



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

**Procedimentos de geração de dados de  
sobrevivência com censura à direita**

**por**

**Flávia Adriane Pestana de Oliveira**

Brasília, 29 de Outubro de 2021

# **Procedimentos de geração de dados de sobrevivência com censura à direita**

**por**

**Flávia Adriane Pestana de Oliveira**

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Eduardo Yoshio Nakano

Brasília, 29 de Outubro de 2021

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Prof. Dr. Eduardo Yoshio Nakano  
Orientador, EST/UnB

Prof. Dr. Helton Saulo Bezerra dos Santos  
EST/UnB

Prof. Dra. Juliana Cobre  
ICMC/USP

... *when pain is over, the remembrance of it often becomes a pleasure.*

(Jane Austen)

Ao meu noivo e minha família que me dão razão para viver.

# Agradecimentos

Em primeiro lugar, agradeço ao meu orientador Professor Dr. Eduardo Yoshio Nakano por todo o ensinamento passado e dedicação ao longo desse estudo. Também, direciono meus agradecimentos ao Professor Dr. Helton Saulo Bezerra dos Santos (Departamento de Estatística, Universidade de Brasília) e a Professora Dra. Juliana Cobre (Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo) que compõem a banca de defesa da dissertação. Em segundo lugar, agradeço ao meu noivo Guilherme Gomes e minha amiga Marina Oliveira que foram essências nessa jornada. Por último, agradeço a minha família e amigos que estiveram ao meu lado.

# Resumo

Este trabalho tem como objetivo apresentar metodologias para a geração de dados de sobrevivência com censura à direita. Foram apresentadas diversas metodologias para a geração de dados de sobrevivência contínuos e discretos segundo os mecanismos de censura do Tipo I, II e aleatório. Ademais, este trabalho também apresentou uma forma alternativa de gerar dados com censura aleatória, obtendo o indicador de censura por meio de uma distribuição de Bernoulli. Os resultados obtidos neste trabalho mostraram que o mecanismo de censura aleatória é o melhor para a geração de dados censurados à direita. Além disso, a metodologia de obtenção dos indicadores de censura por meio de uma distribuição Bernoulli se mostrou um método simples e eficaz de gerar dados com censura aleatória.

**Palavras-chave:** Dados censurados. Tempos de falha discretos. Simulação. Parâmetro de força-stress.

# Abstract

In this work, we present methodologies for the generation of right censored survival data. Several methodologies for the generation of continuous and discrete survival data according to Type I, II and random censoring mechanisms were presented. We also presented an alternative way to generate data with random censoring, obtaining the censoring indicator through a Bernoulli distribution. The results obtained in this work showed that the mechanism of random censoring is the best for generating right censored data. In addition, the methodology for obtaining censoring indicators through a Bernoulli distribution proved to be a simple and efficient method of generating data with random censoring.

**Keywords:** Censored data. Discrete time failure data. Simulation. Stress-strength parameter.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Revisão bibliográfica</b>	<b>4</b>
2.1	Conceitos básicos em Análise de Sobrevivência . . . . .	4
2.1.1	Tipos de censura . . . . .	5
2.1.2	Representação do tempo de sobrevivência . . . . .	7
2.1.3	Estimador de Kaplan-Meier . . . . .	9
2.1.4	Estimação via Máxima Verossimilhança . . . . .	11
2.1.5	Vício do estimador . . . . .	12
2.2	Parâmetro de força-stress . . . . .	14
2.2.1	Exponencial . . . . .	14
2.2.2	Weibull . . . . .	15
2.2.3	Log-Normal . . . . .	16
2.2.4	Birnbaum-Saunders . . . . .	18
<b>3</b>	<b>Geração de dados de sobrevivência com censura à direita</b>	<b>20</b>
3.1	Censura Tipo I . . . . .	21
3.2	Censura Tipo II . . . . .	22
3.3	Censura Aleatória . . . . .	23
3.3.1	Caso 1 - Distribuição de $C$ similar à $T$ . . . . .	24

3.3.2	Caso 2 - Alternativo numérico . . . . .	27
3.3.3	Caso 3 - Distribuição de $C$ diferente de $T$ . . . . .	29
3.3.4	Caso 4 - Indicador de Censura via Bernoulli . . . . .	30
<b>4</b>	<b>Geração de dados discretos de sobrevivência</b>	<b>32</b>
4.1	Censura Tipo I - Discreto . . . . .	33
4.2	Censura Tipo II - Discreto . . . . .	33
4.3	Censura Aleatória . . . . .	34
4.3.1	Caso 1 - Discreto . . . . .	34
4.3.2	Caso 1 - Discreto - Decisão de Censura via Bernoulli . . . . .	34
4.3.3	Caso 4 - Discreto - Indicador de Censura via Bernoulli . . . . .	35
<b>5</b>	<b>Resultados</b>	<b>36</b>
5.1	Simulações do Algoritmo 1 - Censura Tipo I . . . . .	36
5.2	Simulações do Algoritmo 2 - Censura Tipo II . . . . .	38
5.3	Simulações do Algoritmo 3 - Censura Aleatória: Caso 1 . . . . .	38
5.4	Simulações do Algoritmo 4 - Censura Aleatória: Caso 2 . . . . .	40
5.4.1	Escolha do $M$ e $\varepsilon$ . . . . .	40
5.4.2	Escolha do $\eta$ . . . . .	41
5.4.3	Simulações . . . . .	47
5.5	Simulações do Algoritmo 5 - Censura Aleatória: Caso 3 . . . . .	48
5.5.1	Escolha do $\eta$ . . . . .	49
5.5.2	Simulações . . . . .	49
5.6	Simulações do Algoritmo 6 - Censura Aleatória: Caso 4 . . . . .	52
5.7	Comparação dos métodos . . . . .	53
5.7.1	Comparação entre os Algoritmo 3 e Algoritmo 4 . . . . .	53
5.7.2	Vício dos parâmetros . . . . .	55
5.7.3	Vício da censura . . . . .	57

5.7.4	Distribuição da amostra dos dados censurados . . . . .	57
5.8	Comparação dos métodos - Discreto . . . . .	58
5.8.1	Vício dos parâmetros . . . . .	59
5.8.2	Vício da censura . . . . .	60
5.8.3	Distribuição da amostra dos dados discretos censurados . . . . .	61
<b>6</b>	<b>Síntese dos resultados</b>	<b>62</b>
<b>7</b>	<b>Considerações finais</b>	<b>64</b>
<b>A</b>	<b>Simulações do Algoritmo 1</b>	<b>68</b>
<b>B</b>	<b>Simulações do Algoritmo 2</b>	<b>70</b>
<b>C</b>	<b>Simulações do Algoritmo 3</b>	<b>72</b>
<b>D</b>	<b>Algoritmo 4 - Étas escolhidos</b>	<b>74</b>
<b>E</b>	<b>Simulações do Algoritmo 4</b>	<b>79</b>
<b>F</b>	<b>Algoritmo 5 - Étas escolhidos</b>	<b>81</b>
<b>G</b>	<b>Simulações do Algoritmo 6</b>	<b>88</b>
<b>H</b>	<b>Distribuição da amostra dos dados censurados</b>	<b>90</b>
<b>I</b>	<b>Distribuição da amostra dos dados discretos censurados</b>	<b>100</b>
	<b>Referências Bibliográficas</b>	<b>100</b>

# Lista de Tabelas

5.1	Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ), diferença absoluta entre a média do percentual de censura e o percentual de censura desejado ( $E(\hat{p}) - p$ ), erro quadrático médio ( $EQM(\hat{p})$ ) e média de tempo em segundos por simulação de 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim Exp(5)$ para diferentes interações de Monte Carlo ( $M$ ) e precisão ( $\varepsilon$ ) com $\eta = 0,5$ , $p = 0,1$ e $n = 100$ . . . . .	41
5.2	Valor de $\eta$ escolhido para gerar $\lambda_2$ da distribuição de censura $C \sim Exp(\lambda_2)$ a partir da distribuição de tempo $T \sim Exp(\lambda_1)$ para diferentes parâmetros a partir das 1000 simulações do Algoritmo 4 com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	43
5.3	Valor de $\eta$ escolhido para gerar $\lambda_2$ da distribuição de censura $C \sim Weibull(\alpha; \lambda_2)$ a partir da distribuição de tempo $T \sim Weibull(\alpha; \lambda_1)$ para diferentes parâmetros a partir das 1000 simulações do Algoritmo 4 com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	44
5.4	Valor de $\eta$ escolhido para gerar $\mu_2$ da distribuição de censura $C \sim Lognormal(\mu_2; \sigma^2)$ a partir da distribuição de tempo $T \sim Lognormal(\mu_1; \sigma^2)$ para diferentes parâmetros a partir das 1000 simulações do Algoritmo 4 com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	46
5.5	Valor de $\eta$ escolhido para gerar $\mu_2$ da distribuição de censura $C \sim BS(\alpha; \beta_2)$ a partir da distribuição de tempo $T \sim BS(\alpha; \beta_1)$ para diferentes parâmetros a partir das 1000 simulações do Algoritmo 4 com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	47

5.6	Valor de $\eta$ escolhido para gerar o parâmetro da distribuição de censura $C$ diferente da distribuição de tempo $T$ para diferentes parâmetros a partir das 1000 simulações do Algoritmo 5 com $M = 1000, \varepsilon = 0,001, p = 0,1$ e $n = 1000$ . . . . .	49
5.7	Vício absoluto ( $IB(\hat{\theta})$ ) dos parâmetros $\alpha$ e $\lambda$ em 1000 simulações da distribuição $Weibull(2; 5)$ para os Algoritmos 1 a 6 com amostra ( $n$ ) de tamanho igual a 10, 100 e 1000 e percentual ( $p$ ) de censura igual a 0,1, 0,3 e 0,5 . . . . .	55
5.8	Erro Quadrático Médio ( $EQM(\hat{\theta})$ ) dos parâmetros $\alpha$ e $\lambda$ em 1000 simulações da distribuição $Weibull(2; 5)$ para os Algoritmos 1 a 6 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5 . . . . .	56
5.9	Diferença absoluta entre a média do percentual de censura e o percentual de censura desejado ( $ E(\hat{p}) - p $ ) e Erro Quadrático Médio ( $EQM(\hat{p})$ ) do percentual de censura desejado ( $p$ ) em 1000 simulações da distribuição $Weibull(2; 5)$ para os Algoritmos 1 e 3 a 6 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5 . . . . .	57
5.10	Vício ( $IB(\hat{\theta})$ ) dos parâmetros $\beta$ e $q$ em 1000 simulações da distribuição $WeibullDiscreta(2; 0, 95)$ para os Algoritmos 7 a 11 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5 . . . . .	59
5.11	Erro Quadrático Médio ( $EQM(\hat{\theta})$ ) dos parâmetros $\beta$ e $q$ em 1000 simulações da distribuição $WeibullDiscreta(2; 0, 95)$ para os Algoritmos 7 a 11 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5 . . . . .	60
5.12	Diferença absoluta entre a média do percentual de censura e o percentual de censura desejado ( $ E(\hat{p}) - p $ ) e Erro Quadrático Médio ( $EQM(\hat{\theta})$ ) do percentual de censura desejado ( $p$ ) em 1000 simulações da distribuição $WeibullDiscreta(2; 0, 95)$ para os Algoritmos 7 e 9 a 11 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5 . . . . .	60
6.1	Síntese dos resultados da geração de dados, contínuos e discretos, de sobrevivência com censura à direita para diferentes tipos de censura - Algoritmos 1 a 11 . . . . .	63

A.1	Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Tipo I - Algoritmo 1 para as distribuições de tempo $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	69
B.1	Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Tipo II - Algoritmo 2 para as distribuições de tempo $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	71
C.1	Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Aleatória Caso 1; distribuição de $C$ similar à $T$ - Algoritmo 3 para as distribuições: Exponencial, Weibull e Log-Normal e , com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	73
E.1	Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Aleatória Caso 2, alternativo numérico (tentativa e erro), - Algoritmo 4 para as distribuições: Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	80
G.1	Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Aleatória Caso 4, indicador de Censura via Bernoulli - Algoritmo 6 para as distribuições de tempo $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	89

# Lista de Figuras

5.1	Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Tipo I - Algoritmo 1 para as distribuições de tempo $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	37
5.2	Média do percentual de censura em 1000 simulações de Censura Tipo II - Algoritmo 2 para as distribuições de tempo $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	38
5.3	Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 1, distribuição de $C$ similar à $T$ - Algoritmo 3 para as distribuições: Exponencial, Weibull e Log-Normal e , com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	39
5.4	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim Exp(5)$ para diferentes $\eta$ com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	42
5.5	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim Weibull(2; 5)$ para diferentes $\eta$ com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	44
5.6	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim Lognormal(0, 5; 1, 5)$ para diferentes $\eta$ com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	45

5.7	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim BS(1, 5; 3, 5)$ para diferentes $\eta$ com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	46
5.8	Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 2, alternativo numérico (tentativa e erro), - Algoritmo 4 para as distribuições: Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	48
5.9	Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 para a distribuição de tempo $T \sim Weibull(2; 5)$ para distribuição de censura, $C$ , Exponencial, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	50
5.10	Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 para a distribuição de tempo $T \sim Lognormal(0, 5; 1, 5)$ para as distribuições de censura, $C$ , Exponencial e Weibull, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	50
5.11	Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 para a distribuição de tempo $T \sim BS(1, 5; 3, 5)$ para as distribuições de censura, $C$ , Exponencial, Weibull e Log-Normal, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	51
5.12	Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 4, indicador de Censura via Bernoulli - Algoritmo 6 para as distribuições de tempo $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 . . . . .	52
5.13	Comparação da média do percentual de censura em 1000 simulações dos Algoritmos 3 e Algoritmo 4 para $\varepsilon$ igual a 0,1, 0,05, 0,01 e 0,001 para as distribuições $Weibull(2; 5)$ com diferentes tamanhos de amostra . . . . .	54



5.14	Comparação do log da média de tempo da simulação dos Algoritmos 3 e Algoritmo 4 para $\varepsilon$ igual a 0,1, 0,05, 0,01 e 0,001 para a distribuição $Weibull(2; 5)$ com diferentes tamanhos de amostra . . . . .	55
D.1	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim Exp$ para diferentes parâmetros e $\eta$ com $M = 1000, \varepsilon = 0,001, p = 0,1$ e $n = 1000$ . . .	75
D.2	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim Weibull$ para diferentes parâmetros e $\eta$ com $M = 1000, \varepsilon = 0,001, p = 0,1$ e $n = 1000$	76
D.3	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim Lognormal$ para diferentes parâmetros e $\eta$ com $M = 1000, \varepsilon = 0,001, p = 0,1$ e $n = 1000$	77
D.4	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo $T \sim BS$ para diferentes parâmetros e $\eta$ com $M = 1000, \varepsilon = 0,001, p = 0,1$ e $n = 1000$ . . .	78
F.1	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 da distribuição de tempo $T \sim Weibull(2; 5)$ para distribuição de censura, $C$ , Exponencial, com diferentes parâmetros e $\eta$ com $M = 1000, \varepsilon = 0,001, p = 0,1$ e $n = 1000$ . . . . .	82
F.2	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 da distribuição de tempo $T \sim Lognormal(0, 5; 1, 5)$ para distribuição de censura, $C$ , Exponencial, com diferentes parâmetros e $\eta$ com $M = 1000, \varepsilon = 0,001, p = 0,1$ e $n = 1000$ . . . . .	83

F.3	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 da distribuição de tempo $T \sim BS(1, 5; 3, 5)$ para distribuição de censura, $C$ , Exponencial, com diferentes parâmetros e $\eta$ com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	84
F.4	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 da distribuição de tempo $T \sim Lognormal(0, 5; 1, 5)$ para distribuição de censura, $C$ , Weibull, com diferentes parâmetros e $\eta$ com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	85
F.5	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 da distribuição de tempo $T \sim BS(1, 5; 3, 5)$ para distribuição de censura, $C$ , Weibull, com diferentes parâmetros e $\eta$ com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	86
F.6	Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de $C$ diferente de $T$ - Algoritmo 5 da distribuição de tempo $T \sim BS(1, 5; 3, 5)$ para distribuição de censura, $C$ , Log-Normal, com diferentes parâmetros e $\eta$ com $M = 1000$ , $\varepsilon = 0,001$ , $p = 0,1$ e $n = 1000$ . . . . .	87
H.1	Ajuste do modelo $Weibull(2; 5)$ a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 10 e percentual de censura de 10% . . . . .	91
H.2	Ajuste do modelo $Weibull(2; 5)$ a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 10 e percentual de censura de 30% . . . . .	92
H.3	Ajuste do modelo $Weibull(2; 5)$ a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 10 e percentual de censura de 50% . . . . .	93
H.4	Ajuste do modelo $Weibull(2; 5)$ a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 100 e percentual de censura de 10% . . . . .	94
H.5	Ajuste do modelo $Weibull(2; 5)$ a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 100 e percentual de censura de 30% . . . . .	95

H.6	Ajuste do modelo <i>Weibull</i> (2; 5) a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 100 e percentual de censura de 50% . . . . .	96
H.7	Ajuste do modelo <i>Weibull</i> (2; 5) a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 1000 e percentual de censura de 10% . . . . .	97
H.8	Ajuste do modelo <i>Weibull</i> (2; 5) a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 1000 e percentual de censura de 30% . . . . .	98
H.9	Ajuste do modelo <i>Weibull</i> (2; 5) a Função de Sobrevida estimada para os Algoritmos 1 a 6, amostra de tamanho 1000 e percentual de censura de 50% . . . . .	99
I.1	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 10 e percentual de censura de 10% . . . . .	101
I.2	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 10 e percentual de censura de 30% . . . . .	102
I.3	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 10 e percentual de censura de 50% . . . . .	103
I.4	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 100 e percentual de censura de 10% . . . . .	104
I.5	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 100 e percentual de censura de 30% . . . . .	105
I.6	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 100 e percentual de censura de 50% . . . . .	106
I.7	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 1000 e percentual de censura de 10% . . . . .	107
I.8	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 1000 e percentual de censura de 30% . . . . .	108
I.9	Ajuste do modelo <i>WeibullDiscreta</i> (2; 0, 95) a Função de Sobrevida estimada para os Algoritmos 7 a 11, amostra de tamanho 1000 e percentual de censura de 50% . . . . .	109

# Lista de Algoritmos

1	Para gerar amostras com $p \times 100\%$ de censura do Tipo I . . . . .	22
2	Para gerar amostras com $p \times 100\%$ de censura do Tipo II . . . . .	23
3	Para gerar amostras com censura Aleatória Caso 1 . . . . .	27
4	Para gerar amostras com censura Aleatória Caso 2 (Para distribuição Birnbaum-Saunders)	28
5	Para gerar amostras com censura Aleatória Caso 3 (Tempos da distribuição de interesse com censura Exponencial) . . . . .	30
6	Para gerar amostras com censura Aleatória Caso 4 . . . . .	31
7	Para gerar amostras de tempos discretos com $p \times 100\%$ de censura do Tipo I . . . . .	33
8	Para gerar amostras de tempos discretos com $p \times 100\%$ de censura do Tipo II . . . . .	33
9	Para gerar amostras de tempos discretos com censura Aleatória Caso 1 . . . . .	34
10	Para gerar amostras de tempos discretos com censura Aleatória Caso 1 - Decisão de Censura via Bernoulli . . . . .	35
11	Para gerar amostras de tempos discretos com censura Aleatória Caso 4 . . . . .	35

# Capítulo 1

## Introdução

Em Análise de Sobrevivência, o tempo de vida (ou tempo até a falha) é caracterizado pelo tempo até a ocorrência de um evento de interesse. Segundo Colosimo e Giolo, 2006, a principal característica em dados de sobrevivência é a presença de censura nos dados, que ocorre por existir algum fator que impossibilite a observação exata do dado. O mecanismo de censura à direita, em geral, é a situação mais comum e a mesma pode ser subdividida em três tipos: Tipo I, Tipo II e Aleatória.

Este trabalho focará na geração de dados de sobrevivência com censura à direita. As censuras dos Tipos I e II podem ser realizadas facilmente, ao gerar as observações segundo a distribuição de interesse e ao limitar o tempo de duração (pré-definido) do experimento ou o número de observações não censuradas, respectivamente. O mecanismo de censura aleatória é caracterizado por considerar uma variável aleatória, tempo até a ocorrência do evento de interesse, representado por  $T$ , e também uma variável aleatória  $C$ , independente de  $T$ , sendo o tempo até a censura. A censura ocorre quando  $C < T$ .

Assim, a maior dificuldade na geração de dados com censura aleatória se deve ao fato de que nem sempre é possível obter analiticamente o valor de  $R = P(C < X)$ . A quantidade  $R$  é também conhecida como parâmetro de força-stress e, no contexto de geração de dados censurados, ele pode ser utilizado para definir os parâmetros da distribuição de  $C$  de forma

obter o percentual de censura desejado na amostra.

