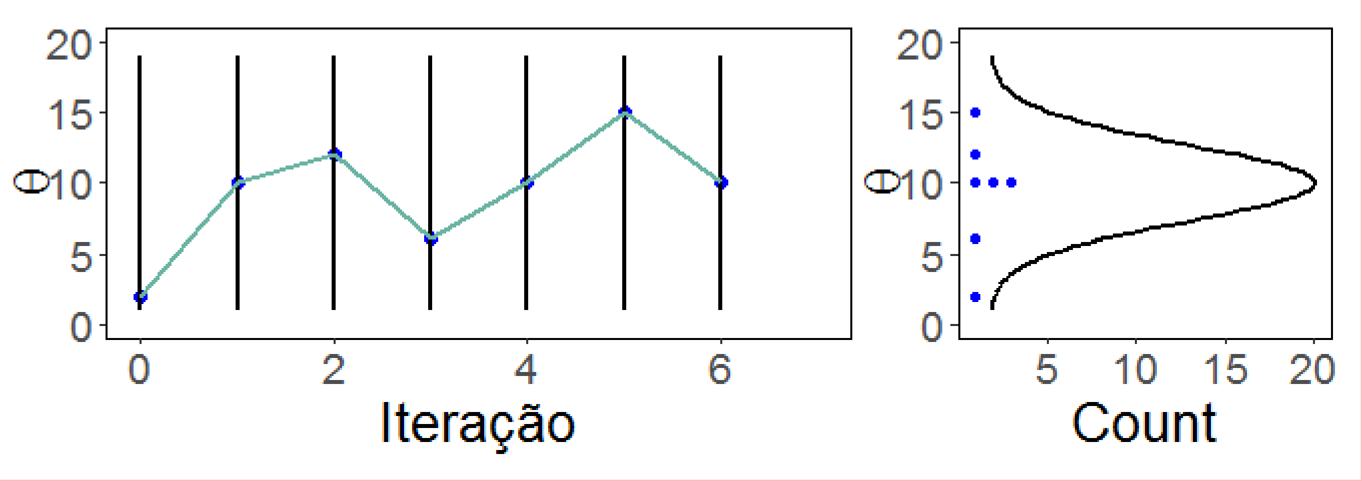
Geração de variáveis aleatórias independentes com mesma distribuição

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



- Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis $i.i.d.$

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \rightarrow \mu$$

Convergindo quase certo

$$n \rightarrow \infty$$

$$P\left(\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

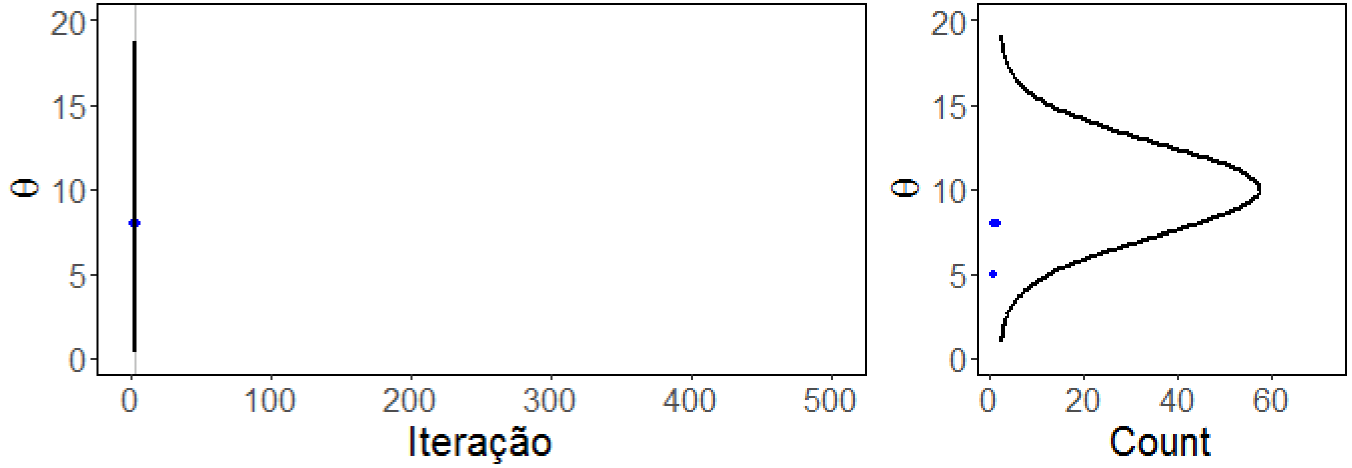
Geração de variáveis aleatórias independentes com mesma distribuição

Monte Carlo é um método de integração numérico conhecido:

- **Integração de Monte Carlo;**

ou

- **Método de Monte Carlo.**



- Lei Forte dos Grandes Números

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis $i.i.d.$

$$\mu = E[Y] < \infty$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu$$

Convergindo quase certo

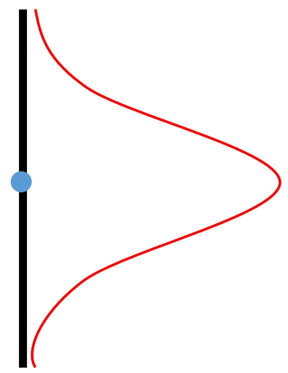
$$n \to \infty$$

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1$$

Geração de variáveis aleatórias independentes com mesma distribuição

Monte Carlo Via Cadeia de Markov:

$$\theta_t \sim N(\theta_{t-1}, \sigma^2)$$



$t-1$

A posteriori: Bimodal

Distribuição Normal mista

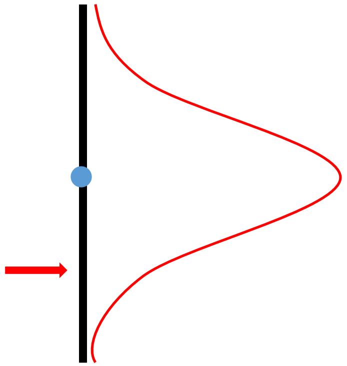
$$f(x) = 0{,}4 * N(17; 3) + 0{,}6 * N(30; 2)$$

