Disciplina: Estatística Computacional e Otimização

Prof.: Luiz Duczmal

Lista: 4

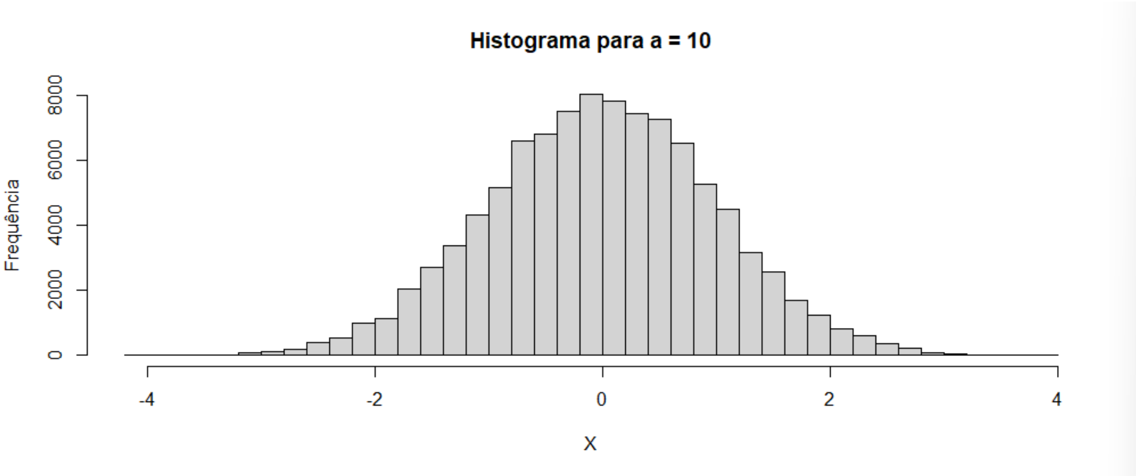
Data de entrega: 05/06/24

Estudante: Pedro Mateus Moraes de Almeida

Matrícula: 2023668250

Lista 4

(1) Implemente o algoritmo Metropolis-Hastings (MCMC) para gerar a distribuição normal. Simule com diferentes valores de a para mostrar os problemas que ocorrem se a for muito grande ou muito pequeno.



Histograma e Estatísticas Descritivas

a = 0.1

- Média: -0.1386
- Desvio Padrão: 1.0827
- Assimetria: 0.0615
- Curtose: 2.9064
- Teste de Shapiro-Wilk: p-valor = 0.04019 (rejeitamos a normalidade)
- Teste de Kolmogorov-Smirnov: p-valor < 2.2e-16 (rejeitamos a normalidade)

O histograma mostra que a distribuição está bem centralizada, mas o valor de a muito pequeno faz com que o algoritmo tenha uma alta taxa de aceitação, o que resulta em pequenas variações nos valores de X

a = 2

- Média: -0.0034

- Desvio Padrão: 1.0048
- Assimetria: -0.0024
- Curtose: 2.9854

O valor de a = 2 parece ser ideal, pois o histograma se assemelha muito a uma distribuição normal. As estatísticas descritivas estão próximas dos valores esperados para uma distribuição normal padrão.

Teste de Shapiro-Wilk: p-valor = 0.3404 (não rejeitamos a normalidade)