Apesar dos dados com censuras dos Tipos I e II serem mais simples de gerar, nem sempre a escolha de um desses mecanismos de censura pode ser vista como a melhor opção. De fato, o pesquisador deve escolher o melhor tipo de censura ao gerar os seus dados levando em conta características que ele deseja em sua amostra simulada, por exemplo: menor (ou ausência de) vício, maior controle do percentual de censura desejado, concordância entre a distribuição de interesse e distribuição dos dados gerados, etc.

Neste contexto, o objetivo principal deste trabalho é apresentar os métodos usuais e propor alguns métodos alternativos de geração de dados com censura à direita e realizar um estudo que auxilie o pesquisador no sentido de indicar qual método é o mais adequado, em determinada situação, ao simular dados censurados. Os métodos propostos são a Censura Aleatória Alternativo numérico (parâmetro força-stress obtido analiticamente), Censura Aleatória com a distribuição de  $C$  diferente de  $T$  (obtido analiticamente), Censura Aleatória cujo o indicador de censura é obtido por meio da distribuição Bernoulli. Além disso, são utilizados alguns desses métodos propostos para o caso de dados discretos e é proposto o método da Censura Aleatória cuja a decisão de censura é tomada via a distribuição Bernoulli.

O desenvolvimento dos algoritmos e resultados foram realizados na versão 4.0.0 do software R<sup>1</sup> por meio do ambiente de desenvolvimento integrado R-Studio. O sistema operacional utilizado foi o Windows 10, processador AMD Ryzen 5 3400G com Radeon Vega Graphics (3.70 Ghz), memória instalada (RAM) de 32,0 GB, tipo de sistema operacional de 64 bits.

No Capítulo 2 é apresentada a revisão bibliográfica dos conceitos básicos em Análise de Sobrevivência e do parâmetro força-stress para distribuições, de interesse desse estudo, Exponencial, Weibull, Log-Normal e Birnbaum-Saunders. Em seguida, no Capítulo 3 é apresentado os métodos usuais para gerar dados de sobrevivência com censura à direita dos Tipo I, Tipo II e Aleatória, além disso, são propostos alguns métodos alternativos para a geração de dados com censura aleatória que dispensa a necessidade do cálculo do parâmetro de força-stress. Na

---

<sup>1</sup>The R Project for Statistical Computing - Software.

sequência, o Capítulo 4 estende alguns dos procedimentos apresentados no Capítulo 3 para a geração de dados discretos de sobrevivência com censura à direita. No Capítulo 5 é realizada a simulação e comparação do desempenho dos métodos usuais e alternativos de geração de dados (contínuos e discretos) de sobrevivência com censura à direita, por meio, da análise do vício das estimativas, percentual de censura obtido e ajuste do modelo. Por último, no Capítulo 6 é realizada uma síntese dos resultados obtidos nesse estudo.

# Capítulo 2

## Revisão bibliográfica

Neste capítulo, será realizada a revisão bibliográfica dos tópicos utilizados nos procedimentos de geração de dados de sobrevivência com censura à direita. Na Seção 2.1 é feita a revisão dos conceitos básicos em Análise de Sobrevivência, de tipos de censura, representação do tempo de sobrevivência, estimador de Kaplan-Meier, estimação via Máxima Verossimilhança e vício do estimador. Por fim, na Seção 2.2 é apresentado o parâmetro de força-stress para as distribuições de interesse desse estudo.

### 2.1 Conceitos básicos em Análise de Sobrevivência

A Análise de Sobrevivência consiste em utilizar um conjunto de técnicas estatísticas, na análise do tempo até a falha (ou evento de interesse). O tempo de falha é caracterizado pelo tempo inicial, a escala de medida e o evento de interesse (falha).

Além disso, outro componente importante do conjunto de dados de sobrevivência é a censura e esse elemento surge em decorrência da observação parcial da resposta. Isto é, nos estudos ou experimentos, devido a variável resposta ser o tempo de falha e como diversos fatores podem interferir na observação do evento de interesse a partir do tempo inicial, quase sempre há a presença de dados censurados.



Neste capítulo, será realizada a revisão dos tópicos básicos em Análise de Sobrevida. Na Subseção 2.1.1 é realizada a revisão dos tipos de censura. Em seguida, na Subseção 2.1.2 é apresentada algumas representações do tempo de sobrevida para entender a técnica estatística da análise. Por fim, nas Subseções 2.1.3 a 2.1.5 são apresentadas técnicas de estimação do parâmetros e da sobrevida na presença de censura e vício do estimador.

### **2.1.1 Tipos de censura**

O principal conceito na Análise de Sobrevida é o tempo de falha. Esse conceito caracteriza-se em observar, a partir de um ponto inicial, o tempo até ocorrer o evento de interesse no indivíduo ou objeto.

Em geral, a escala de medida desse tempo é em dias, horas ou minutos (tempo de relógio), entretanto outras escalas podem ser utilizadas como, por exemplo, quilometragem ou número de ciclos. O evento de interesse são diversos: o tempo até a morte de um paciente; o número de sessões de tratamento até a cura de uma doença; o tempo até uma lâmpada falhar; o tempo até ocorrer um sinistro.

Nesse tipo de experimento, quase sempre não é possível obter a resposta completa do tempo de falha, que resulta na censura da informação. Segundo Colosimo e Giolo, 2006, se não houver censuras nos dados é possível, após algumas transformações na variável resposta, utilizar técnicas estatísticas clássicas, como regressão ou planejamento de experimento, na análise.

Entretanto, omitir os dados a respeito da censura gera estimativas viesadas, dessa forma, utilizar as técnicas de análise de sobrevida é fundamental nesse tipo de estudo. As causas para censurar o tempo pode ocorrer por diversas razões, como o paciente decidir não ir até o fim da pesquisa, seja porque mudou ou perdeu o interesse, também pode ocorrer a morte ou falha do paciente e/ou observação, por um fator alheio ao estudo. Ainda, em um experimento em que o número de falhas é pré-determinado, ao atingir a meta, o restante dos objetos não são mais acompanhados.

A censura é classificada em três tipos, de acordo com a forma que ocorre: à direita, à esquerda e intervalar. No caso da censura à direita, o tempo de ocorrência do evento de interesse do indivíduo sempre vai estar à direita do tempo registrado (tempo censurado), ou seja, o evento de interesse não é observado até o fim do estudo ou experimento.

Na censura à esquerda, ao iniciar o estudo ou experimento, o indivíduo já experimentou o evento de interesse, dessa forma, o tempo de ocorrência está à esquerda tempo registrado. Outro tipo de censura que ocorre é a intervalar. Nesta, o indivíduo é observado em intervalos de tempo e em um desses momentos o evento de interesse acontece. Assim, o tempo de falha dessa observação será um intervalo de tempo entre a última e penúltima visita.

De acordo com Lindsey e Ryan, 1998, a censura à direita e à esquerda podem ser vistas como casos particulares da censura intervalar. Dado que o tempo de falha, na censura intervalar, pertence ao intervalo  $(L, U]$ , na censura à direita esse intervalo será  $(L, \infty)$  e na censura à esquerda  $(0, U]$ , isto é, na censura à direita não se sabe em que momento o evento de interesse será observado depois do momento  $L$  e na censura à esquerda o evento de interesse ocorreu antes do momento  $U$ .

Nesse trabalho, o foco será na censura à direita. Por conseguinte, os três principais tipos de censura à direita são (Colosimo e Giolo, 2006):

- Censura do Tipo I - O período de estudo é pré-fixado, portanto são censurados os indivíduos que não experimentaram o evento de interesse até esse tempo. Nesta situação, o percentual de censura observado é aleatório e o tempo máximo do estudo é pré-fixado pelo pesquisador;
- Censura do Tipo II - A quantidade de observações censuradas (ou não censuradas) é pré-fixada, portanto são censurados os indivíduos que não experimentaram o evento de interesse até a ocorrência da quantidade de observações pré-fixada;
- Censura aleatória - É o tipo de censura mais geral e engloba os do Tipo I e II como casos particulares. Ocorre no caso de um ou mais indivíduos não serem acompanhados até

o final do estudo, ou ainda, se esses experimentaram o evento de interesse por motivos distintos do interesse do estudo (por exemplo, o desfecho é a morte por uma doença e o paciente morre em um acidente de trânsito). A censura aleatória ocorre de forma natural e aleatória, sem a interferência do pesquisador.

### 2.1.2 Representação do tempo de sobrevivência

O tempo até a falha é descrito pela variável, geralmente contínua, não negativa  $T$ . O comportamento da variável  $T$  é caracterizado, por meio, de várias funções matemáticas. Na análise de sobrevivência, a função densidade de probabilidade, a função de sobrevivência, a função de taxa de falha (ou risco) e funções derivadas destas são bastante utilizadas.

#### Função densidade de probabilidade

A função densidade de probabilidade de  $T$ ,  $f(t)$ , é a probabilidade de um indivíduo ou objeto experimentar o evento de interesse no intervalo de tempo  $[t, t + \Delta t)$ , dividido pelo tamanho do intervalo ( $\Delta t$ ). Assim,

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}, \quad t > 0. \quad (2.1)$$

A função  $f(t)$  satisfaz as seguintes condições (Meyer, 1983):

- i)  $f(t) \geq 0$ , tal que  $t \geq 0$ .
- ii)  $\int_0^{\infty} f(t) dt = 1$ .
- iii)  $P(a \leq T \leq b) = \int_a^b f(t) dt$ , tal que  $0 \leq a < b$ .

#### Função de Sobrevivência

A função de sobrevivência,  $S(t)$ , é uma das principais funções probabilísticas para representar o tempo de sobrevivência. A  $S(t)$  representa a probabilidade de uma observação não experimentar

o evento de interesse até um certo tempo  $t$ . A  $f(t)$  pode ser usada para definir a probabilidade do indivíduo sobreviver além do tempo  $t$ , desse modo, a função de sobrevivência pode ser obtida da seguinte forma:

$$S(t) = P[T > t] = \int_t^{\infty} f(t)dt. \quad (2.2)$$

A função de sobrevivência é uma função não crescente, absolutamente contínua, tal que  $\lim_{t \rightarrow 0} S(t) = 1$  e  $\lim_{t \rightarrow \infty} S(t) = 0$ .

### Função de Taxa de Falha (ou Risco)

A função de taxa de falha (ou risco),  $h(t)$ , pode ser vista como o risco instantâneo do indivíduo falhar ou morrer em um determinado tempo  $t$ . A função é expressa por:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t}, \quad t \geq 0. \quad (2.3)$$

Dessa forma,  $h(t)$  é a probabilidade de experimentar o evento de interesse no intervalo  $[t, t + \Delta t)$  dado que o indivíduo sobreviveu até o tempo  $t$ . Portanto,  $h(t)$  só assume valores reais positivos e não é limitada superiormente (Lawless, 2003).

### Relação entre $f(t)$ , $S(t)$ e $h(t)$

As funções apresentadas anteriormente para representação do tempo de sobrevivência são relacionadas matematicamente. A Taxa de Falha,  $h(t)$ , pode ser obtida a partir de  $f(t)$  e  $S(t)$ :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \cap T \geq t)}{\Delta t P(T \geq t)} = \quad (2.4)$$

$$= \frac{\lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}}{P(T > t)} = \frac{f(t)}{S(t)}. \quad (2.5)$$

A função densidade de probabilidade,  $f(t)$ , pode ser obtida a partir da derivada da função

de distribuição acumulada,  $F(t)$ :

$$f(t) = \frac{d}{dt}F(t), \quad (2.6)$$

e  $F(t) = 1 - S(t)$ , assim:

$$f(t) = \frac{d}{dt}[1 - S(t)] = -\frac{d}{dt}S(t) = -S'(t), \quad (2.7)$$

e, portanto

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log S(t). \quad (2.8)$$

Ao integrar ambos os termos:

$$\log S(t) = -\int_0^t h(u)du = -H(t), \quad (2.9)$$

o que implica que

$$S(t) = \exp \left\{ -\int_0^t h(u)du \right\} = \exp \{-H(t)\}. \quad (2.10)$$

Além disso,

$$f(t) = h(t)S(t) = h(t) \exp \left\{ -\int_0^t h(u)du \right\}. \quad (2.11)$$

### 2.1.3 Estimador de Kaplan-Meier

As representações do tempo de sobrevivência apresentadas na Subseção 2.1.2 não consideram a presença de censura nos cálculos. Porém, na prática o conjunto de dados de tempo de falhas podem apresentar censuras, assim é necessário utilizar técnicas estatísticas que considerem a presença de censuras.

Proposto por Kaplan e Meier, 1958, o estimador não-paramétrico de Kaplan-Meier é um

dos mais utilizadas para estimar a função de sobrevivência. Esse estimador considera o número de falhas em determinado tempo e o número de observações que não falharam ou não foram censurados nesse mesmo tempo.

Pressuponha que uma amostra de tamanho  $n$  e  $r$  falhas distintas ( $r < n$ ) nos tempos  $t_1 < t_2 < \dots < t_r$ ,  $S(t)$  será uma função de probabilidade nos tempos de falha  $t_j$  ( $j = 1, \dots, r$ ), assim:

$$S(t) = (1 - q_1)(1 - q_2) \cdots (1 - q_j), \quad (2.12)$$

sendo  $q_j$  a probabilidade de um indivíduo falhar no intervalo  $[t_{j-1}, t_j)$  dado que não falhou até o momento  $t_{j-1}$  (considere  $t_0 = 0$ ), ou seja,  $q_j = P(T \in [t_{j-1}, t_j) | T \geq t_{j-1})$ . A estimativa de  $q_j$  será:

$$\hat{q}_j = \frac{\text{número de falhas em } t_j}{\text{número de observações sob risco em } t_{j-1}}, \quad j = 1, \dots, r. \quad (2.13)$$

Por conseguinte, a estimativa da  $S(t)$  é escrita em termos de probabilidades condicionais. Desse modo, o estimador não-paramétrico de Kaplan-Meier é expresso da seguinte forma:

$$\hat{S}_{KM}(t) = \prod_{j:t_j \leq t} (1 - q_j). \quad (2.14)$$

Considerar  $t_1 < t_2 < \dots < t_r$  os  $r$  tempos ordenados e distintos de falha e  $q_j = \frac{d_j}{n_j}$ . Sendo  $d_j$  o número de falhas em  $t_j$  e  $n_j$  o número de elementos da amostra sob risco em  $t_j$ , seja porque não falharam ou não foram censurados até o momento antes de  $t_j$ .

As principais propriedades do estimador de Kaplan-Meier são (Colosimo e Giolo, 2006):

- i) é não viciado para amostras grandes;
- ii) é fracamente consistente (Breslow e Crowley, 1974) (Meier, 1975);
- iii) converge assintoticamente para um processo gaussiano (Breslow e Crowley, 1974) (Meier, 1975);

iv) é estimador de máxima verossimilhança de  $S(t)$  (Kaplan e Meier, 1958).

#### 2.1.4 Estimação via Máxima Verossimilhança

O método de máxima verossimilhança trabalha a amostra como uma distribuição de probabilidade, de forma que, os resultados são analisados na perspectiva de qual distribuição é mais provável de ter gerado aqueles dados.

Em termos matemáticos, para se obter estimadores para os parâmetros considera-se uma amostra aleatória (independente e identicamente distribuída),  $t_1, \dots, t_n$  sem censuras e a população é caracterizada pela função de densidade (ou probabilidade),  $f(t)$ , de uma distribuição. A função de verossimilhança para um parâmetro  $\theta$  desta população será:

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta), \quad (2.15)$$

o objetivo é, portanto, encontrar em qual distribuição o valor  $\theta$  maximiza a função  $L(\theta)$ , isto é, maximize a probabilidade da amostra observada ocorrer.

Na presença de censuras à direita, um tempo censurado contribui para a função de verossimilhança com a função de sobrevivência. Além disso, considerar (sem perda de generalidade) que as  $r$  primeiras observações da amostra são não-censuradas e as  $(n - r)$  seguintes censuradas, tem-se que a função de verossimilhança será:

$$L(\theta) \propto \prod_{i=1}^r f(t_i; \theta) \prod_{i=r+1}^n S(t_i; \theta), \quad (2.16)$$

em que  $S(\cdot)$  é a função de sobrevivência. É possível reescrever a função de verossimilhança ao incluir uma variável indicadora  $\delta_i$ , em que:

$$\delta_i = \begin{cases} 0, & \text{se } t_i \text{ foi censurado} \\ 1, & \text{se } t_i \text{ foi observado} \end{cases}, \quad i = 1, \dots, n, \quad (2.17)$$

a amostra será baseada em  $(t_1, \delta_1), \dots, (t_n, \delta_n)$  com os  $t_i$ 's independentes e da mesma distribuição de probabilidades com parâmetro  $\theta$ . Sendo assim, a função de verossimilhança pode ser escrita como:

$$L(\theta) \propto \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i}. \quad (2.18)$$

De modo que, ao encontrar o  $\hat{\theta}$  que maximiza a função  $L(\theta)$  esse será o Estimador de Máxima Verossimilhança (EMV) de  $\theta$  (Colosimo e Giolo, 2006) (Cox e Oakes, 1984).

### 2.1.5 Vício do estimador

Uma questão importante ao se estimar um parâmetro é avaliar o erro médio deste, visto que um estimador razoável irá apresentar um valor de erro menor. Existem algumas técnicas estatísticas para avaliar e comparar os estimadores de um parâmetro, entre eles, o Erro Absoluto Médio e o Erro Quadrático Médio.

#### Erro Absoluto Médio

Uma função crescente da distância absoluta  $|\hat{\theta} - \theta|$ , quase sempre, serve para mensurar a eficiência de um estimador. O Erro Absoluto Médio (EAM) é um desses casos, representado a seguir (Casella e Berger, 2010):

$$EAM(\hat{\theta}) = E(|\hat{\theta} - \theta|). \quad (2.19)$$

#### Erro Quadrático Médio

O Erro Quadrático Médio (EQM) de uma estimador  $\hat{\theta}$  de um parâmetro  $\theta$  mede a diferença quadrática média entre o estimador e o parâmetro, cuja medida tem desempenho razoável para



um estimador pontual. O cálculo do EQM é dada por (Casella e Berger, 2010):

$$EQM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2. \quad (2.20)$$

Como o vício do estimador  $\hat{\theta}$  é denominado  $B(\hat{\theta}) = (E(\hat{\theta}) - \theta)$ . Portanto, EQM de  $\hat{\theta}$  é:

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + (B(\hat{\theta}))^2. \quad (2.21)$$

Note que, se  $E(\hat{\theta}) = \theta$ , o estimador  $\hat{\theta}$  é não viciado para  $\theta$  e assim,  $B(\hat{\theta}) = 0$ . Nesse caso, o EQM de um estimador não viesado será:

$$EQM(\hat{\theta}) = Var(\hat{\theta}). \quad (2.22)$$

O EQM possui algumas vantagens em relação a outras medidas de distância, pois incorpora dois componentes. Sendo um, a medida de variabilidade do estimador (precisão) e o outro a medida do viés (exatidão). Portanto, um bom estimador nas propriedades de EQM terá a variância e o viés controlado (Casella e Berger, 2010).

Cabe, também, mencionar que frequentemente o EQM é utilizado para comparar estimativas. O estimador  $\hat{\theta}_1$  será melhor que  $\hat{\theta}_2$  para todo  $\theta$  se

$$EQM(\hat{\theta}_1) \leq EQM(\hat{\theta}_2). \quad (2.23)$$

Se existir um estimador  $\hat{\theta}^*$  tal que para todo  $\hat{\theta}$  de  $\theta$  for

$$EQM(\hat{\theta}^*) \leq EQM(\hat{\theta}), \quad (2.24)$$

então,  $\hat{\theta}^*$  é um estimador ótimo para  $\theta$ . Além disso, se  $\hat{\theta}^*$  for não viciado, logo  $\hat{\theta}^*$  é o estimador

não viciado de variância uniformemente mínima para todo  $\theta$  se (Bolfarine e Sandoval, 2001).

$$Var(\hat{\theta}^*) \leq Var(\hat{\theta}). \quad (2.25)$$

## 2.2 Parâmetro de força-stress

O parâmetro de força-stress consiste em  $T$ , uma variável aleatória que representa o tempo de falha (stress), e  $C$ , uma variável aleatória, a censura do objeto do experimento (força). De forma que  $T$  está submetido a  $C$ , em outras palavras, a falha irá ocorrer se a força for menor que o stress (ou, equivalentemente, a censura irá ocorrer quando o stress for menor que a força). A força-stress é dada pela confiabilidade calculada por

$$R = P(C < T), \quad (2.26)$$

sendo que  $T$  e  $C$  são independentes ao se considerar um mecanismo de censura aleatório. A seguir, nas Subseções 2.2.1 a 2.2.4, são apresentados os cálculos para obter o parâmetro força-stress com as distribuições de interesse desse estudo.