Geração de variáveis aleatórias

~~independentes~~ dependentes

Monte Carlo Via Cadeia de Markov:

$$\theta_t \sim N(\theta_{t-1}, \sigma^2)$$



$t-1$

A posteriori: Bimodal

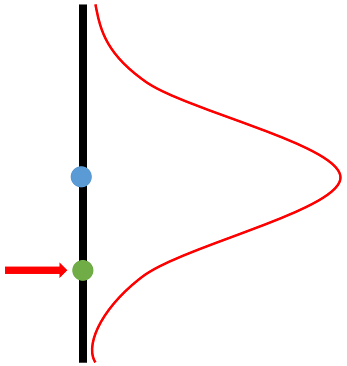
Distribuição Normal mista

$$f(x) = 0{,}4 * N(17;3) + 0{,}6 * N(30;2)$$

Geração de variáveis aleatórias
~~independentes~~ dependentes

Monte Carlo Via Cadeia de Markov:

$$\theta_t \sim N(\theta_{t-1}, \sigma^2)$$



$t-1$

A posteriori: Bimodal

Distribuição Normal mista

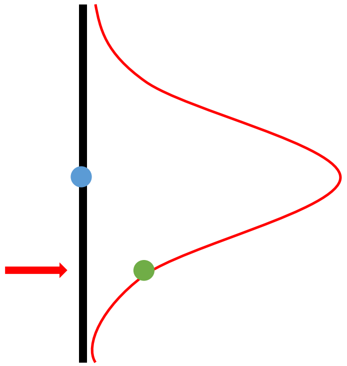
$$f(x) = 0{,}4 * N(17; 3) + 0{,}6 * N(30; 2)$$

Geração de variáveis aleatórias
~~independentes~~ dependentes

carloszarzar_@hotmail.com

Monte Carlo Via Cadeia de Markov:

$\theta_t \sim N(\theta_{t-1}, \sigma^2)$



$t-1$

A posteriori: Bimodal

Distribuição Normal mista

$$f(x) = 0{,}4 * N(17; 3) + 0{,}6 * N(30; 2)$$

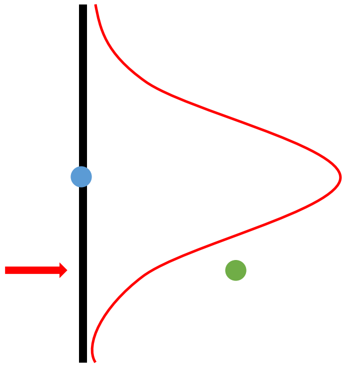
Geração de variáveis aleatórias

~~independentes~~ dependentes

carloszarzar_@hotmail.com

Monte Carlo Via Cadeia de Markov:

$$\theta_t \sim N(\theta_{t-1}, \sigma^2)$$



$t - 1$

A posteriori: Bimodal

Distribuição Normal mista

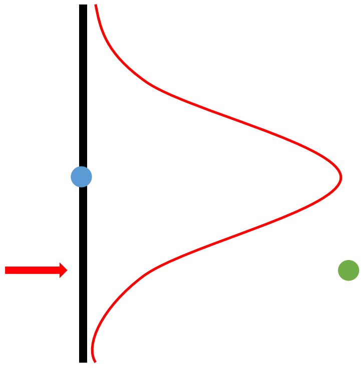
$$f(x) = 0{,}4 * N(17;3) + 0{,}6 * N(30;2)$$

Geração de variáveis aleatórias

~~independentes~~ dependentes

Monte Carlo Via Cadeia de Markov:

$$\theta_t \sim N(\theta_{t-1}, \sigma^2)$$



$t-1$

A posteriori: Bimodal

Distribuição Normal mista

$$f(x) = 0{,}4 * N(17;3) + 0{,}6 * N(30;2)$$

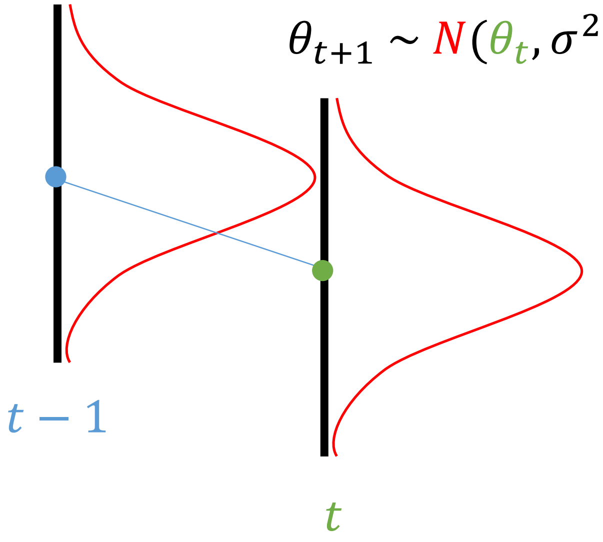
Geração de variáveis aleatórias ~~independentes~~ dependentes

carloszarzar_@hotmail.com

Monte Carlo Via Cadeia de Markov:

$\theta_t \sim N(\theta_{t-1}, \sigma^2)$

$\theta_{t+1} \sim N(\theta_t, \sigma^2)$



A posteriori: Bimodal

Distribuição Normal mista

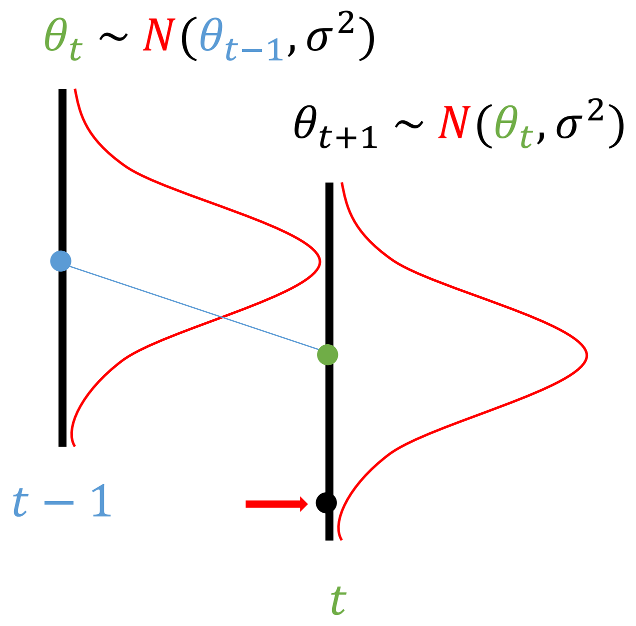
$$f(x) = 0{,}4 * N(17;3) + 0{,}6 * N(30;2)$$

Geração de variáveis aleatórias

~~independentes~~ dependentes

carloszarzar_@hotmail.com

Monte Carlo Via Cadeia de Markov:



A posteriori: Bimodal

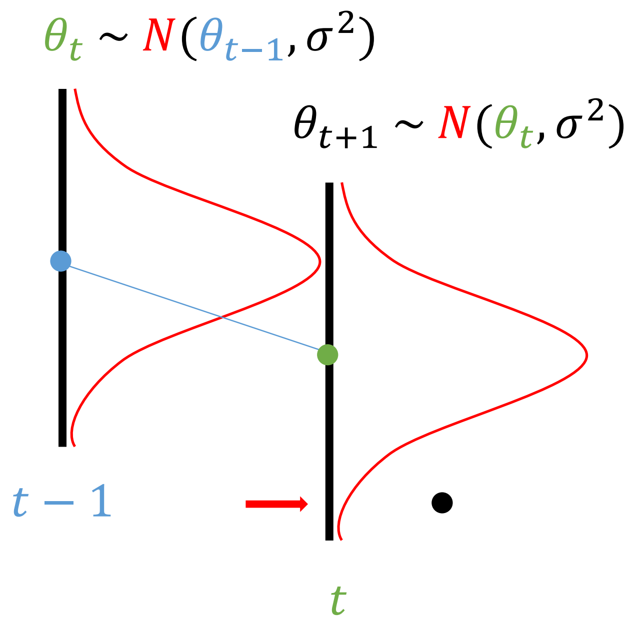
Distribuição Normal mista

$$f(x) = 0{,}4 * N(17; 3) + 0{,}6 * N(30; 2)$$

Geração de variáveis aleatórias

~~independentes~~ dependentes

Monte Carlo Via Cadeia de Markov:



A posteriori: Bimodal

Distribuição Normal mista

$$f(x) = 0{,}4 * N(17; 3) + 0{,}6 * N(30; 2)$$

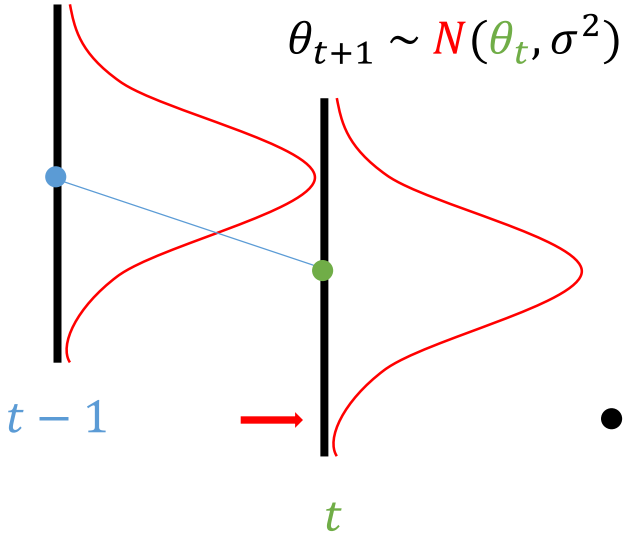
Geração de variáveis aleatórias
~~independentes~~ dependentes

carloszarzar_@hotmail.com

Monte Carlo Via Cadeia de Markov:

$\theta_t \sim N(\theta_{t-1}, \sigma^2)$

$\theta_{t+1} \sim N(\theta_t, \sigma^2)$



A posteriori: Bimodal

Distribuição Normal mista

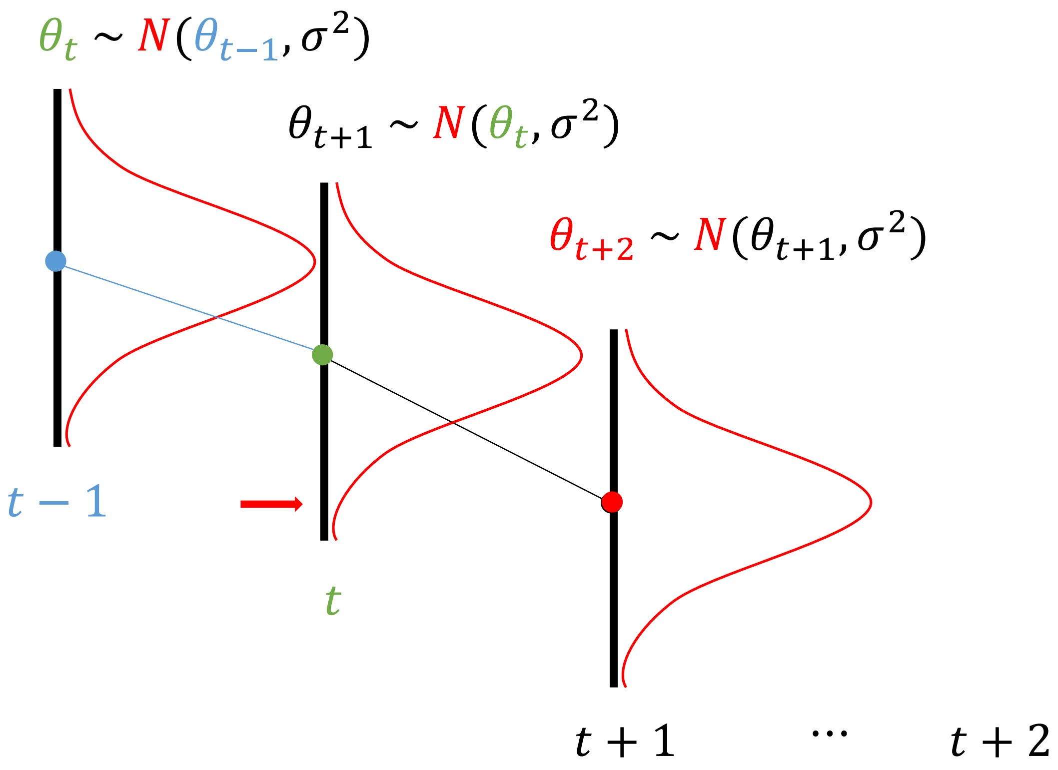
$$f(x) = 0{,}4 * N(17; 3) + 0{,}6 * N(30; 2)$$