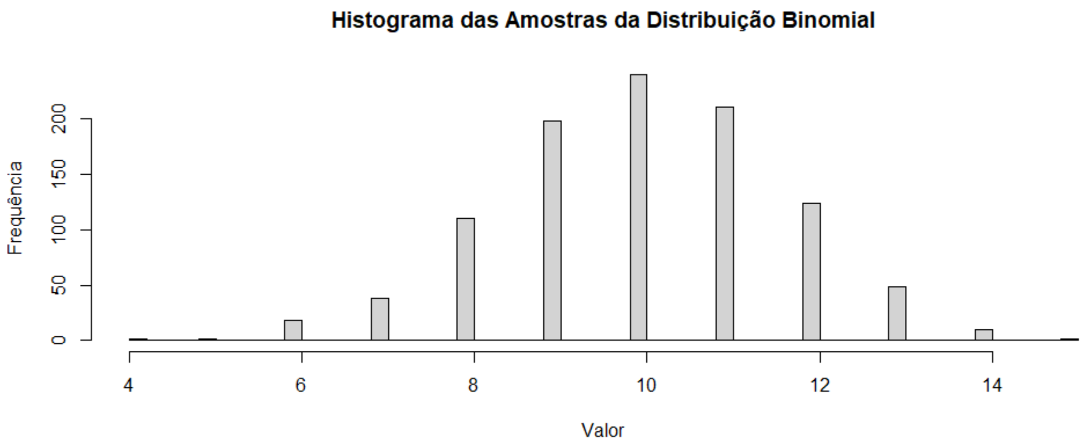
Teste de Kolmogorov-Smirnov: p-valor = 0.06491 (não rejeitamos a normalidade)

a = 10

- Média: -0.0102
- Desvio Padrão: 0.9941
- Assimetria: -0.0225
- Curtose: 2.9760
- Teste de Shapiro-Wilk: p-valor = 0.0548 (não rejeitamos a normalidade)
- Teste de Kolmogorov-Smirnov: p-valor = 6.039e-05 (rejeitamos a normalidade)

Quando a é muito grande, o algoritmo tende a ter uma alta taxa de rejeição, o que resulta em menos movimentos e maior variabilidade entre os valores aceitos. O histograma ainda se assemelha a uma distribuição normal, mas pode apresentar alguns problemas de convergência.

(2) Implemente o algoritmo Metropolis-Hastings (MCMC) para gerar a distribuição binomial.



O histograma mostra a distribuição das amostras geradas pela implementação do algoritmo Metropolis-Hastings para a distribuição binomial com parâmetros n=20 e p=0. A distribuição aparenta ser aproximadamente simétrica e centrada em torno de 10, o que é consistente com a média esperada de uma distribuição binomial com esses parâmetros.

As estatísticas descritivas calculadas para as amostras geradas são:

- Média: 10.032
- Desvio Padrão: 1.644
- Assimetria: -0.134
- Curtose: 3.035

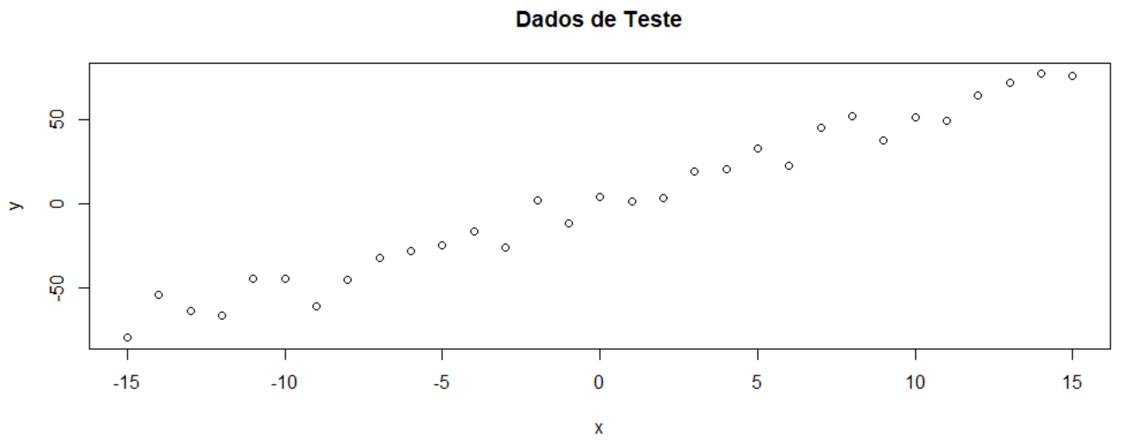
A média das amostras é próxima do valor esperado de 10. Isso indica que o algoritmo está funcionando corretamente em termos de gerar amostras com a média esperada.