### 2.2.1 Exponencial

Uma variável aleatória positiva  $T$  com distribuição Exponencial tem a função densidade de probabilidade:

$$f_T(t) = \frac{1}{\lambda} e^{-\frac{t}{\lambda}}, \quad t \geq 0, \quad \lambda > 0, \quad (2.27)$$

e Função de Sobrevivência e Taxa de Falha

$$S(t) = e^{-\frac{t}{\lambda}}, \quad t \geq 0, \quad (2.28)$$

$$h(t) = \frac{1}{\lambda}, \quad t \geq 0. \quad (2.29)$$

Suponha, as variáveis aleatórias independentes  $T$  (tempo de falha) e  $C$  (censura), sendo  $T \sim Exp(\lambda_1)$  e  $C \sim Exp(\lambda_2)$ . O parâmetro força-stress,  $R = P(C < T)$ , é obtido da seguinte maneira (Kim, Moon e Kang, 2003):

$$\begin{aligned}
 R = P(C < T) &= \int_0^{\infty} \int_0^t \frac{1}{\lambda_1} e^{-\frac{t}{\lambda_1}} \frac{1}{\lambda_2} e^{-\frac{c}{\lambda_2}} dc dt = \int_0^{\infty} \left( \int_0^t \frac{1}{\lambda_2} e^{-\frac{c}{\lambda_2}} dc \right) \frac{1}{\lambda_1} e^{-\frac{t}{\lambda_1}} dt = \\
 &= \int_0^{\infty} F_C(t) \frac{1}{\lambda_1} e^{-\frac{t}{\lambda_1}} dt = \int_0^{\infty} (1 - e^{-\frac{t}{\lambda_2}}) \frac{1}{\lambda_1} e^{-\frac{t}{\lambda_1}} dt = \int_0^{\infty} \frac{1}{\lambda_1} e^{-\frac{t}{\lambda_1}} dt - \int_0^{\infty} \frac{1}{\lambda_1} e^{-\frac{t}{\lambda_1}} e^{-\frac{t}{\lambda_2}} dt = \\
 &= 1 - \frac{\lambda_1^{-1}}{\lambda_1^{-1} + \lambda_2^{-1}} = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \tag{2.30}
 \end{aligned}$$

### 2.2.2 Weibull

Uma variável aleatória  $T$  com distribuição Weibull tem a função densidade de probabilidade igual a:

$$f_T(t) = \frac{\alpha}{\lambda} \left( \frac{t}{\lambda} \right)^{\alpha-1} \exp \left\{ - \left( \frac{t}{\lambda} \right)^{\alpha} \right\}, \quad t > 0, \lambda > 0, \alpha > 0, \tag{2.31}$$

e a Função de Sobrevivência e Taxa de Falha

$$S(t) = \exp \left\{ - \left( \frac{t}{\lambda} \right)^{\alpha} \right\}, \quad t > 0, \tag{2.32}$$

$$h(t) = \frac{\alpha}{\lambda} \left( \frac{t}{\lambda} \right)^{\alpha-1}, \quad t > 0. \tag{2.33}$$

Nesse contexto, leve em conta que o tempo de falha  $T$  e a censura  $C$  são variáveis aleatórias independentes com distribuição de Weibull,  $T \sim Weibull(\alpha, \lambda_1)$  e  $C \sim Weibull(\alpha, \lambda_2)$ , assim, o parâmetro força-stress será (Amiri et al., 2013):

$$R = P(C < T) = \int_0^{\infty} \int_0^t f_T(t) f_C(c) dc dt =$$

$$\begin{aligned}
&= \int_0^{\infty} \int_0^t \frac{\alpha}{\lambda_1} \left(\frac{t}{\lambda_1}\right)^{\alpha-1} \exp\left\{-\left(\frac{t}{\lambda_1}\right)^{\alpha}\right\} \frac{\alpha}{\lambda_2} \left(\frac{c}{\lambda_2}\right)^{\alpha-1} \exp\left\{-\left(\frac{c}{\lambda_2}\right)^{\alpha}\right\} dc dt = \\
&= \int_0^{\infty} \frac{\alpha}{\lambda_1} \left(\frac{t}{\lambda_1}\right)^{\alpha-1} \exp\left\{-\left(\frac{t}{\lambda_1}\right)^{\alpha}\right\} \left(\int_0^t \frac{\alpha}{\lambda_2} \left(\frac{c}{\lambda_2}\right)^{\alpha-1} \exp\left\{-\left(\frac{c}{\lambda_2}\right)^{\alpha}\right\} dc\right) dt = \\
&= \int_0^{\infty} \frac{\alpha}{\lambda_1^{1+\alpha-1}} t^{\alpha-1} \exp\left\{-\left(\frac{t}{\lambda_1}\right)^{\alpha}\right\} (F_C(t)) dt = \\
&= \frac{\alpha}{\lambda_1^{\alpha}} \int_0^{\infty} t^{\alpha-1} \exp\left\{-\left(\frac{t}{\lambda_1}\right)^{\alpha}\right\} \left(1 - \exp\left\{-\left(\frac{t}{\lambda_2}\right)^{\alpha}\right\}\right) dt = \\
&= \frac{\alpha}{\lambda_1^{\alpha}} \left(\int_0^{\infty} t^{\alpha-1} \exp\left\{-\left(\frac{t}{\lambda_1}\right)^{\alpha}\right\} dt - \int_0^{\infty} t^{\alpha-1} \exp\left\{-\left(\frac{t}{\lambda_1}\right)^{\alpha}\right\} \exp\left\{-\left(\frac{t}{\lambda_2}\right)^{\alpha}\right\} dt\right) = \\
&= \frac{\alpha}{\lambda_1^{\alpha}} \left(\frac{1}{\alpha} \frac{1}{\lambda_1^{-\alpha}} - \frac{1}{\alpha} \frac{1}{(\lambda_1^{-\alpha} \lambda_2^{-\alpha})}\right) = \left(\frac{\alpha}{\lambda_1^{\alpha}} \frac{\lambda_1^{\alpha}}{\alpha} - \frac{\alpha \lambda_1^{-\alpha}}{\alpha(\lambda_1^{-\alpha} \lambda_2^{-\alpha})}\right) = \\
&= 1 - \frac{\lambda_1^{-\alpha}}{(\lambda_1^{-\alpha} + \lambda_2^{-\alpha})} = \frac{\lambda_1^{-\alpha} + \lambda_2^{-\alpha} - \lambda_1^{-\alpha}}{(\lambda_1^{-\alpha} + \lambda_2^{-\alpha})} = \\
&= \frac{1}{\lambda_2^{\alpha}} \frac{\lambda_1^{\alpha} \lambda_2^{\alpha}}{(\lambda_1^{\alpha} + \lambda_2^{\alpha})} = \frac{\lambda_1^{\alpha}}{(\lambda_1^{\alpha} + \lambda_2^{\alpha})}. \tag{2.34}
\end{aligned}$$

### 2.2.3 Log-Normal

Uma variável aleatória positiva  $T$  tem a distribuição Log-Normal se  $\log(T) \sim N(\mu, \sigma^2)$ , então a  $f_T(t)$  é obtido por (considere  $\Phi$  e  $\varphi$ , respectivamente, a função de probabilidade acumulada e função de densidade de probabilidade da normal padrão):

$$\begin{aligned}
f_T(t) &= \frac{d}{dt} P(T \leq t) = \frac{d}{dt} P(\ln T \leq \ln t) = \frac{d}{dt} \Phi\left(\frac{\ln t - \mu}{\sigma}\right) = \\
&= \varphi\left(\frac{\ln t - \mu}{\sigma}\right) \frac{d}{dt} \left(\frac{\ln t - \mu}{\sigma}\right) = \varphi\left(\frac{\ln t - \mu}{\sigma}\right) \frac{1}{\sigma t} =
\end{aligned}$$

$$= \frac{1}{\sigma t \sqrt{2\pi}} \exp \left\{ \frac{-(\log t - \mu)^2}{(2\sigma^2)} \right\}, \quad t \geq 0, \quad -\infty < \mu < \infty, \quad \sigma > 0. \quad (2.35)$$

A função de Sobrevivência e Taxa de Falha:

$$S(t) = 1 - \Phi \left( \frac{\log t - \mu}{\sigma} \right), \quad t > 0, \quad (2.36)$$

$$h(t) = \frac{\frac{1}{\sigma t \sqrt{2\pi}} \exp \left\{ \frac{-(\log t - \mu)^2}{(2\sigma^2)} \right\}}{1 - \Phi \left( \frac{\log t - \mu}{\sigma} \right)}, \quad t > 0. \quad (2.37)$$

De modo que, seja  $T$  o tempo de falha, sujeito a censura  $C$ . Assume-se que  $T \sim \text{Lognormal}(\mu_1, \sigma_1^2)$  e  $C \sim \text{Lognormal}(\mu_2, \sigma_2^2)$ , a função de confiabilidade  $R = P(\log C < \log T)$  será obtida por meio da função de confiabilidade  $R = P(C < T)$ . Portanto, considere que a função densidade de probabilidade conjunta  $f_{T,C}(t, c)$ , para o parâmetro força-stress, então:

$$R = P(C < T) = \int_{-\infty}^{\infty} \int_{-\infty}^t f_{T,C}(t, c) dc dt, \quad (2.38)$$

visto que as variáveis aleatórias são independentes, então  $f_{T,C}(t, c) = f_T(t)f_C(c)$ , assim:

$$R = P(C < T) = \int_{-\infty}^{\infty} \int_{-\infty}^t f_T(t)f_C(c) dc dt. \quad (2.39)$$

Se  $F_C(t) = \int_{-\infty}^t f_C(c) dc$  e como a distribuição não assume valores negativos, portanto  $F_C(t) = \int_0^t f_C(c) dc$ , desse modo:

$$R = P(C < T) = \int_0^{\infty} F_C(t)f_T(t) dt. \quad (2.40)$$

Suponha que  $F_C(t)$  seja  $\Phi_C$  (tem-se perda de generalidade), logo

$$R = P(C < T) = P(C - T < 0) = \Phi(\gamma). \quad (2.41)$$

Além disso, suponha uma variável aleatória Log-Normal  $Z$ , sendo  $Z = C - T$ , então  $Z \sim \text{Lognormal}(\mu_z, \sigma_z^2)$ , no qual  $\mu_z = \mu_2 - \mu_1$  e  $\sigma_z^2 = \sigma_1^2 + \sigma_2^2 \Rightarrow \sqrt{\sigma_1^2 + \sigma_2^2}$ , portanto:

$$R = P(Z < 0) = 1 - \Phi\left(-\frac{\mu_z}{\sigma_z}\right) = \Phi\left(\frac{\mu_z}{\sigma_z}\right). \quad (2.42)$$

Por conseguinte do apresentado (Nkemnole e Samiyu, 2017):

$$R = \Phi(\gamma) = \Phi\left(\frac{\mu_z}{\sigma_z}\right) = \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right). \quad (2.43)$$

#### 2.2.4 Birnbaum-Saunders

Uma variável aleatória  $T$  com distribuição Birnbaum-Saunders tem a função densidade de probabilidade igual a:

$$f_T(t) = \frac{1}{2\sqrt{2\pi\alpha\beta}} \left[ \left(\frac{\beta}{t}\right)^{\frac{1}{2}} + \left(\frac{\beta}{t}\right)^{\frac{3}{2}} \right] \exp \left[ -\frac{1}{2\alpha^2} \left( \frac{t}{\beta} + \frac{\beta}{t} - 2 \right) \right], \quad t > 0, \alpha > 0, \beta > 0, \quad (2.44)$$

e a função de Sobrevivência

$$S(t) = 1 - \Phi \left[ \frac{1}{\alpha} \left\{ \left(\frac{t}{\beta}\right)^{\frac{1}{2}} - \left(\frac{\beta}{t}\right)^{\frac{1}{2}} \right\} \right], \quad t > 0. \quad (2.45)$$

Considere,  $T$  e  $C$  variáveis aleatórias independentes com distribuição Birnbaum-Saunders, representando, respectivamente, tempo de falha e censura. Portanto, sendo  $T \sim BS(\alpha_2, \beta_2)$  e

$C \sim BS(\alpha_1, \beta_1)$ , por conseguinte o parâmetro força-stress é:

$$\begin{aligned}
 R = P(C < T) &= \int_0^\infty \int_0^t f_T(t) f_C(c) dc dt = \\
 &= \int_0^\infty \int_0^t \frac{1}{2\sqrt{2\pi\alpha_2\beta_2}} \left[ \left( \frac{\beta_2}{t} \right)^{\frac{1}{2}} + \left( \frac{\beta_2}{t} \right)^{\frac{3}{2}} \right] \exp \left[ -\frac{1}{2\alpha_2^2} \left( \frac{t}{\beta_2} + \frac{\beta_2}{t} - 2 \right) \right] \\
 &\quad \times \frac{1}{2\sqrt{2\pi\alpha_1\beta_1}} \left[ \left( \frac{\beta_1}{c} \right)^{\frac{1}{2}} + \left( \frac{\beta_1}{c} \right)^{\frac{3}{2}} \right] \exp \left[ -\frac{1}{2\alpha_1^2} \left( \frac{c}{\beta_1} + \frac{\beta_1}{c} - 2 \right) \right] dc dt = \\
 &= \int_0^\infty \left( \int_0^t \frac{1}{2\sqrt{2\pi\alpha_1\beta_1}} \left[ \left( \frac{\beta_1}{c} \right)^{\frac{1}{2}} + \left( \frac{\beta_1}{c} \right)^{\frac{3}{2}} \right] \exp \left[ -\frac{1}{2\alpha_1^2} \left( \frac{c}{\beta_1} + \frac{\beta_1}{c} - 2 \right) \right] dc \right) \\
 &\quad \times \frac{1}{2\sqrt{2\pi\alpha_2\beta_2}} \left[ \left( \frac{\beta_2}{t} \right)^{\frac{1}{2}} + \left( \frac{\beta_2}{t} \right)^{\frac{3}{2}} \right] \exp \left[ -\frac{1}{2\alpha_2^2} \left( \frac{t}{\beta_2} + \frac{\beta_2}{t} - 2 \right) \right] dt = \\
 &= \int_0^\infty \Phi \left[ \frac{1}{\alpha_1} \left\{ \left( \frac{\beta_1}{t} \right)^{\frac{1}{2}} - \left( \frac{\beta_1}{t} \right)^{\frac{3}{2}} \right\} \right] \frac{1}{2\sqrt{2\pi\alpha_2\beta_2}} \left[ \left( \frac{\beta_2}{t} \right)^{\frac{1}{2}} + \left( \frac{\beta_2}{t} \right)^{\frac{3}{2}} \right] \\
 &\quad \times \exp \left[ -\frac{1}{\alpha_2^2} \left( \frac{t}{\beta_2} + \frac{\beta_2}{t} - 2 \right) \right] dt, \tag{2.46}
 \end{aligned}$$

a equação acima não pode ser expressa explicitamente, somente com a ajuda de solução numérica (Naznin e Rahman, 2015) (Balakrishnan e Kundu, 2018).

## Capítulo 3

# Geração de dados de sobrevivência com censura à direita

Como visto na Subseção 2.1.1 a censura do Tipo I ocorre nos casos em que o experimento tem um tempo de duração pré-determinado e ao final deste tempo um ou mais elementos em estudo não falharam e a informação sobre o tempo de falha destes elementos será incompleta. O percentual de censuras neste caso não é controlado pelo pesquisador.

A censura Tipo II ocorre nos casos em que o experimento é conduzido até que um número  $r \leq n$  de elementos falhe e as demais observações que não apresentaram o evento de interesse são censuradas no tempo em que a  $r$ -ésima falha ocorre. As  $r$  falhas são pré-determinadas pelo pesquisador antes do início do experimento, de modo que o percentual de censura seja uma constante.

A censura Aleatória é a mais geral das censuras e engloba as demais. Na censura Aleatória os elementos são censurados por causas alheias ao estudo, seja porque falharam por motivos diferentes ao interesse do estudo ou não puderam ser acompanhados até o final do experimento.

Os algoritmos para gerar cada um dos tipos de dados censurados será apresentado nas Seções 3.1 a 3.3.



### 3.1 Censura Tipo I

Para geração de dados censurados do Tipo I, inicialmente define-se o tempo de duração (pré-definido) do experimento,  $t^*$ . Esse tempo depende do percentual de censura desejado e pode ser facilmente obtido por meio do quantil correspondente da distribuição que será considerada para gerar os dados.

Assim, o tempo de duração do experimento,  $t^*$ , para gerar uma amostra com um percentual esperado de censura de  $p \times 100\%$  ( $0 \leq p < 1$ ) pode ser definido pelo quantil  $(1 - p)$ ,  $t_{(1-p)}$  da variável aleatória  $T$ , isto é,

$$t^* = t_{(1-p)} = S^{-1}(p), \quad (3.1)$$

em que  $S^{-1}(\cdot)$  é o inverso da função de sobrevivência  $S(t)$ .

Como exemplo, considere que o objetivo é gerar amostras de uma variável aleatória  $T$ , com distribuição exponencial com média  $\lambda$  e  $p \times 100\%$  de censura do Tipo I. A função de sobrevivência de  $T$  é dado por  $S(t) = e^{-\frac{t}{\lambda}}$ . Assim, para este exemplo, o tempo de duração  $t^*$ , é definido por

$$\begin{aligned} t^* &= t_{(1-p)} = S^{-1}(p) \\ &\Downarrow \\ p &= e^{-\frac{t^*}{\lambda}} \\ &\Downarrow \\ t^* &= -\lambda \log(p). \end{aligned} \quad (3.2)$$

Dessa forma, todos os valores gerados de  $T$ , que forem maiores do que  $t^*$ , serão substituídos por  $t^*$  e serão considerados censurados.

Um procedimento para gerar amostras de tempo de sobrevivência com censura do Tipo I é apresentado pelo Algoritmo 1, a seguir:

---

**Algoritmo 1:** Para gerar amostras com  $p \times 100\%$  de censura do Tipo I

---

**Passo 1:** Obter  $t^* = t_{(1-p)}$ , quantil  $(1 - p)$  da distribuição de interesse;

**Passo 2:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse;

**Passo 3:** Gerar o indicador de censura  $\delta_i, i = 1, \dots, n$ :

se  $T_i \leq t^*$ , fazer  $\delta_i = 1$ ;

se  $T_i > t^*$ , fazer  $\delta_i = 0$  e  $T_i = t^*$ .

---

As demais distribuições de interesse desse estudo possuem os seguintes tempos de duração,  $t^*$ :

- Weibull:  $t^* = \lambda[-\log(p)]^{\frac{1}{\alpha}}$ ;
- Log-Normal:  $t^* = \exp[\mu + \sigma Z_p]$ , sendo  $Z_p$  o quantil  $p$  da distribuição Normal Padrão;
- Birnbaum-Sanders:  $t^* = \frac{\beta}{4} \left[ \alpha Z_p + \sqrt{(\alpha Z_p)^2 + 4} \right]^2$ , sendo  $Z_p$  o quantil  $p$  da distribuição Normal Padrão.

Note que o percentual de censura obtido pode não ser igual ao valor pré-fixado,  $p$ , de fato o percentual de censura gerado é uma quantidade aleatória.

### 3.2 Censura Tipo II

No caso de censura do Tipo II, é preciso pré-definir o número de falhas do experimento,  $r$ . Portanto, para gerar os dados, é necessário definir o percentual de censura  $p \times 100\%$  (ou percentual de falhas  $(1 - p) \times 100\%$ ), ou seja, o percentual de censura será uma constante. O número de falhas é obtido por meio do percentual  $(1 - p)$  desejado e o tamanho da amostra, isto é,

$$r \propto n(1 - p), \quad (3.3)$$

com  $r \in \mathbb{Z}_+^*$ , assim se o resultado tiver casas decimais, o número de falhas deve ser arredondado para o próximo número inteiro superior.

Dessa forma, ao gerar amostras de uma variável aleatória  $T$ , com a distribuição de interesse, as observações que tiverem tempo superior ao apresentado pela falha do quantil  $(1 - p)$  dos

valores gerados  $t_{(1-p)}$ , serão censurados. A seguir é apresentado o Algoritmo 2 que leva em consideração a geração de dados censurados do Tipo II:

---

**Algoritmo 2:** Para gerar amostras com  $p \times 100\%$  de censura do Tipo II

---

**Passo 1:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse;

**Passo 2:** Definir tempo limite:  $t^* = t_{(1-p)}$  da amostra gerada;

**Passo 3:** Gerar o indicador de censura  $\delta_i$ ,  $i = 1, \dots, n$ :

se  $T_i \leq t^*$ , fazer  $\delta_i = 1$ ;

se  $T_i > t^*$ , fazer  $\delta_i = 0$  e  $T_i = t^*$ .

---

Note que, o percentual de censura gerado é fixo, porém, pode não ser exatamente igual a  $p \times 100\%$ . Isso ocorre pelo fato de o percentual  $p$  não ser múltiplo do tamanho da amostra e/ou presença de empates  $t^*$ .