Geração de variáveis aleatórias

~~independentes~~ dependentes

carloszarzar_@hotmail.com

Monte Carlo Via Cadeia de Markov:



A posteriori: Bimodal

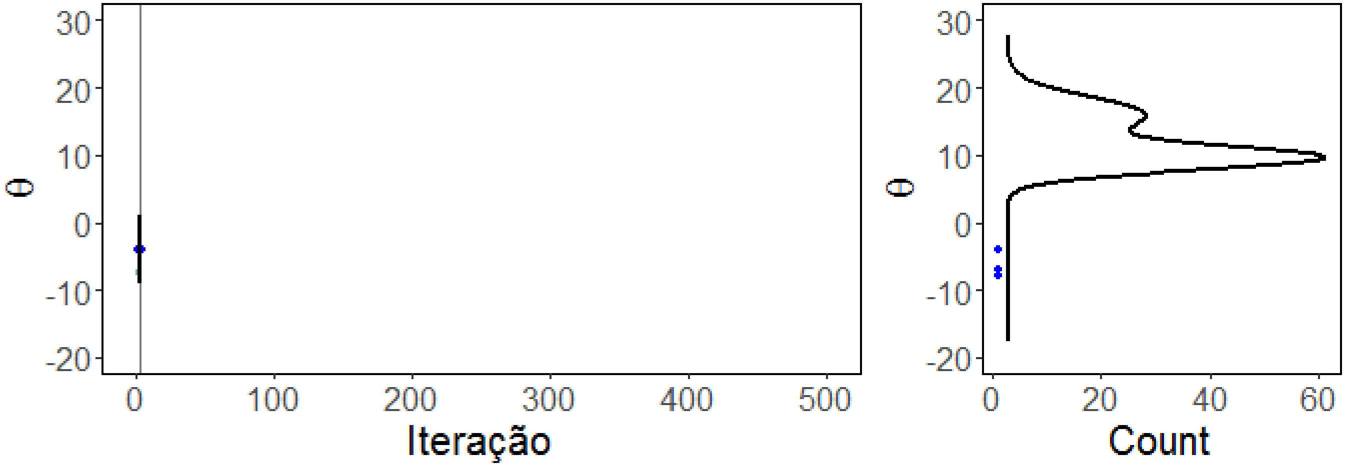
Distribuição Normal mista

$$f(x) = 0{,}4 * N(17;3) + 0{,}6 * N(30;2)$$

Geração de variáveis aleatórias

~~independentes~~ dependentes

Monte Carlo Via Cadeia de Markov:



A posteriori: Bimodal

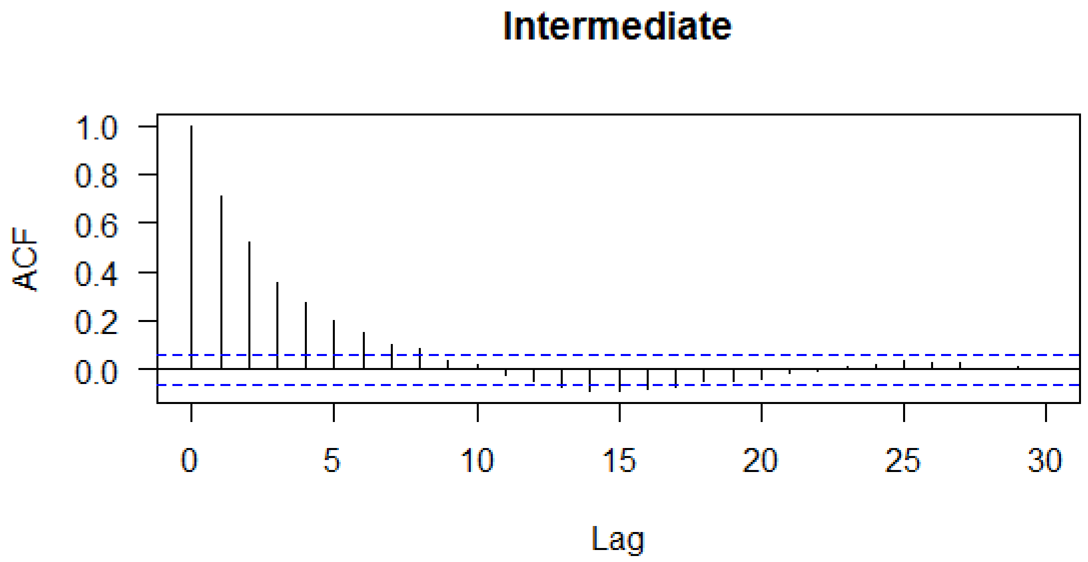
Distribuição Normal mista

$$f(x) = 0{,}4 * N(17;3) + 0{,}6 * N(30;2)$$

$$\theta_t \sim N(\theta_{t-1};\sigma)$$

Geração de variáveis aleatórias ~~independentes~~ dependentes

Monte Carlo Via Cadeia de Markov:



A posteriori: Bimodal

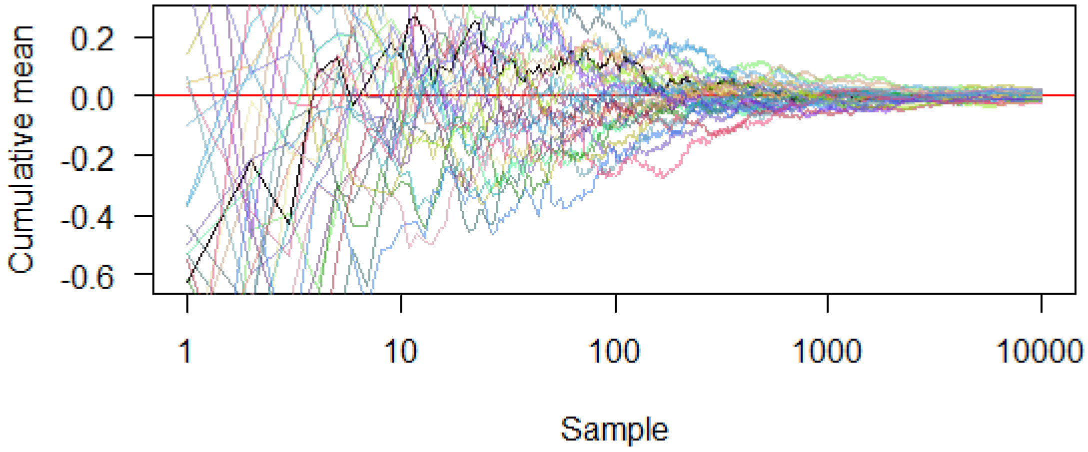
Distribuição Normal mista

$$f(x) = 0{,}4 * N(17; 3) + 0{,}6 * N(30; 2)$$

$$\theta_t \sim N(\theta_{t-1}; \sigma)$$

Geração de variáveis aleatórias ~~independentes~~ dependentes

Monte Carlo Via Cadeia de Markov:



A posteriori: Bimodal

Distribuição Normal mista

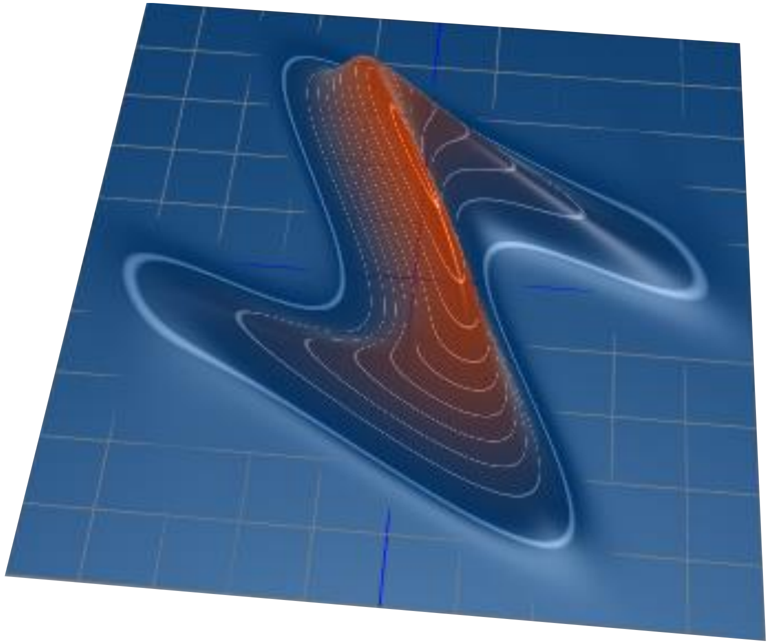
$$f(x) = 0{,}4 * N(17;3) + 0{,}6 * N(30;2)$$

Geração de variáveis aleatórias

~~independentes~~ dependentes

carloszarzar_@hotmail.com

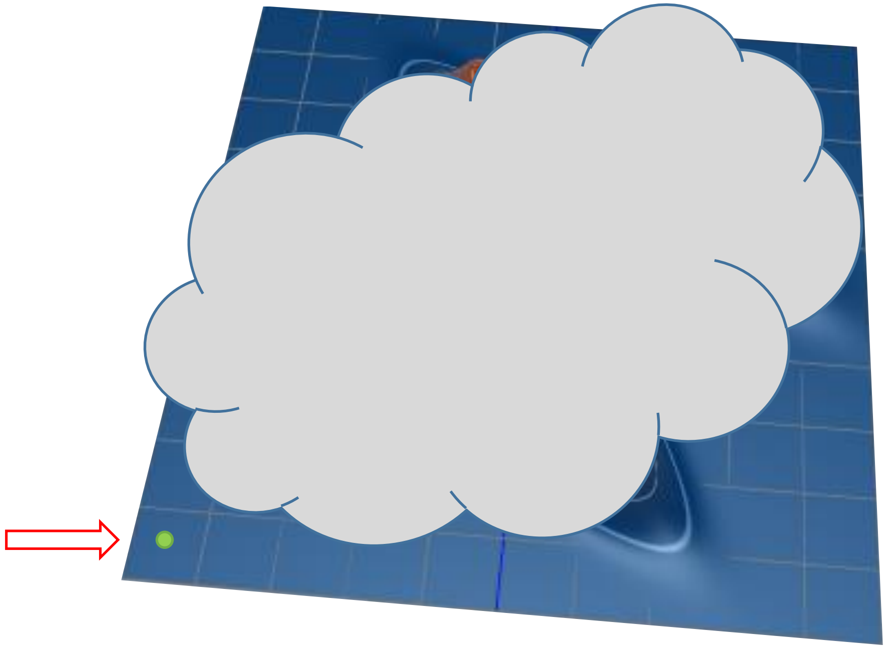
Espaço paramétrico da
distribuição *a posteriori*