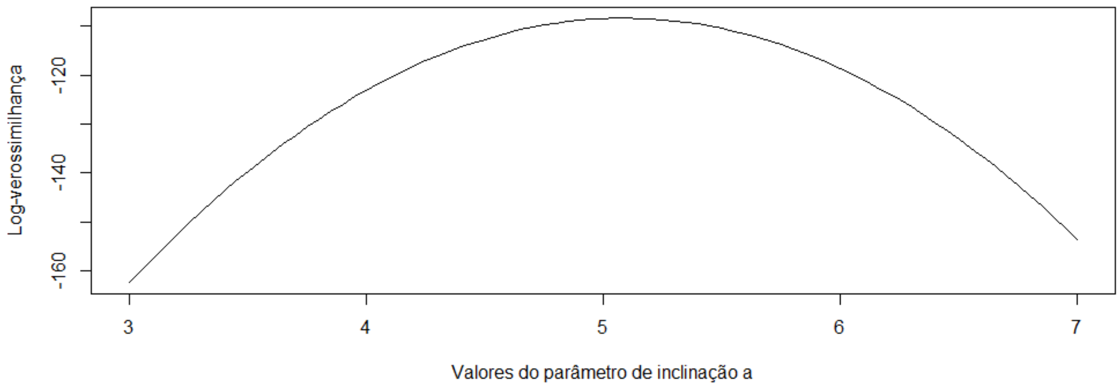
O desvio padrão das amostras é próximo do valor esperado. Embora haja uma pequena diferença, ela pode ser atribuída à variabilidade das amostras geradas.

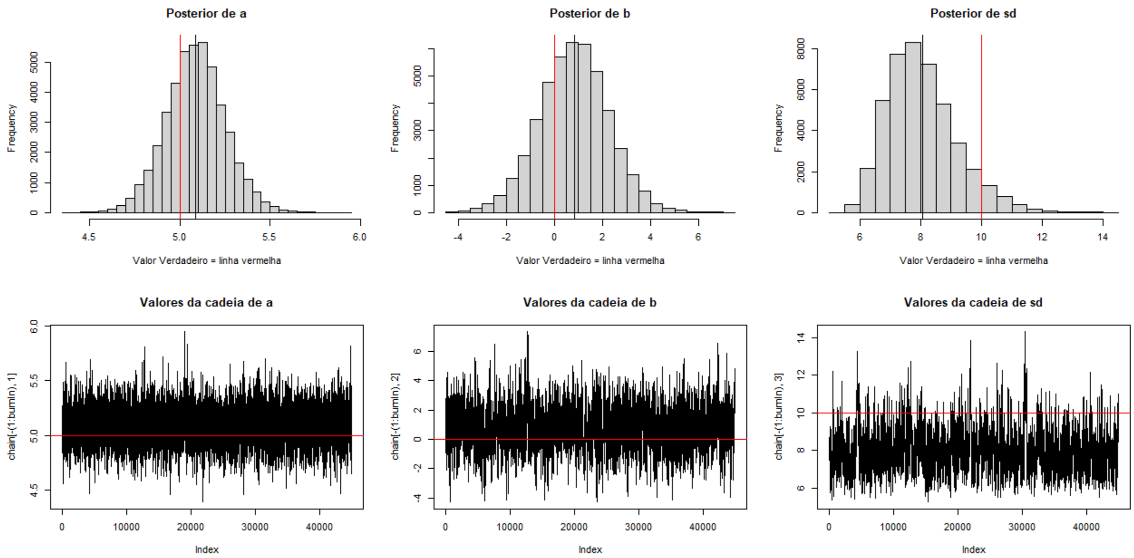
A assimetria ligeiramente negativa sugere que a distribuição das amostras tem uma leve inclinação para a esquerda, embora esteja bastante próxima de 0

A curtose é próxima de 3, o que é típico de uma distribuição normal.