### 3.3 Censura Aleatória

O mecanismo de censura aleatória considera que a censura ocorre de forma independente à falha. O mecanismo é similar ao parâmetro de força-stress apresentado na Seção 2.2. Considere que o tempo de falha é representado por  $T$  uma v.a. e o tempo até a censura por  $C$  também uma v.a. e independente de  $T$ . Portanto, a observação será:

$$t = \min(T, C), \quad (3.4)$$

e a variável indicadora de censura é:

$$\delta = \begin{cases} 1, & \text{se } T \leq C \\ 0, & \text{se } T > C \end{cases}. \quad (3.5)$$

Abaixo são incluídos os casos específicos da censura aleatória. Na Subseção 3.3.1 é apresentado o caso mais conhecido de censura aleatória, quando  $C$  e  $T$  têm a mesma distribuição. Na sequência (Subseção 3.3.2) é explorado um método numérico para determinar os valores dos

parâmetros da distribuição de  $C$ , quando os mesmos não podem ser (ou são difíceis de serem) obtidos analiticamente.

Na Subseção 3.3.3 é proposta a alternativa de considerar a distribuição  $C$  diferente da distribuição do tempo de sobrevivência e por fim, na Subseção 3.3.4 é uma proposta uma alternativa de obter o indicador de censura por meio de uma distribuição de Bernoulli.

### 3.3.1 Caso 1 - Distribuição de $C$ similar à $T$

Uma alternativa para geração de dados com censura aleatória é considerar distribuições similares para o tempo até a falha,  $T$ , e o tempo de censura,  $C$ , isto é, a mesma família de distribuições, porém com parâmetros diferentes.

Suponha que o interesse é gerar valores censurados (censura aleatória) de uma distribuição qualquer com (vetor de) parâmetro(s)  $\theta$ . Neste caso, como  $C$  é da mesma família de distribuições de  $T$ , o procedimento se resume a encontrar qual é o (vetor de) parâmetro(s) da distribuição de  $C$ .

Este vetor pode ser obtido (quando existir solução analítica) por meio da resolução de um sistema de equações  $p = R = P(C < T)$ , em que  $p$  é o percentual de censura desejada na amostra.

Um procedimento para geração de amostras de tempo de sobrevivência com censura Aleatória Caso 1 é apresentado pelo Algoritmo 3. A seguir são apresentados esse procedimento para as distribuições Exponencia, Weibull e Log-normal.

### Exponencial

Considere o caso em que o interesse é obter uma amostra com  $p \times 100\%$  de censura de uma distribuição Exponencial com parâmetro  $\lambda_1$ . Se  $T \sim Exp(\lambda_1)$  e  $C \sim Exp(\lambda_2)$ , dessa forma

(Equação 2.30 da Subseção 2.2.1):

$$p = R = P(C < T) = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \quad (3.6)$$

$$p = \frac{\lambda_1}{\lambda_1 + \lambda_2} \Rightarrow p(\lambda_1 + \lambda_2) = \lambda_1 \Rightarrow p\lambda_2 = \lambda_1(1 - p),$$

então

$$\lambda_2 = \frac{1 - p}{p} \lambda_1. \quad (3.7)$$

Portanto, a distribuição de censura  $C$  é uma  $Exp(\frac{1-p}{p} \lambda_1)$ .

## Weibull

Considere gerar uma amostra com  $p \times 100\%$  de censura de uma distribuição Weibull com parâmetros  $\alpha$  e  $\lambda_1$ . Se  $T \sim Weibull(\alpha, \lambda_1)$  e  $C \sim Weibull(\alpha, \lambda_2)$ , então (Equação 2.34 da Subseção 2.2.2):

$$p = R = P(C < T) = \frac{\lambda_1^\alpha}{(\lambda_1^\alpha + \lambda_2^\alpha)} \quad (3.8)$$

$$p = \frac{\lambda_1^\alpha}{(\lambda_1^\alpha + \lambda_2^\alpha)} \Rightarrow p(\lambda_1^\alpha + \lambda_2^\alpha) = \lambda_1^\alpha$$

$$p\lambda_2^\alpha = \lambda_1^\alpha(1 - p) \Rightarrow \lambda_2^\alpha = \lambda_1^\alpha \frac{(1 - p)}{p},$$

devido a  $\lambda_1, \lambda_2 > 0$  e  $p > 0$ , então

$$\lambda_2 = \lambda_1 \left( \frac{1 - p}{p} \right)^{\frac{1}{\alpha}}. \quad (3.9)$$

Assim, a distribuição do tempo de censura  $C$  deve ser  $Weibull(\alpha, (\frac{1-p}{p})^{1/\alpha} \lambda_1)$ .

### Log-Normal

Para gerar uma amostra com  $p \times 100\%$  de censura de uma distribuição Log-Normal com parâmetros  $\mu_1$  e  $\sigma_1^2$ , sendo  $T \sim \text{Lognormal}(\mu_1, \sigma_1^2)$  e  $T \sim \text{Lognormal}(\mu_2, \sigma_2^2)$ , temos que (Equação 2.43 da Subseção 2.2.3):

$$p = R = \Phi(\gamma) = \Phi\left(\frac{\mu_z}{\sigma_z}\right) = \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right), \quad (3.10)$$

sendo  $\Phi(\cdot)$  a função de distribuição acumulada da normal padrão.

Para isolar os parâmetros  $\mu_2$  e  $\sigma_2^2$  será usada a inversa da função de distribuição acumulada da normal padrão,  $\Phi^{-1}(\cdot)$ , portanto:

$$p = \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \Rightarrow \Phi^{-1}(p) = \frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \quad (3.11)$$

sendo as soluções para  $\mu_2$  e  $\sigma_2^2$ , respectivamente:

$$\mu_2 = \Phi^{-1}(p)\sqrt{\sigma_1^2 + \sigma_2^2} + \mu_1, \quad (3.12)$$

e

$$\sigma_2^2 = \left(\frac{\mu_2 - \mu_1}{\Phi^{-1}(p)}\right)^2 - \sigma_1^2. \quad (3.13)$$

Note que a Equação 3.12 e Equação 3.13 são linearmente dependentes. De acordo com Biazatti e Nakano, 2020,  $\sigma_1^2$  é conhecida e pode-se considerar  $\sigma_2^2 = \sigma_1^2$  então, a Equação 3.12 se reduz a:

$$\mu_2 = \Phi^{-1}(p)\sqrt{2\sigma_1^2} + \mu_1. \quad (3.14)$$

Por conseguinte, o tempo de censura  $C \sim \text{Lognormal}(\Phi^{-1}(p)\sqrt{2\sigma_1^2} + \mu_1, \sigma_1^2)$ , no qual  $\Phi^{-1}(p)$  é o quantil  $p$  de uma distribuição normal padrão.

Note que, assim como no método de geração de dados com censura do Tipo I, o percentual

---

**Algoritmo 3:** Para gerar amostras com censura Aleatória Caso 1

---

**Passo 1:** Obter os valores dos parâmetros da distribuição do tempo de censura ( $C$ ) a partir dos parâmetros da distribuição do tempo de sobrevivência desejado, proporção de censuras ( $p$ ) e do parâmetro força-stress ( $R$ );

**Passo 2:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse;

**Passo 3:** Gerar  $C_1, C_2, \dots, C_n$  da distribuição de censura cujos parâmetros foram definidos no Passo 1;

**Passo 4:** Gerar o indicador de censura  $\delta_i, i = 1, \dots, n$ :

se  $T_i \leq C_i$ , fazer  $\delta_i = 1$ ;

se  $T_i > C_i$ , fazer  $\delta_i = 0$  e  $T_i = C_i$ .

---

de censura obtido pode não ser igual ao valor pré-fixado,  $p$ , de fato, o percentual de censura gerado é uma quantidade aleatória.

### 3.3.2 Caso 2 - Alternativo numérico

Note que nem sempre é possível obter analiticamente o parâmetro de força-stress de uma distribuição de probabilidades. Na Subseção 2.2.4 foi visto o caso da distribuição Birnbaum-Saunders, cujo valor de  $R = P(C < T)$  não tem fórmula analítica fechada.

Nestas situações é possível encontrar numericamente a solução do sistema  $p = R = P(C < T)$ . Uma alternativa é obter os valores dos parâmetros da distribuição do tempo de censuras variando os parâmetros da distribuição de  $C$  até que o percentual de censura gerado seja próximo de  $p$ .

A média das distribuições Exponencial, Weibull, Log-Normal e Birnbaum-Saunders para diferentes parâmetros são apresentadas, respectivamente, a seguir:

- $X \sim Exp(\lambda): E(X) = \lambda$ ;
- $X \sim Weibull(\alpha, \lambda): E(X) = \lambda \Gamma\left(1 + \frac{1}{\alpha}\right)$ ;
- $X \sim Lognormal(\mu, \sigma^2): E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$ ;
- $X \sim BS(\alpha, \beta): E(X) = \beta \left(1 + \frac{\alpha^2}{2}\right)$ .

Nota-se que a média das distribuições Exponencial e Weibull aumentam à medida que o valor de  $\lambda$  (parâmetro de escala) cresce. O mesmo ocorre com a distribuição Log-Normal e Birnbaum-Saunders quanto maiores os valores dos parâmetros  $\mu$  (parâmetro de localização) e  $\beta$  (parâmetro de escala), respectivamente.

Por conseguinte, um procedimento para encontrar numericamente a solução do sistema  $p = R = P(C < T)$  consiste no fato de que o percentual de censura tende a aumentar à medida que a média (ou mediana) da distribuição do tempo de censura é maior.

Tomando como exemplo a distribuição Birnbaum-Saunders, um procedimento para gerar uma amostra de tempos com  $p \times 100\%$  de censura é apresentado no Algoritmo 4.

---

**Algoritmo 4:** Para gerar amostras com censura Aleatória Caso 2 (Para distribuição Birnbaum-Saunders)

---

Tempo de vida segundo uma  $BS(\alpha, \beta)$  e  $p \times 100\%$  de censura.

**Passo 1:** Obter os valores dos parâmetros da distribuição do tempo de censura ( $C \sim BS(\alpha, \beta_2)$ ) a partir dos parâmetros da distribuição do tempo de sobrevivência desejado ( $T \sim BS(\alpha, \beta_1)$ ), proporção de censuras ( $p$ ) e do parâmetro força-stress ( $R$ );

**Passo 1.0:** Definir  $\beta_2 = \beta_1$  ( $\beta_1$  é o chute inicial de  $\beta_2$ );

**Passo 1.1:** Gerar  $X_1, X_2, \dots, X_M$  valores de  $X \sim BS(\alpha, \beta_1)$  ( $M$  grande);

**Passo 1.2:** Gerar  $Y_1, Y_2, \dots, Y_M$  valores de  $Y \sim BS(\alpha, \beta_2)$  ( $M$  grande);

**Passo 1.3:** Calcular  $R = \frac{1}{M} \sum_{i=1}^M I(Y_i < X_i)$  ( $R = P(Y < X)$  empírico);

**Passo 1.4:** Para  $\varepsilon > 0$  e  $\eta > 0$ :

se  $R - p > \varepsilon$ , fazer  $\beta_2 = \beta_2 + \eta$  e voltar ao **Passo 1.1**;

se  $R - p < -\varepsilon$ , fazer  $\beta_2 = \beta_2 - \eta$  e voltar ao **Passo 1.1**;

se  $-\varepsilon < R - p < \varepsilon$ , ir ao **Passo 2**;

**Passo 2:** Gerar  $T_1, T_2, \dots, T_n$  de  $T \sim BS(\alpha, \beta_1)$ ;

**Passo 3:** Gerar  $C_1, C_2, \dots, C_n$  de  $C \sim BS(\alpha, \beta_2)$ ;

**Passo 4:** Gerar o indicador de censura  $\delta_i$ ,  $i = 1, \dots, n$ :

se  $T_i \leq C_i$ , fazer  $\delta_i = 1$ ;

se  $T_i > C_i$ , fazer  $\delta_i = 0$  e  $T_i = C_i$ .

---

Note no **Passo 1.4** que, se o  $R = P(Y < X) \gg p$ , significa que os tempos de  $Y$  devem ser maiores e, portanto, o valor de  $\beta_2$  deve ser maior que o considerado. Analogamente, se  $R = P(Y < X) \ll p$ , o valor de  $\beta_2$  deve ser menor.

O Algoritmo 4 pode ser facilmente adaptado para as demais distribuições apresentadas nesse



estudo realizando as seguintes modificações nos Passos 1 a 3:

- $Exp(\lambda)$ :  $\lambda_1 = \lambda_2$ ;  $X \sim Exp(\lambda_1)$ ;  $Y \sim Exp(\lambda_2)$ ;  
 $T \sim Exp(\lambda_1)$  e  $C \sim Exp(\lambda_2)$  ( $\lambda_2$  obtido a partir do **Passo 1.0 a Passo 1.4**).
- $Weibull(\alpha, \lambda)$ :  $\lambda_1 = \lambda_2$ ;  $X \sim Weibull(\alpha, \lambda_1)$ ;  $Y \sim Weibull(\alpha, \lambda_2)$ ;  $T \sim Weibull(\alpha, \lambda_1)$   
e  $C \sim Weibull(\alpha, \lambda_2)$  ( $\lambda_2$  obtido a partir do **Passo 1.0 a Passo 1.4**).
- $Lognormal(\mu, \sigma^2)$ :  $\mu_1 = \mu_2$ ;  $X \sim Lognormal(\mu_1, \sigma^2)$ ;  $Y \sim Lognormal(\mu_2, \sigma^2)$ ;  
 $T \sim Lognormal(\mu_1, \sigma^2)$  e  $C \sim Lognormal(\mu_2, \sigma^2)$  ( $\mu_2$  obtido a partir do **Passo 1.0 a Passo 1.4**).

### 3.3.3 Caso 3 - Distribuição de $C$ diferente de $T$

Uma alternativa para gerar dados com censura aleatória é considerar uma família de distribuições diferente para gerar os tempos de censura,  $C$ . De fato, essa é uma alternativa útil, principalmente se a distribuição na qual se deseja gerar os tempos é complexa.

Neste sentido, devido à complexidade da distribuição do tempo de sobrevivência  $T$ , descartamos, neste caso, o cálculo analítico de  $R = P(C < T)$ . Ademais, nesta situação, o processo de geração de dados é semelhante ao Caso 2 apresentado na Subseção 3.3.2, diferindo apenas em utilizar uma distribuição diferente para gerar o tempo de censura  $C$ .

Por exemplo, tome a distribuição Exponencial com média  $\lambda$ , se  $C \sim Exp(\lambda)$  então o percentual de censura tende a aumentar (diminuir) à medida que a  $\lambda$  aumenta (diminui). Para aplicar esse método de geração de amostra com censura, o intuito é utilizar uma distribuição para tempo de censura,  $C$ , menos complexa que a distribuição de tempo,  $T$ .

Um procedimento para gerar tempos de vida segundo uma distribuição qualquer com  $p \times 100\%$  de censura e considerando o tempo de censura uma Exponencial é apresentado no Algoritmo 5.

---

**Algoritmo 5:** Para gerar amostras com censura Aleatória Caso 3 (Tempos da distribuição de interesse com censura Exponencial)

---

Tempos da distribuição de interesse com censura Exponencial e  $p \times 100\%$  de censura.

**Passo 1:** Obter o valor de  $\lambda$  da distribuição Exponencial para o tempo de censura ( $C \sim Exp(\lambda)$ ) a partir dos parâmetros da distribuição do tempo de sobrevivência desejado, proporção de censuras ( $p$ ) e do parâmetro força-stress ( $R$ );

**Passo 1.0:** Definir  $\lambda$ ;

**Passo 1.1:** Gerar  $X_1, X_2, \dots, X_M$  da distribuição de interesse ( $M$  grande);

**Passo 1.2:** Gerar  $Y_1, Y_2, \dots, Y_M$  valores de  $Y \sim Exp(\lambda)$  ( $M$  grande);

**Passo 1.3:** Calcular  $R = \frac{1}{M} \sum_{i=1}^M I(Y_i < X_i)$  ( $R = P(Y < X)$  empírico);

**Passo 1.4:** Para  $\varepsilon > 0$  e  $\eta > 0$ :

se  $R - p > \varepsilon$ , fazer  $\lambda = \lambda + \eta$  e voltar ao **Passo 1.1**;

se  $R - p < -\varepsilon$ , fazer  $\lambda = \lambda - \eta$  e voltar ao **Passo 1.1**;

se  $-\varepsilon < R - p < \varepsilon$ , ir ao **Passo 2**;

**Passo 2:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse;

**Passo 3:** Gerar  $C_1, C_2, \dots, C_n$  de  $C \sim Exp(\lambda)$ ;

**Passo 4:** Gerar o indicador de censura  $\delta_i$ ,  $i = 1, \dots, n$ :

se  $T_i \leq C_i$ , fazer  $\delta_i = 1$ ;

se  $T_i > C_i$ , fazer  $\delta_i = 0$  e  $T_i = C_i$ .

---

### 3.3.4 Caso 4 - Indicador de Censura via Bernoulli

Outra alternativa para gerar dados com censura aleatória, além da Subseção 3.3.3, é considerar a distribuição Bernoulli para gerar o indicador de Censura,  $\delta$ . Considere o tempo de falha,  $T$ , uma distribuição de interesse e, a variável indicadora de censura,  $C$ , independente de  $T$  e com distribuição Bernoulli com probabilidade de sucesso  $(1 - p)$ , isto é

$$P(C = c) = (1 - p)^c p^{1-c}, \quad 0 \leq p < 1, \quad c = 0, 1, \quad (3.15)$$

assim, a variável indicadora de censura é dada por:

$$\delta = \begin{cases} 1, & \text{se } c = 1 \\ 0, & \text{se } c = 0 \end{cases}. \quad (3.16)$$

Um procedimento para geração de dados com censura Aleatória Caso 4 é apresentada pelo

---

Algoritmo 6, a seguir:

---

**Algoritmo 6:** Para gerar amostras com censura Aleatória Caso 4

---

**Passo 1:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse;

**Passo 2:** Gerar  $\delta_1, \delta_2, \dots, \delta_n$  da distribuição  $Bernoulli(1 - p)$ .

---

Note que o percentual de censura obtido pode não ser igual ao valor pré-fixado,  $p$ , de fato, o percentual de censura gerado é uma quantidade aleatória. Alguns trabalhos têm adotado a metodologia de obtenção do indicador de censura pela distribuição Bernoulli. Podemos citar, por exemplo, os trabalhos de Brunello e Nakano, 2015, e Cardial, Fachini-Gomes e Nakano, 2020.

# Capítulo 4

## Geração de dados discretos de sobrevivência

Os métodos de geração de dados de sobrevivência com censura à direita apresentado no Capítulo 3 por meio dos Algoritmos 1 a 6 podem ser utilizados para gerar dados discretos, sendo necessárias algumas adaptações.

Todos os modelos de variáveis contínuas podem ser usados para obter modelos discretos, agrupados os tempos em intervalos unitários. A variável aleatória discreta é então definida como a parte inteira da variável, isto é, o maior valor inteiro menor ou igual ao valor observado (Nakano e Carrasco, 2006) (Vila, Nakano e Saulo, 2019).

Desta forma, uma maneira simples de gerar dados discretos de sobrevivência é gerar os dados de sua correspondente distribuição contínua e então discretizar o valor obtido. As Seções 4.1 a 4.3 apresentam, os Algoritmos 7 a 11 para geração de dados discretos de sobrevivência com censura à direita a partir dos Algoritmos 1, 2, 3 e 6 com as devidas alterações e adaptações.

#### 4.1 Censura Tipo I - Discreto

O procedimento para gerar amostras de tempo discretos de sobrevivência com censura do Tipo I é realizado a partir do Algoritmo 1 (Seção 3.1), apresentado a seguir pelo Algoritmo 7. Note que, a diferença consiste no **Passo 4** do Algoritmo 7, no qual é introduzido a discretização dos tempos da amostra gerada. Os tempos de duração,  $t^*$ , para as distribuições Exponencial, Weibull, Log-Normal e Birnbaum-Saunders foram apresentados na Seção 3.1.

---

**Algoritmo 7:** Para gerar amostras de tempos discretos com  $p \times 100\%$  de censura do Tipo I

---

**Passo 1:** Obter  $t^* = t_{(1-p)}$ , quantil  $(1 - p)$  da distribuição de interesse;

**Passo 2:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse (análoga contínua);

**Passo 3:** Gerar o indicador de censura  $\delta_i$ ,  $i = 1, \dots, n$ :

se  $T_i \leq t^*$ , fazer  $\delta_i = 1$ ;

se  $T_i > t^*$ , fazer  $\delta_i = 0$  e  $T_i = t^*$ ;

**Passo 4:** Discretizar  $T_i$ ,  $i = 1, \dots, n$  (considerar o maior número inteiro menor ou igual a  $T_i$ ).

---

#### 4.2 Censura Tipo II - Discreto

Para gerar amostras de tempo discretos de sobrevivência com censura do Tipo II basta adaptar o Algoritmo 2 (Seção 3.2), apresentado a seguir pelo Algoritmo 8. Observa-se que a diferença consiste no **Passo 4** do Algoritmo 8, no qual é introduzida a discretização dos tempos. Aqui, o tempo limite,  $t^*$ , é o quantil  $(1 - p)$  dos valores gerados de  $T$  (Seção 3.2.).