Espaço paramétrico da distribuição *a posteriori* ⇨ Desconhecido

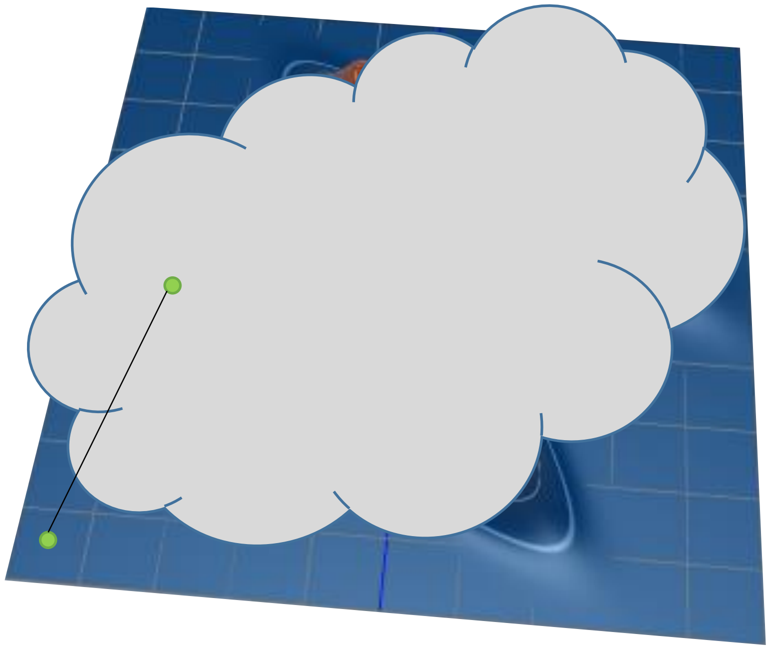
carloszarzar_@hotmail.com

Espaço paramétrico da
distribuição *a posteriori*



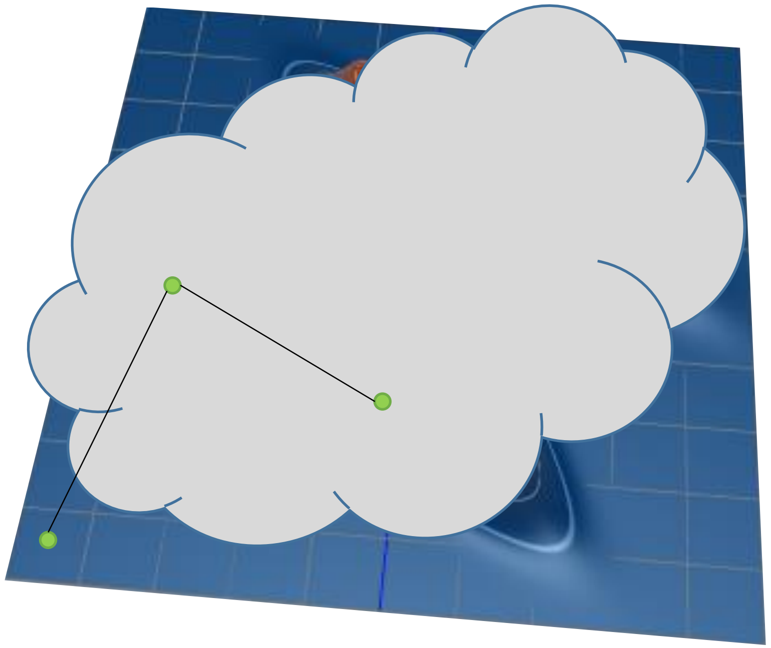
- *Random walk MCMC*

Espaço paramétrico da distribuição *a posteriori*



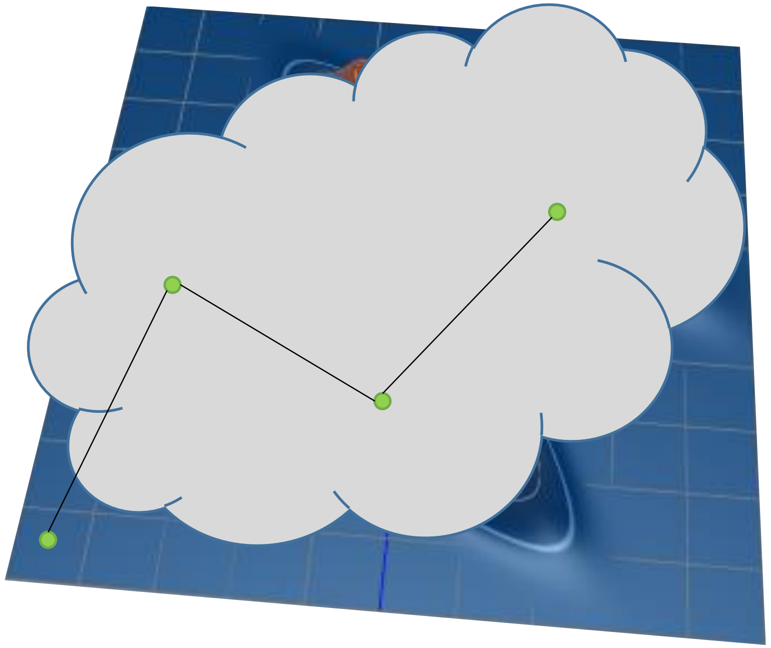
- *Random walk MCMC*

Espaço paramétrico da
distribuição *a posteriori*



- *Random walk MCMC*

Espaço paramétrico da distribuição *a posteriori*

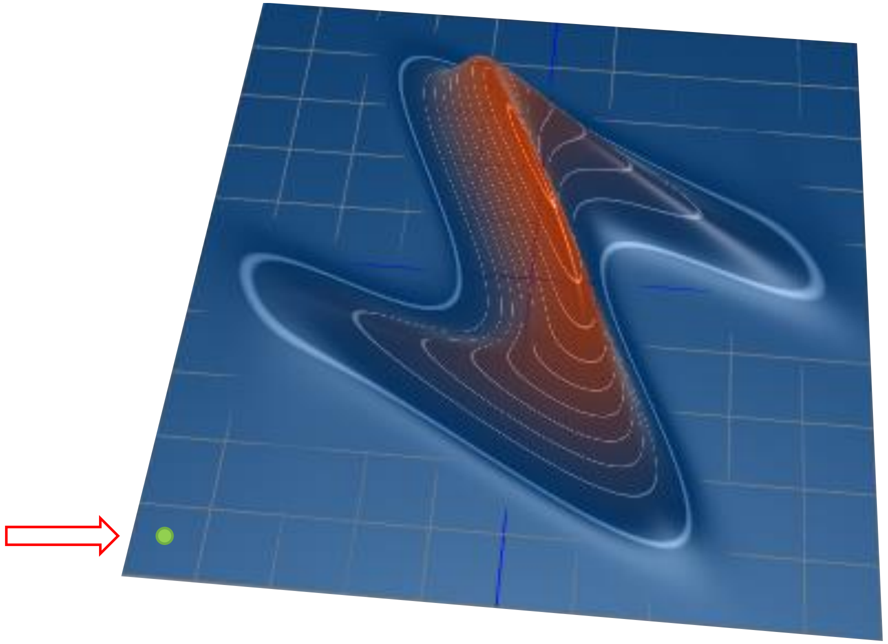