(3) Implemente o algoritmo Metropolis-Hastings (MCMC) para o problema de regressão linear visto em aula
https://theoreticalecology.wordpress.com/2010/09/17/metropolis-hastings-mcmc-in-r/
e compare com a rotina tradicional de regressão linear para a amostra utilizada.







Resultados:

Estatísticas Descritivas:

Para o parâmetro a:

- Média: 5.0866
- Desvio Padrão: 0.1646
- Assimetria: 0.0845
- Curtose: 3.4111

Para o parâmetro b:

- Média: 0.8189
- Desvio Padrão: 1.4548
- Assimetria: 0.0536
- Curtose: 3.3616

Para o parâmetro sd:

- Média: 8.0546
- Desvio Padrão: 1.1612
- Assimetria: 0.8259
- Curtose: 4.0871

Intercepto (b):

- Estimativa: 0.7865
- Erro Padrão: 1.3667
- Valor t: 0.576
- Valor p: 0.569

Inclinação (a):

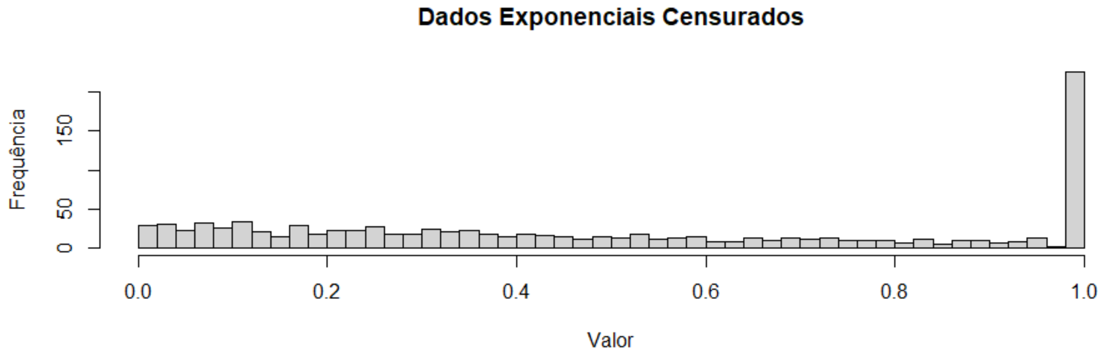
- Estimativa: 5.0886
- Erro Padrão: 0.1528
- Valor t: 33.303
- Valor p: < 2e-16

Os resultados do algoritmo Metropolis-Hastings para a inclinação a e o intercepto b são bastante próximos dos obtidos pela regressão linear tradicional, o que indica que o MCMC está funcionando corretamente.

Já média do desvio padrão (sd) obtida pelo MCMC é maior do que a estimativa do erro padrão residual da regressão linear tradicional, indicando que o MCMC pode estar capturando uma maior variabilidade nos dados.

Por fim as distribuições posteriores geradas pelo MCMC fornecem uma visão mais completa sobre a incerteza dos parâmetros, ao contrário da regressão linear tradicional, que fornece apenas estimativas pontuais e seus erros padrão.

(4) Implemente o algoritmo Expectation-Maximization (EM) para resolver o problema 1 (Censored data) no início do
video https://www.youtube.com/watch?v=93mVAAN0R5s
Teste o algoritmo para uma amostra aleatória de dados exponenciais censurados que você deve gerar com o método da função inversa (para gerar a variável aleatória exponencial).



O histograma mostra uma grande concentração de dados no ponto de censura (valor = 1), com uma distribuição decrescente para valores menores que 1, como esperado para dados exponenciais censurados.

Resultado da Estimativa de Lambda

A estimativa de λ pelo algoritmo EM foi 1.51719.

O valor verdadeiro de λ era 1.5. A estimativa obtida pelo algoritmo EM é muito próxima ao valor verdadeiro, indicando que o algoritmo funcionou corretamente para estimar o parâmetro a partir dos dados censurados. O algoritmo EM é eficaz para lidar com dados censurados e fornece boas estimativas dos parâmetros da distribuição subjacente.