---

**Algoritmo 8:** Para gerar amostras de tempos discretos com  $p \times 100\%$  de censura do Tipo II

---

**Passo 1:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse (análoga contínua);

**Passo 2:** Definir tempo limite:  $t^* = t_{(1-p)}$  da amostra gerada;

**Passo 3:** Gerar o indicador de censura  $\delta_i$ ,  $i = 1, \dots, n$ :

se  $T_i \leq t^*$ , fazer  $\delta_i = 1$ ;

se  $T_i > t^*$ , fazer  $\delta_i = 0$  e  $T_i = t^*$ .

**Passo 4:** Discretizar  $T_i$ ,  $i = 1, \dots, n$  (considerar o maior número inteiro menor ou igual a  $T_i$ ).

---

### 4.3 Censura Aleatória

A seguir nas Subseções 4.3.1 a 4.3.3 são apresentadas os diferentes casos de censura aleatória.

#### 4.3.1 Caso 1 - Discreto

Uma alternativa para gerar amostras de tempo discretos de sobrevivência com censura Aleatória Caso 1, é adaptar o Algoritmo 3 (Subseção 3.3.1), apresentado a seguir pelo Algoritmo 9. Nesse Algoritmo é introduzido o **Passo 5** para discretizar os tempos e as censuras.

---

**Algoritmo 9:** Para gerar amostras de tempos discretos com censura Aleatória Caso 1

---

**Passo 1:** Obter os valores dos parâmetros da distribuição do tempo de censura ( $C$ ) a partir dos parâmetros da distribuição do tempo de sobrevivência desejado, proporção de censuras ( $p$ ) e do parâmetro força-stress ( $R$ );

**Passo 2:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse (análoga contínua);

**Passo 3:** Gerar  $C_1, C_2, \dots, C_n$  da distribuição de censura cujos parâmetros foram definidos no Passo 1;

**Passo 4:** Gerar o indicador de censura  $\delta_i, i = 1, \dots, n$ :

se  $T_i \leq C_i$ , fazer  $\delta_i = 1$ ;

se  $T_i > C_i$ , fazer  $\delta_i = 0$  e  $T_i = C_i$ ;

**Passo 5:** Discretizar  $T_i, i = 1, \dots, n$  (considerar o maior número inteiro menor ou igual a  $T_i$ ).

---

#### 4.3.2 Caso 1 - Discreto - Decisão de Censura via Bernoulli

Outra alternativa para gerar amostras de tempo discretos de sobrevivência com censura Aleatória Caso 1, é adaptar o Algoritmo 3 (Subseção 3.3.1) e introduzir a decisão de Censura via Bernoulli, de forma que, em casos de  $T = C$  a decisão de censurar, ou não o tempo de falha, é tomada a partir do resultado de uma  $Bernoulli(1 - p)$  (Biazatti e Nakano, 2020). A seguir é apresentado a construção do Algoritmo 10 feita a partir dessas adaptações:

---

**Algoritmo 10:** Para gerar amostras de tempos discretos com censura Aleatória Caso 1 -  
Decisão de Censura via Bernoulli

---

**Passo 1:** Obter os valores dos parâmetros da distribuição do tempo de censura ( $C$ ) a partir dos parâmetros da distribuição do tempo de sobrevivência desejado, proporção de censuras ( $p$ ) e do parâmetro força-stress ( $R$ );

**Passo 2:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse (análoga contínua);

**Passo 3:** Gerar  $C_1, C_2, \dots, C_n$  da distribuição de censura cujos parâmetros foram definidos no Passo 1;

**Passo 4:** Discretizar  $T_i$  e  $C_i$  com  $i = 1, \dots, n$  (considerar o maior número inteiro menor ou igual a  $T_i$  e  $C_i$ , respectivamente);

**Passo 5:** Gerar o indicador de censura  $\delta_i, i = 1, \dots, n$ :

se  $T_i < C_i$ , fazer  $\delta_i = 1$ ;

se  $T_i > C_i$ , fazer  $\delta_i = 0$ ;

se  $T_i = C_i$ , gerar  $X_i \sim \text{Bernoulli}(1 - p)$ :

se  $X_i = 1$ , fazer  $\delta_i = 1$ ;

se  $X_i = 0$ , fazer  $\delta_i = 0$ .

---

#### 4.3.3 Caso 4 - Discreto - Indicador de Censura via Bernoulli

O procedimento para gerar amostras de tempo discretos de sobrevivência com censura Aleatória Caso 4 realizado a partir do Algoritmo 6 (Subseção 3.3.4), é apresentado a seguir pelo Algoritmo 11. A diferença entre o caso contínuo e discreto, consiste no **Passo 3** no qual é introduzida a discretização dos tempos da amostra gerada.

---

**Algoritmo 11:** Para gerar amostras de tempos discretos com censura Aleatória Caso 4

---

**Passo 1:** Gerar  $T_1, T_2, \dots, T_n$  da distribuição de interesse (análoga contínua);

**Passo 2:** Gerar  $\delta_1, \delta_2, \dots, \delta_n$  da distribuição  $\text{Bernoulli}(1 - p)$ ;

**Passo 3:** Discretizar  $T_i, i = 1, \dots, n$  (considerar o maior número inteiro menor ou igual a  $T_i$ ).

---

# Capítulo 5

## Resultados

Neste capítulo, são apresentados os resultados da aplicação dos Algoritmos de geração de dados de sobrevivência com censura à direita, além da comparação entre eles. As Seções 5.1 a 5.6 são realizadas as simulações dos Algoritmos com o objetivo de avaliar a capacidade de obter o percentual de censura desejado. Nas Seções 5.7 e 5.8 são apresentadas as comparações entre os Algoritmos, por meio do vício das estimativas e censura e, ajuste do modelo.

### 5.1 Simulações do Algoritmo 1 - Censura Tipo I

A Figura 5.1 apresenta os gráficos de 1000 simulações das quatro distribuições de tempo,  $T$ , de interesse desse estudo, ao gerar censura do Tipo I (Seção 3.1), por meio do Algoritmo 1, para diferentes parâmetros e tamanhos de amostra, sendo o percentual esperado de censura desejado igual a 10%<sup>1</sup>.

Observa-se, independente da distribuição de tempo  $T$  e do valor do parâmetros, a média do percentual de censura e o desvio-padrão apresentaram o mesmo comportamento nas amostras de tamanho igual e, foram distintos nas de tamanho diferente.

Mesmo nas amostras pequenas, igual a 10, a média do percentual de censura das simulações

---

<sup>1</sup>ver Apêndice A para mais detalhes da simulações do Algoritmo 1.



atingiu o esperado, porém, o desvio-padrão se apresentou muito alto ( $\approx 9,5\%$ ) e, em mais de 50% das simulações, o percentual de censura, ficou no intervalo de 0% a 20%. No entanto nas amostras de tamanho 100, o desvio-padrão é menor, com média variabilidade ( $\approx 3\%$ ), visto que em várias simulações (mais de 50%) o percentual de censura ficou entre 8% e 12%.

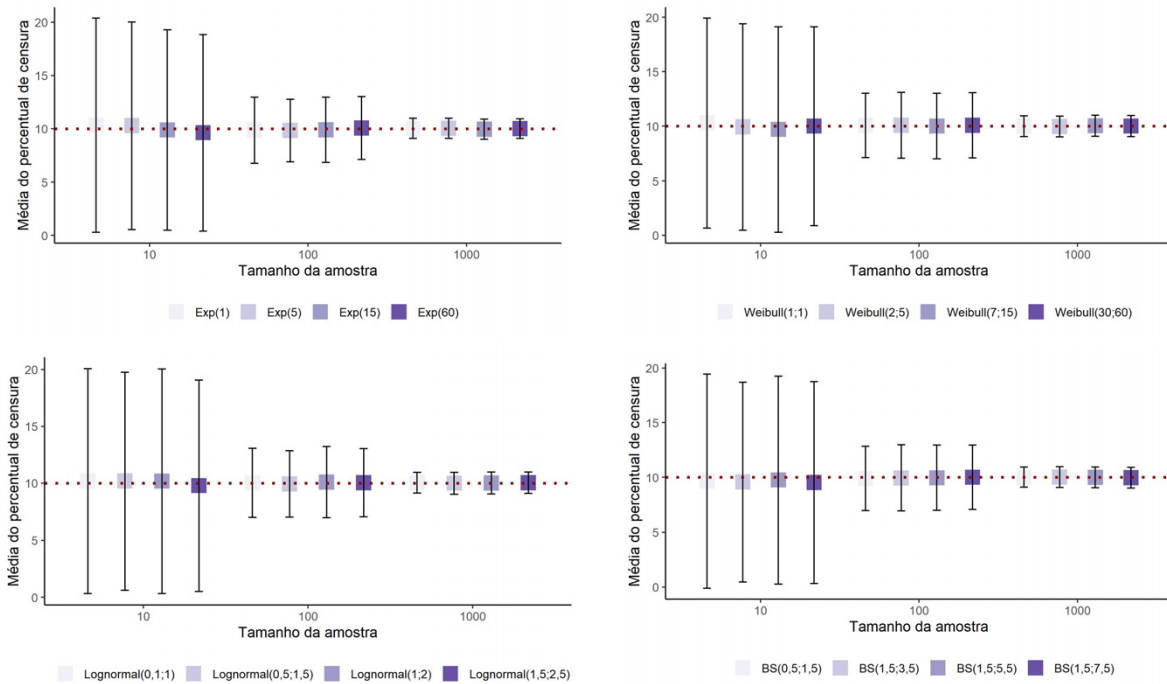


Figura 5.1: Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Tipo I - Algoritmo 1 para as distribuições de tempo  $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

Ainda da Figura 5.1, destaca-se que para as amostras de tamanho 1000, que a média do percentual de censura média ficou muito próximo do esperado e com baixa variabilidade (desvio-padrão  $< 1\%$ ) e, em mais de 50% das simulações o percentual de censura ficou no intervalo de 9,3% e 10,6%, logo, em quase todas as simulações o percentual de censura desejado foi alcançado.

Sendo assim, o Algoritmo 1 é mais adequado para gerar amostras de tamanhos maiores, pois, à medida que o tamanho da amostra aumenta, maior é o controle do percentual de censura desejado.

## 5.2 Simulações do Algoritmo 2 - Censura Tipo II

Com o intuito de avaliar a geração de dados com Censura Tipo II (Seção 3.2), por meio do Algoritmo 2 foram realizadas 1000 simulações de cada uma das distribuições de tempo,  $T$ , para diferentes parâmetros e tamanhos de amostra, com um percentual esperado de censura de 10%<sup>2</sup>.

A Figura 5.2, apresenta um único gráfico para as amostras de tamanho 10, 100 e 1000, visto que, independente distribuição de tempo,  $T$ , e dos parâmetros, o percentual de censura observado foi igual a 10%, sem desvio-padrão.

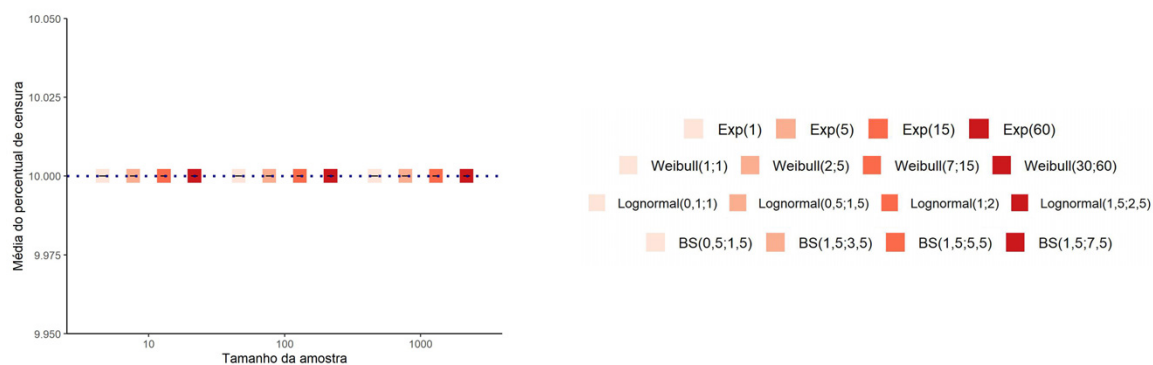


Figura 5.2: Média do percentual de censura em 1000 simulações de Censura Tipo II - Algoritmo 2 para as distribuições de tempo  $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

O resultado das simulações está conforme a construção do Algoritmo 2, que por ter como base a Censura Tipo II, no qual a priori já é definida qual será a censura exata, nas simulações não há eventos aleatórios que possam interferir no percentual de censura definido no início.

## 5.3 Simulações do Algoritmo 3 - Censura Aleatória: Caso 1

A Figura 5.3 mostra os gráficos da censura Aleatória Caso 1 (Subseção 3.3.1), distribuição de  $C$  similar à  $T$ , feita através de 1000 simulações do Algoritmo 3, para cada uma das distribuições de tempo,  $T$ , para diferentes parâmetros e tamanhos de amostra, com percentual de censura

<sup>2</sup>ver Apêndice B para mais detalhes da simulações do Algoritmo 2.

desejado de 10% <sup>3</sup>. Devido aos parâmetros da distribuição de censura,  $C$ , serem obtidos analiticamente, nesse caso não é apresentada a distribuição Birnbaum-Saunders.

Nota-se, assim como na Seção 5.1, dentro das amostras de tamanho igual, a média percentual de censura e o desvio-padrão são semelhantes, independente da distribuição e dos valores do parâmetros utilizados e, entre as amostras de tamanho diferente os resultados são distintos (o desvio-padrão do percentual de censura diminui à medida que o tamanho da amostra aumenta).

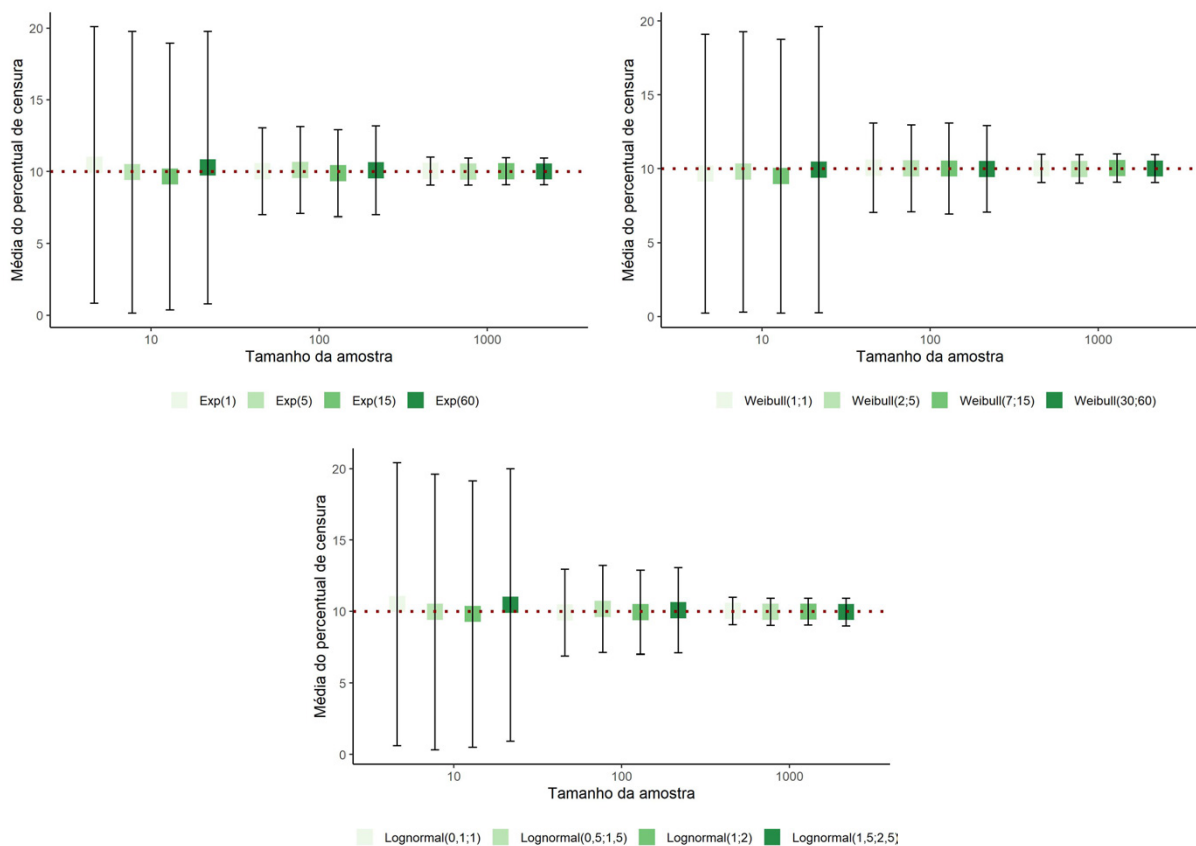


Figura 5.3: Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 1, distribuição de  $C$  similar à  $T$  - Algoritmo 3 para as distribuições: Exponencial, Weibull e Log-Normal e , com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

Ainda da Figura 5.3, nota-se que, nas amostras de tamanho 100 e 1000, a média do percentual e o desvio-padrão têm resultados satisfatório, porque se aproximam do valor desejado e possuem

<sup>3</sup>ver Apêndice C para mais detalhes da simulações do Algoritmo 3.

variabilidade média ( $\approx 3\%$ ) e baixa ( $\approx 1\%$ ), respectivamente. O intervalo do percentual de censura foi igual a (8%, 12%) para amostra de tamanho 100 e (9,3%, 10,6%) para de tamanho 1000.

#### 5.4 Simulações do Algoritmo 4 - Censura Aleatória: Caso 2

Uma forma de gerar censura Aleatória Caso 2, alternativo numérico (tentativa e erro), é por meio do Algoritmo 4. Nesse caso, ao utilizar o algoritmo, é necessário decidir o valor das interações de Monte Carlo ( $M$ ), precisão ( $\varepsilon$ ) e  $\eta$  (valor definido no processo de interações para ser acrescido, ou subtraído, do valor inicial dado ao parâmetro da distribuição de censura  $C$  até obter um parâmetro que satisfaça a relação  $p = R = P(C < T)$ ), de forma a obter um percentual de censura igual ou o mais próximo possível do desejado. Nas Subseções 5.4.1 e 5.4.2 é realizado um estudo para escolha de  $M$ ,  $\varepsilon$  e  $\eta$  e, na Subseção 5.4.3 é apresentada a simulação do Algoritmo 4 com os valores escolhidos.

##### 5.4.1 Escolha do $M$ e $\varepsilon$

Na Tabela 5.1 é apresentado o resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ), diferença absoluta entre a média do percentual de censura e o percentual de censura desejado ( $|E(\hat{p}) - p|$ ), erro quadrático médio ( $EQM(\hat{p})$ ) e média de tempo em segundos por simulação, de 1000 simulações do Algoritmo 4 para  $M$  igual 100, 1000 e 10000 e  $\varepsilon$  igual 0,1, 0,05, 0,01 e 0,001 com  $\eta = 0,5$ , percentual de censura desejado igual a 10% e amostra de tamanho 100. Uma vez que as quatro distribuições de tempo,  $T$ , tiverem resultados semelhantes foi decidido, nessa etapa de decisão de  $M$  e  $\varepsilon$ , apresentar somente a distribuição Exponencial com  $\lambda = 5$ .

Observa-se que para  $M = 100$  em nenhuma das precisões o valor de  $E(\hat{p})$  ficou próximo de 10%. O mesmo ocorre para  $M = 1000$  com  $\varepsilon$  igual a 0,1, 0,05 e 0,01 em que a média percentual de censura não ficou próxima ao desejado. Entretanto, no caso  $M = 1000$  com

Tabela 5.1: Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ), diferença absoluta entre a média do percentual de censura e o percentual de censura desejado ( $|E(\hat{p}) - p|$ ), erro quadrático médio ( $EQM(\hat{p})$ ) e média de tempo em segundos por simulação de 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim Exp(5)$  para diferentes interações de Monte Carlo ( $M$ ) e precisão ( $\varepsilon$ ) com  $\eta = 0,5$ ,  $p = 0,1$  e  $n = 100$

<b>M</b>	$\varepsilon$	$E(\hat{p})$	$DP(\hat{p})$	$CV(\hat{p})$	$ E(\hat{p}) - p $	$EQM(\hat{p})$	<b>Média de tempo em segundos por simulação</b>
100	0,1	0,4988	0,0494	0,0991	0,3988	0,1615	0,000
	0,05	0,2012	0,0459	0,2283	0,1012	0,0123	0,001
	0,01	0,1569	0,0395	0,2516	0,0569	0,0048	0,001
	0,001	0,1433	0,0390	0,2722	0,0433	0,0034	0,001
1000	0,1	0,4980	0,0495	0,0994	0,3980	0,1608	0,000
	0,05	0,1564	0,0357	0,2281	0,0564	0,0044	0,004
	0,01	0,1201	0,0334	0,2779	0,0201	0,0015	0,006
	0,001	0,1053	0,0311	0,2955	0,0053	0,0010	0,008
10000	0,1	0,4990	0,0503	0,1008	0,3990	0,1617	0,001
	0,05	0,1484	0,0358	0,2410	0,0484	0,0036	0,037
	0,01	0,1129	0,0319	0,2827	0,0129	0,0012	0,055
	0,001	0,1032	0,0308	0,2983	0,0032	0,0010	0,062

$\varepsilon = 0,001$  a média do percentual de censura foi próxima do valor desejado com variabilidade média. E, no caso de  $M = 10000$ , o resultado das simulações foi semelhante ao  $M = 1000$ .