- *Random walk MCMC*

Custo computacional muito grande para amostrar todo o espaço paramétrico.

carloszarzar_@hotmail.com

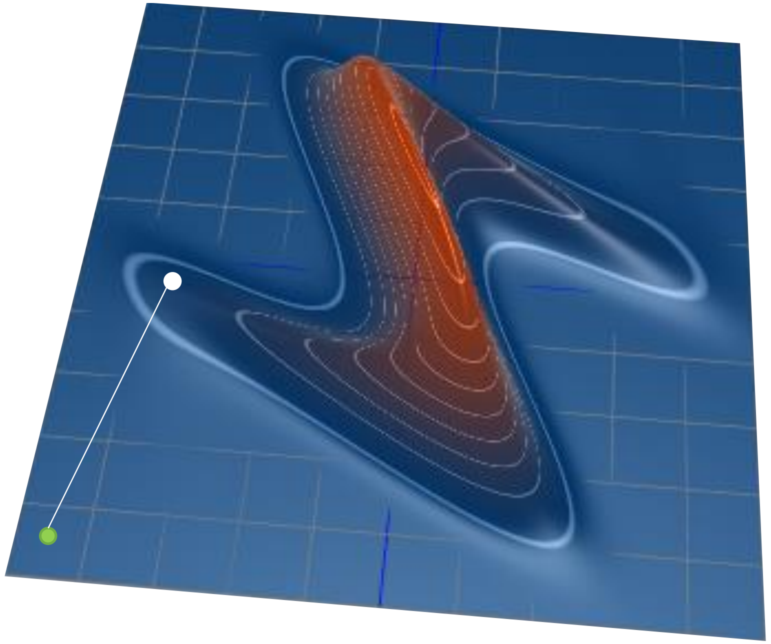
Espaço paramétrico da
distribuição *a posteriori*



- *Random walk MCMC*

Etapa de aceitação ou
rejeição de cada iteração do
passeio aleatório.

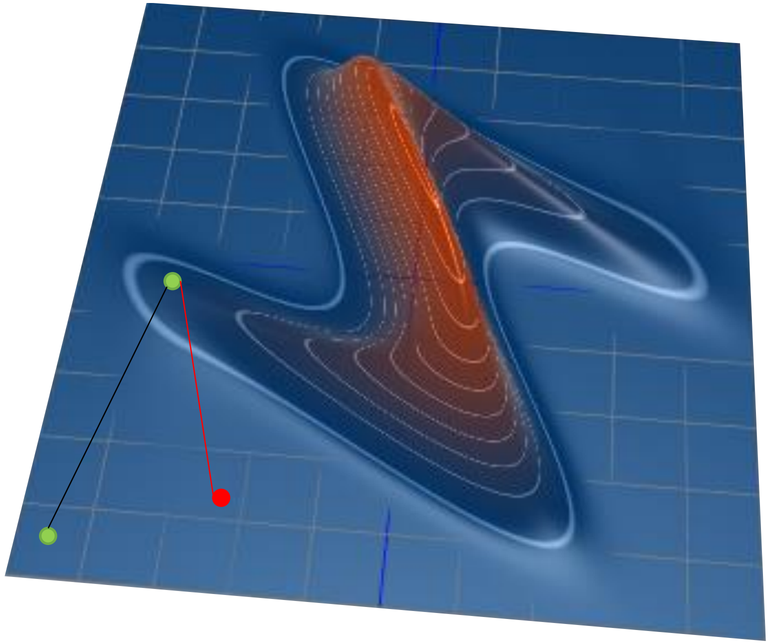
Espaço paramétrico da
distribuição *a posteriori*



- *Random walk MCMC*
- Amostrador em avaliação;

Etapa de aceitação ou
rejeição de cada iteração do
passeio aleatório.