A diferença entre  $M$  igual a 1000 e 10000 consiste na média de tempo das simulações. Enquanto  $M = 1000$  com  $\varepsilon = 0,001$  a média de tempo da simulação é de 0,008 segundos,  $M = 10000$  com  $\varepsilon = 0,001$  é de 0,062 segundos, ou seja, 7,75 vezes mais tempo para obter as mesmas estimativas para média, variação e vício.

Portanto, para obter um percentual de censura desejado, o valor de  $M$  deve ser maior ou igual 1000 e  $\varepsilon \leq 0,001$ . Nesse estudo, para simular dados com censura Aleatória Caso 2, por meio do Algoritmo 4, será adotado  $M = 1000$  e  $\varepsilon = 0,001$ , por obter um percentual adequado e computacionalmente ser mais rápido para gerar as simulações.

### 5.4.2 Escolha do $\eta$

Na Subseção 5.4.1 foi realizada a escolha do  $M$  e  $\varepsilon$ , entretanto, ainda é necessário decidir o valor  $\eta$  do Algoritmo 4. Na Subseção 5.4.2 são apresentadas as quatro distribuições de tempo,

$T$ , de interesse desse estudo e são definidos os valores de  $\eta$  para cada uma delas.

## Exponencial

Na Figura 5.4 são apresentadas as 1000 simulações da distribuição de tempo  $T \sim Exp(5)$ , com  $\varepsilon = 0,001$ ,  $M = 1000$  (conforme decidido na Subseção 5.4.1), percentual de censura desejado de 10% e amostra de tamanho 1000 para diferentes valores de  $\eta$ .

O primeiro gráfico mostra a média de percentual de censura obtida e no segundo gráfico o log da média do tempo da simulação para cada  $\eta$ .

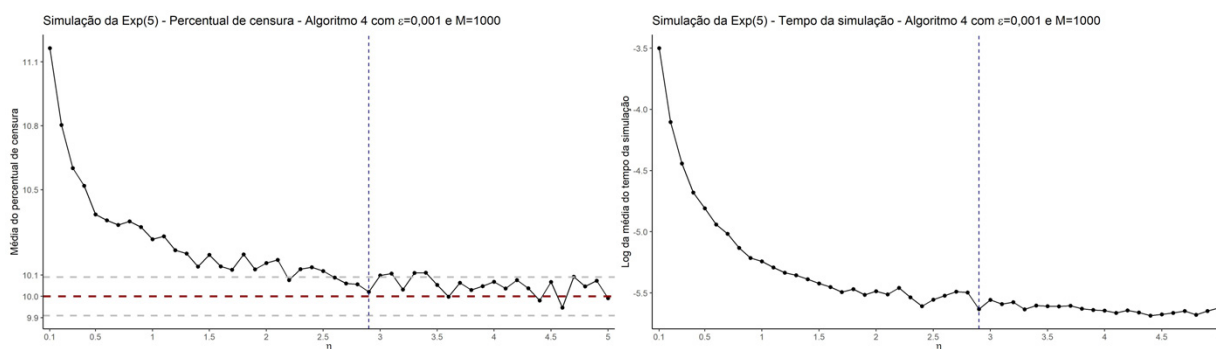


Figura 5.4: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim Exp(5)$  para diferentes  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

Nota-se, no primeiro gráfico (Figura 5.4), que para  $\eta < 0,5$ , a média do percentual de censura fica acima de 10,5% e para  $\eta < 2,5$  a diferença entre a média e o valor desejado fica acima de 0,1%. De acordo com a análise, os valores de  $\eta$  devem estar no intervalo  $(2,5, 5)$  para que o percentual de censura desejado seja alcançado com precisão de 1 casa decimal, ou seja, média do percentual de censura acima de 9,91% e abaixo de 10,01%.

Ainda da Figura 5.4, vale ressaltar, que a média de tempo da simulação acompanha o acerto do percentual de censura, já que quanto mais distante o percentual de censura ficou do desejado, maior o tempo da simulação. Por exemplo, no caso de  $\eta = 0,1$  que foi a maior média do percentual de censura observada (11,16%), a média do tempo da simulação foi o maior entre todos os observados.

A escolha de  $\eta$  será com base na combinação de dois fatores: média do percentual de censura próxima do  $p$  desejado e o menor tempo médio para gerar a censura, por conseguinte, a escolha de  $\eta = 2,9$  (indicado em ambos os gráficos pela linha em azul) parece ser o mais adequada nessa situação.

A Tabela 5.2 mostra o valor de  $\eta$  escolhido<sup>4</sup> para os parâmetros da distribuição Exponencial que são de interesse desse estudo, conforme a lógica utilizada para  $Exp(5)$ .

Tabela 5.2: Valor de  $\eta$  escolhido para gerar  $\lambda_2$  da distribuição de censura  $C \sim Exp(\lambda_2)$  a partir da distribuição de tempo  $T \sim Exp(\lambda_1)$  para diferentes parâmetros a partir das 1000 simulações do Algoritmo 4 com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

Distribuição de $T$	$\eta$
$Exp(1)$	0,84
$Exp(5)$	2,9
$Exp(15)$	11
$Exp(60)$	53

## Weibull

A Figura 5.5 apresenta dois gráficos com informações das 1000 simulações geradas do Algoritmo 4 para a distribuição de tempo  $T \sim Weibull(2; 5)$  considerando diferentes  $\eta$  com  $\varepsilon = 0,001$ ,  $M = 1000$ ,  $p = 0,1$  e  $n = 1000$ . Observa-se, no primeiro gráfico, que, quando  $\eta$  está entre 0,6, 1,5, a média do percentual de censura acerta o valor desejado com precisão de 1 casa decimal. E, à medida que  $\eta$  cresce, a média vai ficando mais distante dos 10% e passa a oscilar em torno desse valor.

No segundo gráfico (Figura 5.5), nota-se que nos casos em que  $\eta$  retornou percentual de censura mais próximo aos 10% também precisou, em média, de um tempo menor, proporcionalmente, para gerar a simulação.

Dado o exposto, para a distribuição  $T \sim Weibull(2; 5)$ , o  $\eta$  indicado é igual a 0,7 (indicado pela linha em azul em ambos os gráficos), pois retornou em média  $\approx 10,02\%$  de censura e com

<sup>4</sup>ver Apêndice D para os gráficos usados para análise e escolha.

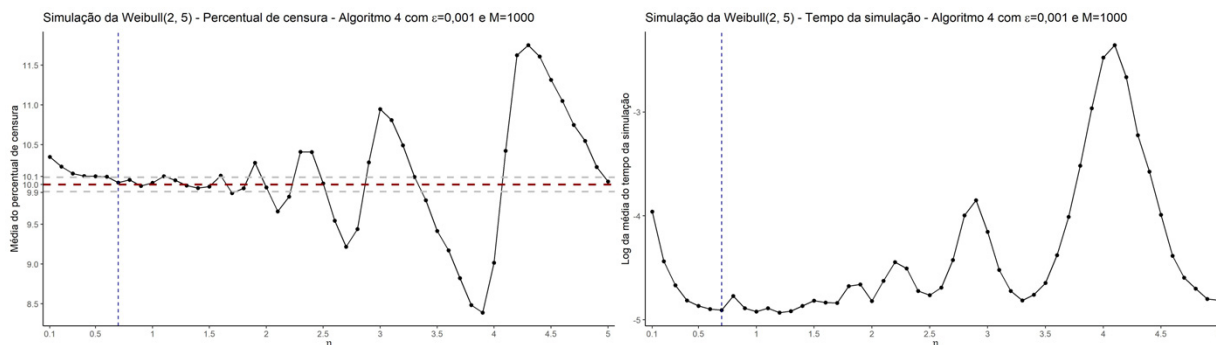


Figura 5.5: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim Weibull(2; 5)$  para diferentes  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

tempo de simulação, proporcionalmente, menor em relação aos demais casos.

A Tabela 5.3 mostra o valor de  $\eta$  escolhido<sup>5</sup> para os parâmetros da distribuição Weibull que são de interesse desse estudo, conforme a lógica utilizada para  $Weibull(2; 5)$ .

Tabela 5.3: Valor de  $\eta$  escolhido para gerar  $\lambda_2$  da distribuição de censura  $C \sim Weibull(\alpha; \lambda_2)$  a partir da distribuição de tempo  $T \sim Weibull(\alpha; \lambda_1)$  para diferentes parâmetros a partir das 1000 simulações do Algoritmo 4 com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

Distribuição de $T$	$\eta$
$Weibull(1; 1)$	0,84
$Weibull(2; 5)$	0,7
$Weibull(7; 15)$	0,5
$Weibull(30; 60)$	0,31

## Log-Normal

A resultado de 1000 simulações do Algoritmo 4 no caso da distribuição de tempo  $T \sim Lognormal(0, 5; 1, 5)$  para diferentes  $\eta$  com  $\varepsilon = 0,001$ ,  $M = 1000$  e  $p = 0,1$  mostra-se nos dois gráficos presentes na Figura 5.6.

A média do percentual de censura obtida para cada  $\eta$  está no primeiro gráfico. Nota-se que para essa distribuição com os parâmetros escolhidos, a média do percentual de censura fica

<sup>5</sup>ver Apêndice D para os gráficos usados para análise e escolha.



mais próximo do valor desejado, com 1 casa decimal de precisão, nos casos em  $\eta$  é pequeno ( $0,07 < \eta < 0,22$ ).

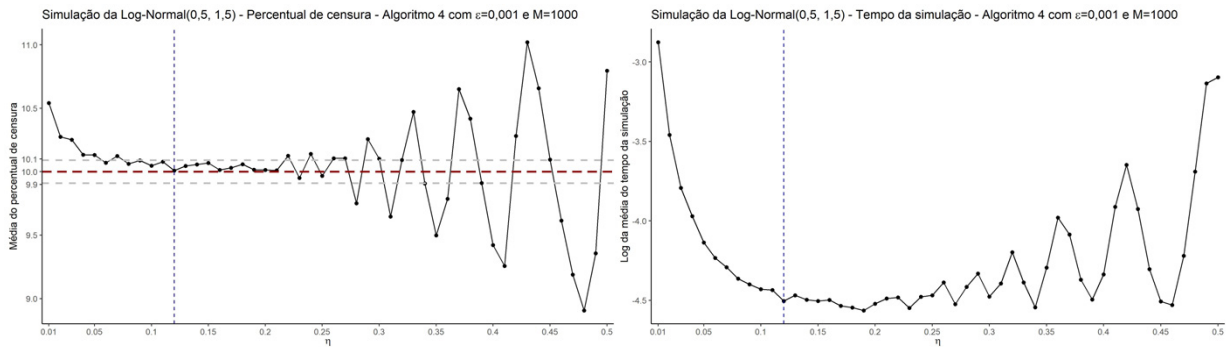


Figura 5.6: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim Lognormal(0, 5; 1, 5)$  para diferentes  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

Para essa distribuição, vale ressaltar, que pequenas mudanças (0,01 para mais ou para menos) no valor de  $\eta$  já altera a precisão do acerto no percentual de censura desejado. Além disso, à medida que  $\eta$  cresce tanto a média fica mais distante do valor desejado, quanto se perde a continuidade no acerto da média, ou seja, o percentual passa a oscilar em valores que ficam muito acima ou abaixo de 10%.

Assim, como nos casos da distribuição Exponencial e Weibull, quando mais distante a média do percentual de censura do valor desejado, maior o tempo necessário para gerar a simulação (segundo gráfico Figura 5.6).

Portanto, uma escolha adequada para o caso  $T \sim Lognormal(0, 5; 1, 5)$  seria  $\eta = 0,12$  (indicado na linha azul em ambos os gráficos) com média do percentual de censura  $\approx 10,009\%$  e com um dos menores tempos de simulação, comparado ao demais.

A Tabela 5.4 apresenta os valores de  $\eta$  escolhidos<sup>6</sup> para os parâmetros da distribuição Log-Normal que são de interesse desse estudo, conforme a lógica utilizada para a  $Lognormal(0, 5; 1, 5)$ .

<sup>6</sup>ver Apêndice D para os gráficos usados para análise e escolha.

Tabela 5.4: Valor de  $\eta$  escolhido para gerar  $\mu_2$  da distribuição de censura  $C \sim \text{Lognormal}(\mu_2; \sigma^2)$  a partir da distribuição de tempo  $T \sim \text{Lognormal}(\mu_1; \sigma^2)$  para diferentes parâmetros a partir das 1000 simulações do Algoritmo 4 com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

Distribuição de $T$	$\eta$
$\text{Lognormal}(0, 1; 1)$	0,12
$\text{Lognormal}(0, 5; 1, 5)$	0,12
$\text{Lognormal}(1; 2)$	0,21
$\text{Lognormal}(1, 5; 2, 5)$	0,23

### Birnbaum-Saunders

A Figura 5.7 apresenta dois gráficos das 1000 simulações do Algoritmo 4 para a distribuição de tempo  $T \sim BS(1, 5; 3, 5)$  com  $\varepsilon = 0,001$ ,  $M = 1000$  e  $p = 0,1$  para diferentes  $\eta$ , sendo o primeiro a respeito da média do percentual de censura e o segundo da média de tempo da simulação.

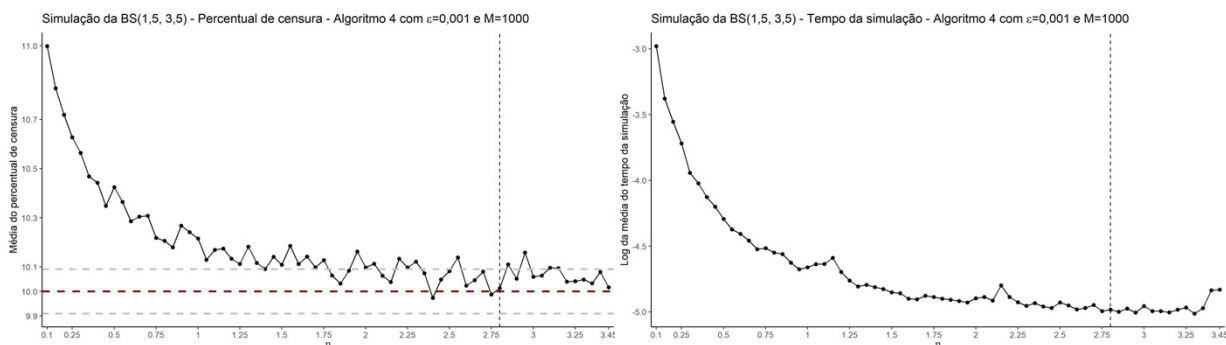


Figura 5.7: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim BS(1, 5; 3, 5)$  para diferentes  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

Observa-se (gráfico 1 - Figura 5.7) que até  $\eta = 1,7$  a média do percentual de censura não atingiu valor desejado com 1 casa decimal de precisão e, esta somente foi alcançada e melhorada à medida que  $\eta$  aumentou. Além disso, a média do tempo da simulação acompanhou os casos em que a média do percentual era mais próxima dos 10% (gráfico 2 - Figura 5.7).

Um valor a ser considerado para simular percentual de censura a partir do Algoritmo 4, usando a distribuição de tempo  $T \sim BS(1, 5; 3, 5)$ , seria  $\eta = 2,8$  (indicado na linha em azul),

devido a média do percentual de censura ser  $\approx 10,01\%$  e a média de tempo da simulação ser um dos menores proporcionalmente, quando comparado aos demais casos.

A Tabela 5.5 apresenta os valores de  $\eta$  escolhidos<sup>7</sup> para os parâmetros da distribuição Birnbaum-Saunders que são de interesse desse estudo, conforme a lógica utilizada para a  $BS(1, 5; 3, 5)$ .

Tabela 5.5: Valor de  $\eta$  escolhido para gerar  $\mu_2$  da distribuição de censura  $C \sim BS(\alpha; \beta_2)$  a partir da distribuição de tempo  $T \sim BS(\alpha; \beta_1)$  para diferentes parâmetros a partir das 1000 simulações do Algoritmo 4 com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

Distribuição de $T$	$\eta$
$BS(0, 5; 1, 5)$	0,19
$BS(1, 5; 3, 5)$	2,8
$BS(1, 5; 5, 5)$	5,2
$BS(1, 5; 7, 5)$	6,6

### 5.4.3 Simulações

A geração de dados censurados do Caso 2 (Subseção 3.3.2) por meio de simulações do Algoritmo 4 é apresentada a seguir. Diferente do Caso 1 (Subseção 3.3.1) que a distribuição de censura,  $C$ , é obtida analiticamente, no Caso 2 é obtida por tentativa e erro, assim, todas as distribuições de interesse desse estudo, independente de possuir ou não solução analítica, podem ser avaliadas.

A Figura 5.8 contém os gráficos com a média do percentual de censura nas 1000 simulações do Algoritmo 4 para  $\varepsilon = 0,001$  e  $M = 1000$  das distribuições Exponencial, Weibull, Log-Normal e Birnbaum-Saunders para diferentes parâmetros e tamanhos de amostra<sup>8</sup>. Em qualquer uma das quatro distribuições, a amostra pequena (de tamanho 10) apresentou o mesmo problema de possuir alta variabilidade ( $CV > 90\%$ ) em torno da média.

Ainda da Figura 5.8, observa-se que nas amostras de tamanho 100 a média do percentual de censura possui uma variabilidade média ( $CV \approx 30\%$ ) e nas amostras de tamanho 1000 essa

<sup>7</sup>ver Apêndice D para os gráficos usados para análise e escolha.

<sup>8</sup>ver Apêndice E para mais detalhes da simulações do Algoritmo 4.

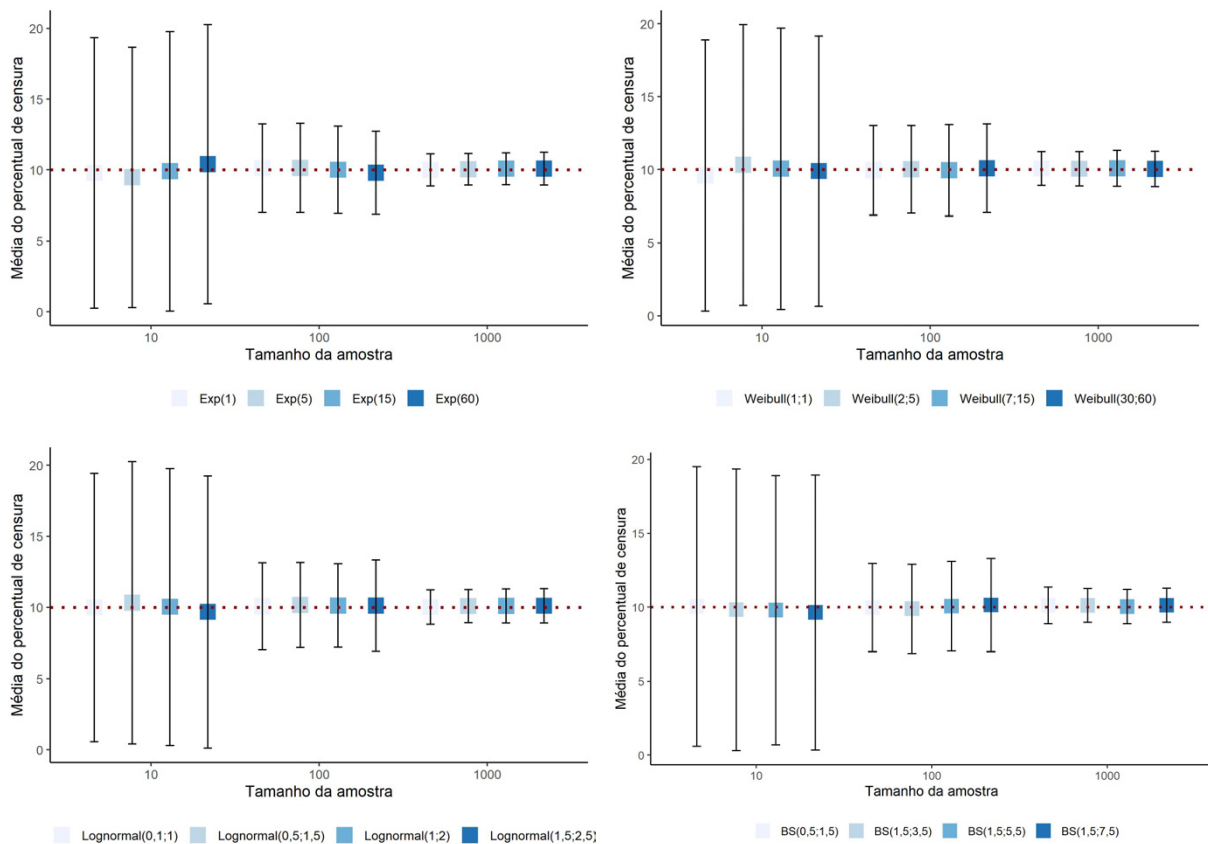


Figura 5.8: Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 2, alternativo numérico (tentativa e erro), - Algoritmo 4 para as distribuições: Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

variabilidade foi pequena ( $CV \approx 11\%$ ).

Vale mencionar que, em qualquer uma das distribuições e parâmetros, o Algoritmo 4 produziu, em média, percentual de censura muito próximo aos 10%.

### 5.5 Simulações do Algoritmo 5 - Censura Aleatória: Caso 3

Nessa seção é apresentada uma forma de gerar censura Aleatória Caso 3, Distribuição de  $C$  diferente de  $T$  feita por simulações do Algoritmo 5. Da mesma maneira que foi realizada a escolha de  $\eta$  adequado para o Algoritmo 4 (censura Aleatória Caso 2) feito na Subseção 5.4.2, na Subseção 5.5.1 exibe-se as informações de cada um dos  $\eta$  escolhidos para o Algoritmo 5.

Por fim, na Subsecção 5.5.2 é apresentado o resultado das simulações desse Algoritmo utilizando os  $\eta$  escolhidos e  $M = 1000$  e  $\varepsilon = 0,001$ .

### 5.5.1 Escolha do $\eta$

A Tabela 5.6 mostra o valor de  $\eta$  escolhido<sup>9</sup> de cada uma das combinações de distribuições de tempo e censura, de forma a obter um percentual de censura próximo ao desejado. Esses valores serão utilizados para gerar as distribuições com o Algoritmo 5 ao longo desse estudo.

Tabela 5.6: Valor de  $\eta$  escolhido para gerar o parâmetro da distribuição de censura  $C$  diferente da distribuição de tempo  $T$  para diferentes parâmetros a partir das 1000 simulações do Algoritmo 5 com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

$\eta$	Distribuição de tempo $T$		
	<i>Weibull</i> (2; 5)	<i>Lognormal</i> (0, 5; 1, 5)	<i>BS</i> (1, 5; 3, 5)
<b>Distribuição de censura <math>C</math></b>			
<i>Exp</i> (5)	4,5	4,6	4,8
<i>Exp</i> (15)	6	5	11,5
<i>Exp</i> (60)	4,5	5	6,5
<i>Weibull</i> (2; 5)	-	1,6	2,5
<i>Weibull</i> (7; 15)	-	0,75	1,5
<i>Weibull</i> (30; 60)	-	1,2	1,8
<i>Lognormal</i> (0, 5; 1, 5)	-	-	0,14
<i>Lognormal</i> (1; 2)	-	-	0,23
<i>Lognormal</i> (1, 5; 2, 5)	-	-	0,24

### 5.5.2 Simulações

A Figura 5.9 mostra o gráfico da média do percentual de censura em 1000 simulações para a distribuição  $T \sim Weibull(2; 5)$  com censura Exponencial com  $\lambda$  igual a 5, 15 e 60 para diferentes tamanhos de amostra.

Para amostras de tamanho 10, independente do valor do parâmetro da Exponencial, a média do percentual de censura ficou abaixo do valor desejado e com alta variabilidade. Nas amostras maiores (100 e 1000) a média conseguiu atingir o valor desejado e o desvio-padrão em torno dessa média diminuiu conforme a amostra aumentou.

<sup>9</sup>ver Apêndice F para os gráficos usados para análise e escolha.

A distribuição de tempo,  $T \sim Lognormal$ , pode ser censurada tanto pela distribuição Exponencial quanto Weibull. No gráfico da Figura 5.10 mostra-se a média do percentual de censura em 1000 simulações para ambos os cenários, com diferentes parâmetros e tamanho de amostra.

Observa-se que  $T \sim Lognormal(0, 5; 1, 5)$  censurada pela Exponencial teve desempenho ligeiramente melhor à censurada pela Weibull, visto que, a média do percentual de censura ficou mais estável e próximo do desejado, mesmo na amostra menor ( $n = 10$ ).

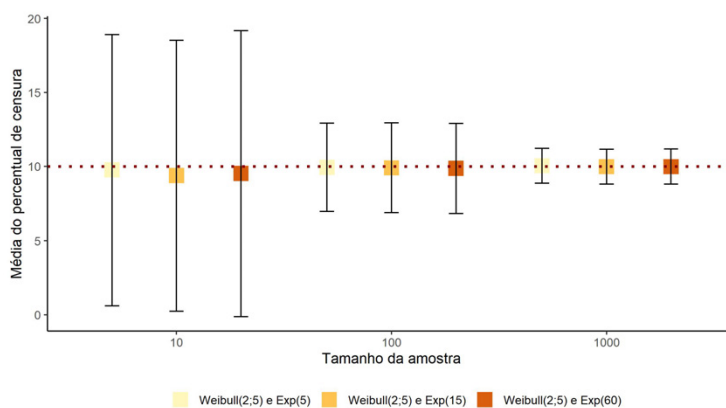


Figura 5.9: Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 para a distribuição de tempo  $T \sim Weibull(2; 5)$  para distribuição de censura,  $C$ , Exponencial, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

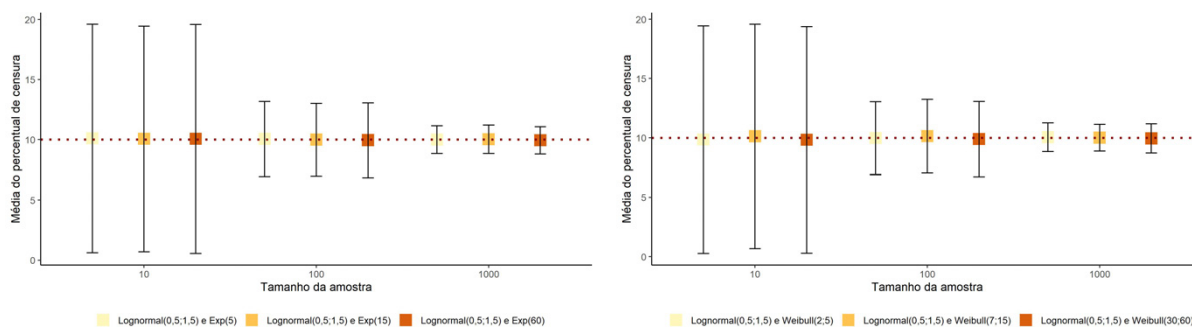


Figura 5.10: Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 para a distribuição de tempo  $T \sim Lognormal(0, 5; 1, 5)$  para as distribuições de censura,  $C$ , Exponencial e Weibull, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

A Figura 5.11 apresenta os gráficos da média do percentual de censura em 1000 simulações

para  $T \sim BS(1, 5; 3, 5)$  com censura Exponencial, Weibull e Log-Normal para diferentes parâmetros nas distribuições de censura e tamanhos de amostra.

Observa-se que  $C \sim Weibull$  destaca-se dos demais casos devido à média do percentual de censura ser instável e ficar acima do desejado nas amostras de tamanho 10 e 100, em especial quando é utilizada a  $C \sim Weibull(30; 60)$  com  $n = 10$ , pois a média percentual de censura foi  $< 3\%$ .

Ainda da Figura 5.11, nota-se que a média do percentual de censura ficou mais próxima do valor desejado quando a censura foi Log-Normal, independente do valor dos parâmetros e do tamanho da amostra.

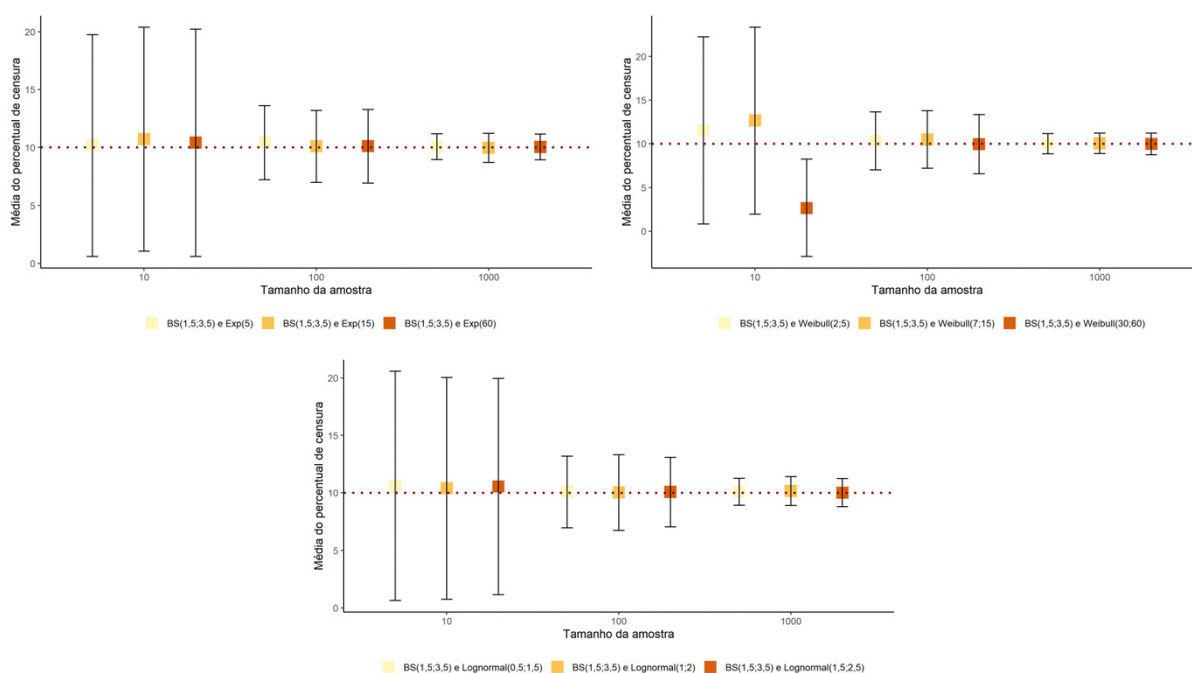


Figura 5.11: Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 para a distribuição de tempo  $T \sim BS(1, 5; 3, 5)$  para as distribuições de censura,  $C$ , Exponencial, Weibull e Log-Normal, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

## 5.6 Simulações do Algoritmo 6 - Censura Aleatória: Caso 4

A Figura 5.12 apresenta os gráficos da censura Aleatória Caso 4 (Subseção 3.3.4), indicador de censura via Bernoulli, por meio de 1000 simulações do Algoritmo 6 para cada uma das distribuições de interesse com diferentes parâmetros e amostras de tamanho 10, 100 e 1000 com percentual de censura desejado de 10% <sup>10</sup>.

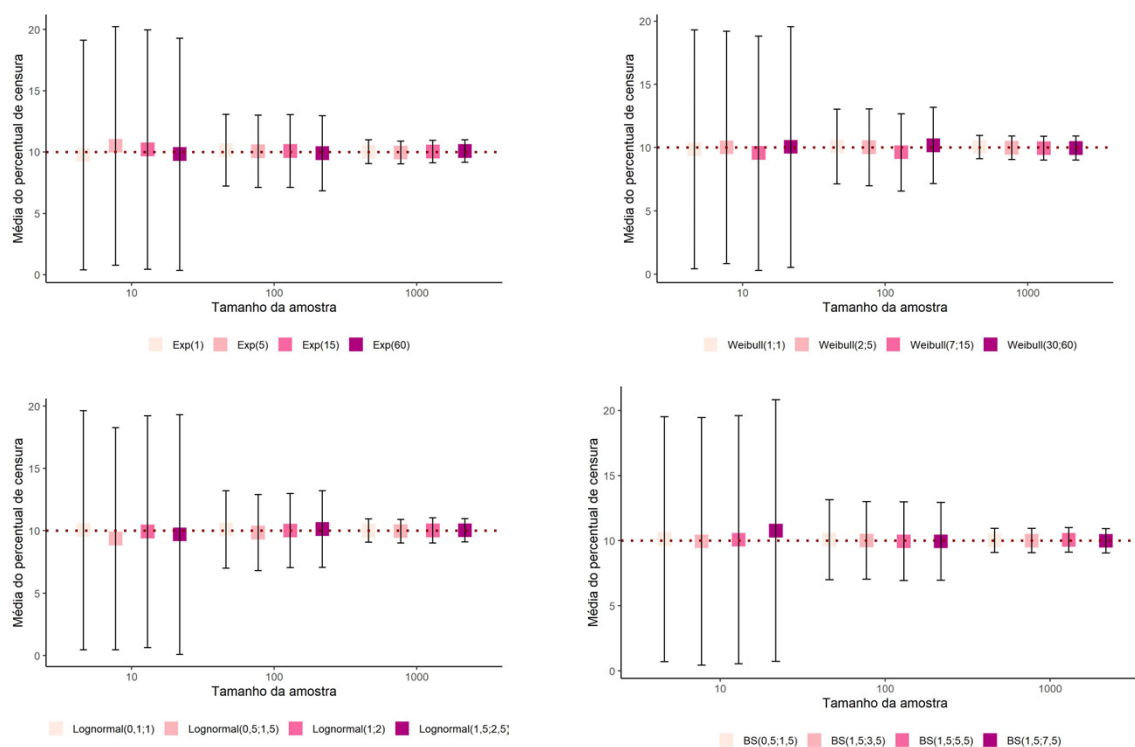


Figura 5.12: Média e desvio-padrão do percentual de censura em 1000 simulações de Censura Aleatória Caso 4, indicador de Censura via Bernoulli - Algoritmo 6 para as distribuições de tempo  $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

Nota-se que a média do percentual de censura, independente da distribuição de tempo, nas amostras de tamanho 10, não atinge o percentual desejado e possui variabilidade alta ( $CV \approx 95\%$ ). O resultado nas amostra  $n = 100$  também é insatisfatório, apesar da variabilidade diminuir ( $CV \approx 30\%$ ) em relação a  $n = 10$ , pois em mais de 50% das simulações o percentual de censura não atingiu o desejado.

<sup>10</sup>ver Apêndice G para mais detalhes da simulações do Algoritmo 6.



Destaca-se (Figura 5.12) que, em amostras grandes ( $n = 1000$ ), a média do percentual de censura é satisfatória, visto que, 50% dos resultados está no intervalo ( $\approx 9,4\%$ ,  $\approx 10,6\%$ ), próximo aos 10% desejado, além disso, a variabilidade observada nesse caso é pequena ( $CV \approx 9,5\%$ ).

## 5.7 Comparação dos métodos

Esse estudo tem como um dos objetivos comparar os métodos de geração de censura à direita propostos nas Seções 3.1, 3.2 e nas Subseções 3.3.1 a 3.3.4. Para isso, nas seções a seguir é realizada a comparação entre os Algoritmos 3 e 4, vício das estimativas e do percentual de censura, além de verificar a divergência entre a distribuição gerada e a teórica.

Segundo as Seções 5.1 a 5.6, o comportamento do percentual de censura obtido, quanto a média e variância, foi similar entre as distribuições de interesse desse estudo, por isso, nas comparações será utilizada somente uma distribuição. Optou-se pela distribuição Weibull, com  $\alpha = 2$  e  $\lambda = 5$ , dado que, no Algoritmo 5, existe apenas uma opção para essa distribuição, Weibull censurada pela Exponencial.

### 5.7.1 Comparação entre os Algoritmo 3 e Algoritmo 4

Nas subseções 3.3.1 e 3.3.2 são apresentados os Caso 1 - Distribuição de  $C$  similar à  $T$  - e Caso 2 - Alternativo numérico (tentativa e erro) sendo a Distribuição de  $C$  similar à  $T$ , dessa forma, é possível comparar esses dois casos e avaliar o comportamento do Algoritmo quando a solução é obtida analiticamente, Algoritmo 3, e quando recorre-se a solução numérica, Algoritmo 4, a fim de verificar qual dos dois é mais eficaz na obtenção do percentual de censura desejado.

A Figura 5.13 apresenta o gráfico da para a distribuição  $Weibull(2; 5)$  gerada, por meio do Algoritmo 3 e do Algoritmo 4 para  $\varepsilon$  igual a 0,1, 0,05, 0,01 e 0,001<sup>11</sup>. Observa-se que

---

<sup>11</sup>Para valores  $\eta < 0,001$  a média do percentual de censura estabiliza no valor de  $p$  desejado.

o desempenho do Algoritmo 3 é superior ao Algoritmo 4 para  $\varepsilon$  igual a 0,1, 0,05 e 0,01, independente do tamanho da amostra, visto que a média do percentual de censura é mais próximo do valor desejado e possui menor variabilidade.

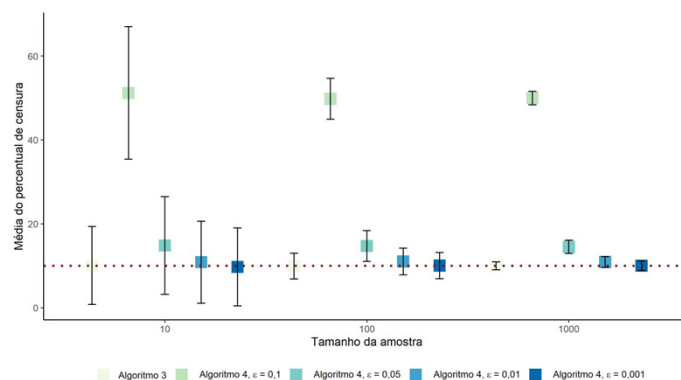


Figura 5.13: Comparação da média do percentual de censura em 1000 simulações dos Algoritmos 3 e Algoritmo 4 para  $\varepsilon$  igual a 0,1, 0,05, 0,01 e 0,001 para as distribuições  $Weibull(2; 5)$  com diferentes tamanhos de amostra

Ainda da Figura 5.13, ao comparar o Algoritmo 3 com o Algoritmo 4 para  $\varepsilon = 0,001$  percebe-se que a média percentual de censura são semelhantes, muito próximo aos 10% e com variabilidade menor, quando comparado aos outros cenários.

A Figura 5.14 mostra o gráfico do tempo decorrido para gerar as 1000 simulações dos Algoritmos 3 e 4 ( $\varepsilon$  igual a 0,1, 0,05, 0,01 e 0,001) da distribuição  $Weibull(2; 5)$ . Ressalta-se que, para facilitar a comparação dos tempos decorridos para gerar as simulações dos casos citados, optou-se em aplicar  $\log$  à média de tempo da simulação.

A média do tempo das simulações do Algoritmo 3 foi menor que a do Algoritmo 4, sendo duas vezes menor em relação ao Algoritmo 4 para  $\varepsilon = 0,001$ , cujo desempenho foi similar.

Portanto, alinhado ao observado no desempenho da média do percentual de censura com a média de tempo da simulação, o Algoritmo 3 parece mais vantajoso frente ao Algoritmo 4. Entretanto, vale lembrar, que o Algoritmo 3 possui limitações, dado que é necessário encontrar a solução analítica dos parâmetros da distribuição de censura  $C$ .

Dessa forma, na ausência de solução analítica, seja qual for o motivo, o Algoritmo 4 com  $\varepsilon = 0,001$  é preferível, uma vez que irá entregar percentual de censura próximo ao valor

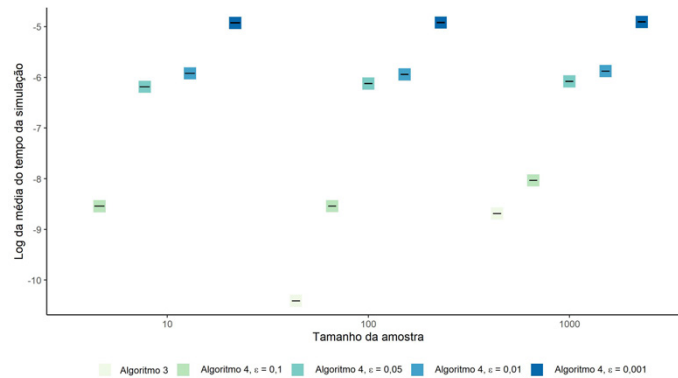


Figura 5.14: Comparação do log da média de tempo da simulação dos Algoritmos 3 e Algoritmo 4 para  $\varepsilon$  igual a 0,1, 0,05, 0,01 e 0,001 para a distribuição  $Weibull(2; 5)$  com diferentes tamanhos de amostra

desejado, sendo necessário somente um tempo adicional na simulação.

### 5.7.2 Vício dos parâmetros

A Tabela 5.7 mostra o vício absoluto dos parâmetros  $\alpha$  e  $\lambda$  em 1000 simulações da distribuição  $Weibull(2; 5)$  para os Algoritmos 1 a 6 para diferentes tamanhos de amostras e percentuais de censura.