Espaço paramétrico da
distribuição *a posteriori*



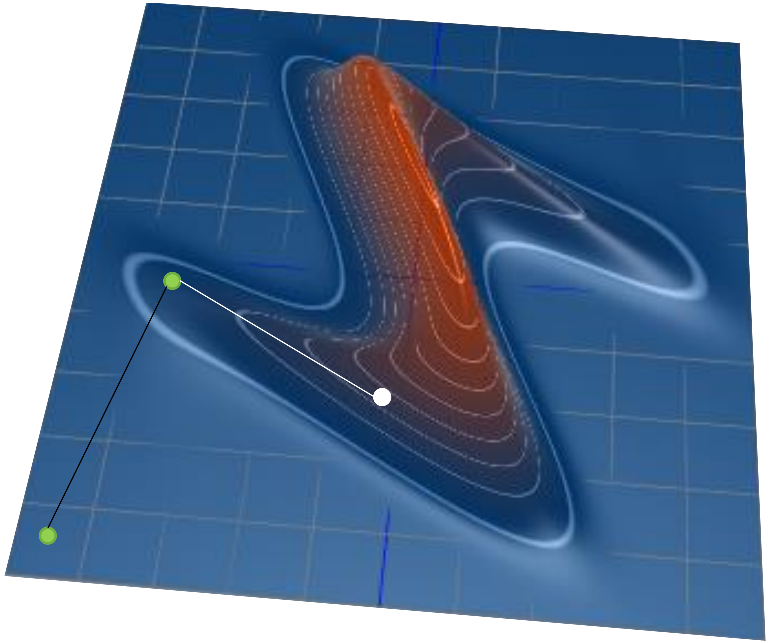
Random walk MCMC

Amostrador em avaliação;

Amostrador rejeitado;

Etapa de aceitação ou
rejeição de cada iteração do
passeio aleatório.

Espaço paramétrico da
distribuição *a posteriori*



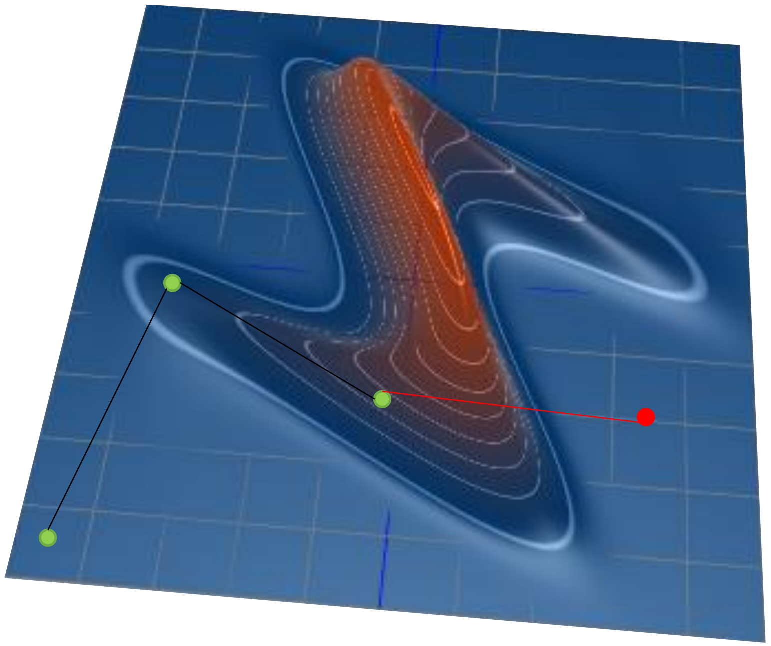
Random walk MCMC

Amostrador em avaliação;

Amostrador rejeitado;

Etapa de aceitação ou
rejeição de cada iteração do
passeio aleatório.

Espaço paramétrico da
distribuição *a posteriori*

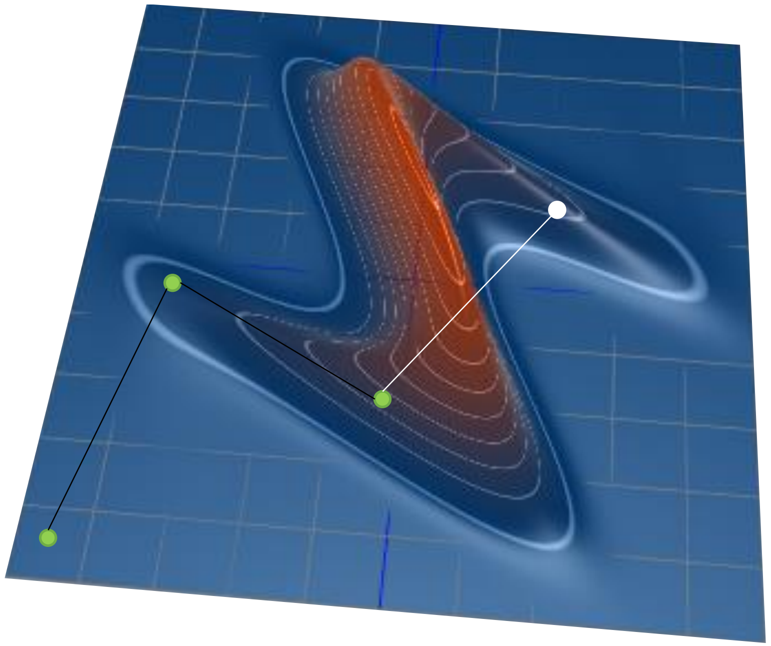


- *Random walk MCMC*
- Amostrador em avaliação;
- Amostrador rejeitado;

Etapa de aceitação ou
rejeição de cada iteração do
passeio aleatório.

carloszarzar_@hotmail.com

Espaço paramétrico da
distribuição *a posteriori*



- *Random walk MCMC*
- Amostrador em avaliação;
- Amostrador rejeitado;

- <u>Método a amostragem:</u>
  - Gibbs
  - Metropolis-Hasting
  - Hamiltoniano MC
  - NUTS

Etapa de aceitação ou
rejeição de cada iteração do
passeio aleatório.

Introdução a linguagem Stan (*rstan*), um software para modelos bayesianos.

Universidade Federal do Oeste do Pará (UFOPA)
Campus de Monte Alegre – Engenharia de Aquicultura

Obrigado !!!

AGRADECIMENTO E COLABORADORES:

UNIVERSIDADE FEDERAL DE LAVRAS

NÚCLEO DE ESTUDOS EM REGRESSÃO NÃO LINEAR APLICADA

UFOPA

Professor: Carlos Antônio Zarzar
E-mail: carloszarzar_@hotmail.com
carlos.zarzar@ufopa.edu.br

Data: 09/03/2022