Tabela 5.7: Vício absoluto ( $|B(\hat{\theta})|$ ) dos parâmetros  $\alpha$  e  $\lambda$  em 1000 simulações da distribuição  $Weibull(2; 5)$  para os Algoritmos 1 a 6 com amostra ( $n$ ) de tamanho igual a 10, 100 e 1000 e percentual ( $p$ ) de censura igual a 0,1, 0,3 e 0,5

$ B(\hat{\alpha}) ;  B(\hat{\lambda}) $		Algoritmo					
n	p	1	2	3	4	5	6
10	0,1	0,11; 0,24	0,1; 0,12	0,2; 0,22	0,19; 0,24	0,26; 0,25	0,36; 0,29
	0,3	0,3; 1,41	0,32; 0,91	0,08; 1,05	0,05; 1,14	0,15; 0,87	0,42; 1,04
	0,5	0,66; 5,14	0,6; 2,77	0,38; 3,8	0,2; 4,26	0,01; 3,95	1,13; 2,27
100	0,1	0,2; 0,19	0,19; 0,18	0,1; 0,21	0,09; 0,22	0,04; 0,24	0,03; 0,27
	0,3	0,51; 1,02	0,51; 0,99	0,35; 0,99	0,34; 0,94	0,19; 0,81	0,03; 0,98
	0,5	0,79; 3,19	0,78; 3,06	0,62; 2,83	0,61; 2,81	0,48; 3,15	0,04; 2,06
1000	0,1	0,21; 0,19	0,21; 0,18	0,12; 0,21	0,12; 0,22	0,08; 0,24	0,01; 0,28
	0,3	0,53; 1	0,53; 1,01	0,37; 0,97	0,36; 0,95	0,22; 0,81	0; 0,98
	0,5	0,8; 3,14	0,8; 3,12	0,64; 2,74	0,63; 2,75	0,51; 3,02	0,01; 2,07

Observa-se (Tabela 5.7) que o vício absoluto de  $\hat{\alpha}$  e  $\hat{\lambda}$ , tende a crescer em distância do valor original do parâmetro, à medida que o percentual de censura aumenta para qualquer um dos Algoritmos, independente do tamanho da amostra. Contudo, em geral, o vício absoluto do

estimador de  $\alpha$  pode ser considerado pequeno.

O vício absoluto do parâmetro  $\lambda$ ,  $|B(\hat{\lambda})|$ , do Algoritmo 2 para  $p = 0, 1$  foi o menor entre os diferentes Algoritmos, independente do tamanho da amostra. No caso de  $p = 0, 3$ , o Algoritmo 5 apresentou os menores vícios (absolutos) para  $\lambda$  e no maior percentual ( $p = 0, 5$ ) foi o Algoritmo 6 que apresentou menor  $|B(\hat{\lambda})|$ .

Ainda da Tabela 5.7, ressalta-se que  $|B(\hat{\alpha})|$  foi muito próximo de 0 para o Algoritmo 6 nas amostras  $n$  igual a 100 e 1000, independente percentual de censura desejado, além disso, o  $|B(\hat{\alpha})|$  foi menor quando comparado aos demais Algoritmos.

Na Tabela 5.8 são apresentados os valores do Erro Quadrático Médio (EQM) dos parâmetros  $\alpha$  e  $\lambda$  calculados a partir da média de 1000 simulações da distribuição *Weibull*(2; 5) para os Algoritmos 1 a 6 para amostra de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5. Nota-se que o  $EQM(\hat{\lambda})$  tende a aumentar conforme o tamanho da amostra diminui e percentual de censura aumenta.

O  $EQM(\hat{\lambda})$  do Algoritmo 2 foi o menor entre os Algoritmos para  $p = 0, 1$ , independente do valor  $n$ ,  $p = 0, 3$  e  $p = 0, 5$  para amostra de tamanho 10. O Algoritmo 5 apresentou o menor erro quadrático médio para  $\lambda$ , quando comparado aos demais, no cenário de  $p = 0, 3$  e amostra de tamanho 100 e 1000 e, para  $p = 0, 5$ , o Algoritmo 6 mostrou o menor EQM com esses tamanhos de amostra.

Tabela 5.8: Erro Quadrático Médio ( $EQM(\hat{\theta})$ ) dos parâmetros  $\alpha$  e  $\lambda$  em 1000 simulações da distribuição *Weibull*(2; 5) para os Algoritmos 1 a 6 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5

EQM( $\hat{\alpha}$ ); EQM( $\hat{\lambda}$ )		Algoritmo					
n	p	1	2	3	4	5	6
10	0,1	0,57; 1,26	0,39; 0,69	0,51; 0,99	0,53; 0,97	0,62; 1,01	0,81; 0,93
	0,3	0,45; 8,06	0,35; 1,69	0,46; 3,78	0,54; 5,06	0,73; 2,94	1,12; 2,84
	0,5	0,62; 179,68	0,53; 9,68	0,56; 99,1	14,26; 96,46	6,62; 75,13	274,12; 11,13
100	0,1	0,07; 0,14	0,06; 0,11	0,03; 0,14	0,04; 0,15	0,03; 0,14	0,03; 0,16
	0,3	0,28; 1,36	0,27; 1,09	0,14; 1,2	0,14; 1,08	0,06; 0,82	0,04; 1,08
	0,5	0,63; 11,68	0,62; 9,55	0,4; 8,99	0,39; 8,8	0,26; 10,9	0,05; 4,51
1000	0,1	0,05; 0,05	0,05; 0,04	0,02; 0,05	0,02; 0,06	0,01; 0,06	0; 0,08
	0,3	0,28; 1,03	0,28; 1,02	0,14; 0,96	0,13; 0,93	0,05; 0,68	0; 0,97
	0,5	0,64; 9,98	0,64; 9,74	0,41; 7,6	0,4; 7,63	0,26; 9,18	0; 4,31

### 5.7.3 Vício da censura

A Tabela 5.9, mostra os valores da diferença absoluta entre a média do percentual de censura, o percentual de censura desejado ( $|E(\hat{p}) - p|$ ) e o Erro Quadrático Médio do  $\hat{p}$  em 1000 simulações da distribuição *Weibull*(2; 5) de cada um dos Algoritmos (1 a 6) para amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5.

O Algoritmo 2 destaca-se por apresentar  $|E(\hat{p}) - p|$  e  $EQM(\hat{p})$  iguais a 0, independente do valor de  $n$  e  $p$ . Em contra partida, o Algoritmo 5 tem os maiores valores para o vício entre todos os Algoritmos estudados para quase todos os tamanhos de amostra e percentuais de censura desejados, sendo as exceções os casos  $p = 0, 1$ .

Tabela 5.9: Diferença absoluta entre a média do percentual de censura e o percentual de censura desejado ( $|E(\hat{p}) - p|$ ) e Erro Quadrático Médio ( $EQM(\hat{p})$ ) do percentual de censura desejado ( $p$ ) em 1000 simulações da distribuição *Weibull*(2; 5) para os Algoritmos 1 e 3 a 6 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5

$ E(\hat{p}) - p ; EQM(\hat{p})$		Algoritmo*				
n	p	1	3	4	5	6
10	0,1	0,0021; 0,0087	0,0035; 0,009	0,0001; 0,0092	0,0012; 0,0092	0,0031; 0,0097
	0,3	0,0011; 0,0205	0,0095; 0,0191	0,0057; 0,022	0,0375; 0,0209	0,0096; 0,0221
	0,5	0,0017; 0,0242	0,0122; 0,024	0,0051; 0,0252	0,0387; 0,0258	0,0021; 0,0254
100	0,1	0,0006; 0,0009	0; 0,0008	0,0007; 0,001	0,0006; 0,0009	0,0001; 0,0009
	0,3	0,0009; 0,002	0,0004; 0,0021	0,0061; 0,0022	0,0357; 0,0034	0,0003; 0,002
	0,5	0,0015; 0,0025	0,0018; 0,0025	0,0012; 0,0026	0,0508; 0,0052	0,0013; 0,0025
1000	0,1	0,0002; 0,0001	0,0002; 0,0001	0,0012; 0,0001	0; 0,0001	0,0001; 0,0001
	0,3	0,0008; 0,0002	0,0001; 0,0002	0,0048; 0,0003	0,0338; 0,0017	0,0003; 0,0002
	0,5	0,0001; 0,0002	0,0001; 0,0002	0,0006; 0,0003	0,0454; 0,0023	0,0006; 0,0002

\*O Algoritmo 2 foi retirado da tabela por apresentar valores de  $|E(\hat{p}) - p|$  e  $EQM(\hat{p})$  igual a 0 para todos os cenários.

Observa-se que o  $EQM(\hat{p})$  tende a ser maior à medida que a amostra diminui e o percentual de censura aumenta, independente do Algoritmo (com exceção do Algoritmo 2). Todos os algoritmos, com exceção ao 5, geram percentuais de censura próximo ao desejado.

### 5.7.4 Distribuição da amostra dos dados censurados

Como visto nas seções anteriores, a presença de censura induz vício nas estimativas dos parâmetros da distribuição. No entanto, o interesse aqui é verificar se as amostras obtidas

pelos algoritmos apresentados nesse trabalho mantém a mesma distribuição teórica (mesmo que apresente parâmetros diferentes).

Uma forma de avaliar se os dados gerados seguem a mesma (família de) distribuição teórica desejada é verificar, por meio de uma análise gráfica, o ajuste da Função de Sobrevida estimada por Kaplan-Meier e pelo modelo paramétrico.

Conforme, o exposto anteriormente, essa análise foi realizada somente para a distribuição  $Weibull(2; 5)$ , para amostras de tamanho 10, 100 e 1000 e percentuais de censura desejado igual 0,1, 0,3 e 0,5.

Os gráficos com as estimativas da função de sobrevivência estimadas são apresentados nas Figuras H.1 a H.9 do Apêndice H. Em geral, os resultados mostraram que amostras geradas de uma distribuição Weibull por meio dos mecanismos de censura Tipo I e II (Algoritmos 1 e 2) não seguem uma distribuição Weibull, principalmente quando o percentual de censura é grande. Por outro lado, amostras geradas por meio de um mecanismo de censura aleatória (Algoritmos 3 a 6) se comportam segundo a família de distribuição teórica.

## 5.8 Comparação dos métodos - Discreto

Com o objetivo de comparar os métodos de geração de dados discretos de sobrevivência com censura à direita propostos nas Seções 4.1 a 4.3 foi realizada, nas seções a seguir, a comparação do vício das estimativas e do percentual de censura, e, ajuste do modelo entre os Algoritmos 7 a 11.

Vale ressaltar que essa comparação foi feita apenas para a distribuição  $Weibull(\alpha; \lambda)$  discretizada, pois o intuito é avaliar o desempenho dos métodos (Algoritmos) propostos. Além disso, conforme é mostrado na construção dos Algoritmos 7 a 11, os dados são gerados a partir de uma distribuição contínua e discretizados posteriormente. Portanto, para gerar uma distribuição  $WeibullDiscreta(\beta; q)$ , deve-se considerar a reparametrização  $q = \exp\{\frac{-1}{\lambda^\beta}\}$  e  $\beta = \alpha$ , logo,  $Weibull(\beta; \lambda = [\log(1/q)]^{-\frac{1}{\beta}})$  (Biazatti e Nakano, 2020).

### 5.8.1 Vício dos parâmetros

A Tabela 5.10 mostra o resultado do vício ( $|\mathbf{B}(\hat{\theta})|$ ) dos parâmetros  $q$  e  $\beta$  em 1000 simulações da distribuição  $WeibullDiscreta(2; 0, 95)$  para os Algoritmos 7 a 11 para diferentes amostras  $n$  e percentuais de censura  $p$ .

Verifica-se que  $B(\hat{\beta})$  cresce principalmente, com aumento no percentual de censura desejado, porém, também cresce conforme a amostra diminui de tamanho. O Algoritmo 11 destaca-se por apresentar o menor vício para estimativa de  $\beta$  em quase todos os casos, quando comparado aos demais Algoritmos.

Tabela 5.10: Vício ( $|\mathbf{B}(\hat{\theta})|$ ) dos parâmetros  $\beta$  e  $q$  em 1000 simulações da distribuição  $WeibullDiscreta(2; 0, 95)$  para os Algoritmos 7 a 11 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5

$ \mathbf{B}(\hat{\beta}) ;  \mathbf{B}(\hat{q}) $		Algoritmo				
n	p	7	8	9	10	11
10	0,1	0,01; 0,02	0,04; 0,02	0,15; 0	0,15; 0,01	0,31; 0
	0,3	0,47; 0,04	0,45; 0,04	0,23; 0,02	0,26; 0,03	0,43; 0,01
	0,5	0,8; 0,05	0,78; 0,05	0,6; 0,03	0,51; 0,03	0,25; 0,02
100	0,1	0,25; 0,02	0,29; 0,02	0,17; 0,01	0,14; 0,01	0; 0
	0,3	0,64; 0,04	0,63; 0,04	0,45; 0,02	0,45; 0,02	0,01; 0,02
	0,5	0,93; 0,06	0,92; 0,05	0,74; 0,03	0,78; 0,03	0,05; 0,02
1000	0,1	0,29; 0,02	0,3; 0,02	0,18; 0,01	0,17; 0,01	0,02; 0
	0,3	0,66; 0,04	0,65; 0,04	0,48; 0,02	0,46; 0,02	0,05; 0,01
	0,5	0,93; 0,05	0,93; 0,05	0,76; 0,03	0,76; 0,03	0,1; 0,02

Ainda da Tabela 5.10, observa-se que o  $B(\hat{q})$  diminui à medida que o percentual de censura desejado diminui e que o resultado do vício de estimativa de  $q$  foi próximo entre os Algoritmos para os tamanho de amostra e percentuais de censura.

A Tabela 5.11 apresenta os valores do Erro Quadrático Médio ( $EQM(\hat{\theta})$ ) dos parâmetros  $\beta$  e  $q$  em 1000 simulações da distribuição  $WeibullDiscreta(2; 0, 95)$  para os Algoritmos 7 a 11 para diferentes amostras e percentuais de censura. Verifica-se que o  $EQM(\hat{q}) \approx 0$  ou igual a 0 para os valores de  $n$  e  $p$ . Cabe, também, mencionar que o Algoritmo 11 apresenta os menores valores de  $EQM(\hat{\beta})$ , exceto para  $n = 10$ .

Tabela 5.11: Erro Quadrático Médio ( $EQM(\hat{\theta})$ ) dos parâmetros  $\beta$  e  $q$  em 1000 simulações da distribuição  $WeibullDiscreta(2; 0, 95)$  para os Algoritmos 7 a 11 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5

$EQM(\hat{\beta}); EQM(\hat{q})$		Algoritmo				
n	p	7	8	9	10	11
10	0,1	0,53; 0	0,29; 0	0,48; 0	0,53; 0	0,69; 0
	0,3	0,5; 0,01	0,44; 0	0,47; 0	0,47; 0	1; 0
	0,5	0,8; 0,01	0,72; 0	0,66; 0	0,71; 0	0,79; 0
100	0,1	0,09; 0	0,1; 0	0,05; 0	0,05; 0	0,03; 0
	0,3	0,42; 0	0,41; 0	0,22; 0	0,23; 0	0,04; 0
	0,5	0,88; 0	0,85; 0	0,57; 0	0,62; 0	0,06; 0
1000	0,1	0,09; 0	0,09; 0	0,03; 0	0,03; 0	0; 0
	0,3	0,43; 0	0,42; 0	0,23; 0	0,22; 0	0,01; 0
	0,5	0,86; 0	0,86; 0	0,58; 0	0,59; 0	0,01; 0

### 5.8.2 Vício da censura

A diferença absoluta entre a média do percentual de censura e o percentual de censura desejado ( $|E(\hat{p}) - p|$ ) e Erro Quadrático Médio ( $EQM(\hat{\theta})$ ) do percentual de censura desejado em 1000 simulações da distribuição  $WeibullDiscreta(2; 0, 95)$  para os Algoritmos 7 e 9 a 11 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5, é apresentado na Tabela 5.12. O Algoritmo 8 não apresenta diferença entre a média e o percentual de censura desejado para todos os  $n$  e  $p$ , logo, não apresenta  $EQM(\hat{p})$ .

Tabela 5.12: Diferença absoluta entre a média do percentual de censura e o percentual de censura desejado ( $|E(\hat{p}) - p|$ ) e Erro Quadrático Médio ( $EQM(\hat{\theta})$ ) do percentual de censura desejado ( $p$ ) em 1000 simulações da distribuição  $WeibullDiscreta(2; 0, 95)$  para os Algoritmos 7 e 9 a 11 com amostras de tamanho 10, 100 e 1000 e percentuais de censura iguais a 0,1, 0,3 e 0,5

$ E(\hat{p}) - p ; EQM(\hat{p})$		Algoritmo*			
n	p	7	9	10	11
10	0,1	0; 0,0081	0,007; 0,0084	0,006; 0,0089	0,002; 0,0085
	0,3	0,012; 0,02	0,001; 0,0243	0,001; 0,0189	0,008; 0,0214
	0,5	0,003; 0,027	0,014; 0,0206	0,001; 0,0252	0,001; 0,0318
100	0,1	0,0006; 0,0011	0,0037; 0,0008	0,0112; 0,001	0,0027; 0,0008
	0,3	0,0036; 0,0023	0,0025; 0,002	0,0226; 0,0039	0,0009; 0,0022
	0,5	0,012; 0,0021	0,0024; 0,0023	0,0062; 0,0073	0,0094; 0,0027
1000	0,1	0,0004; 0,0001	0,001; 0,0001	0,0117; 0,0004	0,0007; 0,0001
	0,3	0,0021; 0,0002	0,0027; 0,0002	0,0195; 0,0027	0,0011; 0,0002
	0,5	0,0001; 0,0002	0,0026; 0,0003	0,0081; 0,0055	0,0008; 0,0003

\*O Algoritmo 8 foi retirado da tabela por apresentar valores de  $|E(\hat{p}) - p|$  e  $EQM(\hat{p})$  igual a 0 para todos os cenários.



O erro do percentual de censura, assim como o vício das estimativas, diminui à medida que o tamanho da amostra aumenta e o  $p$  desejado diminui. O Algoritmo 11 e 9, destacam-se, por apresentarem a menor diferença absoluta e  $EQM(\hat{p})$  entre os Algoritmos (exceto pelo Algoritmo 8). Porém, vale mencionar que todos os Algoritmos apresentam  $|E(\hat{p}) - p|$  e  $EQM(\hat{p})$  pequeno.

### 5.8.3 Distribuição da amostra dos dados discretos censurados

Para verificar se as amostras discretas obtidas pelos algoritmos (7 a 11) apresentados nesse trabalho mantém a mesma distribuição teórica, foi considerado aqui amostras de tamanho 10, 100 e 1000 de uma distribuição  $WeibullDiscreta(2; 0, 95)$  com percentuais de censura desejado iguais 0,1, 0,3 e 0,5.

Os gráficos com as estimativas da função de sobrevivência estimadas são apresentados nas Figuras I.1 a I.9 do Apêndice I. De forma similar ao caso contínuo, os resultados mostraram que amostras geradas dos mecanismos de censura Tipo I e II (Algoritmos 7 e 8) não mantêm a família de distribuição teórica, principalmente quando o percentual de censura é grande. Por outro lado, amostras geradas por meio de um mecanismo de censura aleatória, Algoritmos 10 a 11 (principalmente), se comportam segundo a família de distribuição teórica.

# Capítulo 6

## Síntese dos resultados

A Tabela 6.1 apresenta a síntese dos resultados da geração de dados, contínuos e discretos, de sobrevivência com censura à direita segundo os Algoritmos 1 a 11, indicando as suas vantagens e desvantagens.

Tabela 6.1: Síntese dos resultados da geração de dados, contínuos e discretos, de sobrevivência com censura à direita para diferentes tipos de censura – Algoritmos 1 a 11

Censura	Vantagens	Desvantagens
<b>Dados Contínuos</b>		
Tipo I - Algoritmo 1	Algoritmo simples para definir o indicador de censura	Amostra não segue a família de distribuição desejada
Tipo II - Algoritmo 2	Algoritmo simples para definir o indicador de censura Não apresenta erro na estimativa do percentual de censura Controle do percentual de censura (melhor controle quanto maior o tamanho da amostra)	Amostra não segue a família de distribuição desejada Necessita da solução analítica de $\bar{P}(\bar{C} < \bar{T})$ Imprecisão no percentual de censura em amostras pequenas.
Aleatória Caso 1 - Algoritmo 3	Amostra mantém família de distribuição desejada Não necessita de solução analítica de $\bar{P}(\bar{C} < \bar{T})$	Necessita definir $\eta$ , $M$ e $\varepsilon$ adequado Tempo adicional para gerar os dados Imprecisão no percentual de censura em amostras pequenas.
Aleatória Caso 2 - Algoritmo 4	Controle do percentual de censura (melhor controle quanto maior o tamanho da amostra) Amostra mantém família de distribuição desejada	
Aleatória Caso 3 - Algoritmo 5	Facilidade na definição do parâmetro da distribuição de censura (principalmente quando a distribuição de $T$ é complexa) Controle do percentual de censura (melhor controle quanto maior o tamanho da amostra)	Necessita definir $\eta$ , $M$ e $\varepsilon$ adequado Imprecisão no percentual de censura em amostras pequenas.
Aleatória Caso 4 - Algoritmo 6	Amostra mantém família de distribuição desejada Algoritmo simples para definição da censura Controle do percentual de censura (melhor controle quanto maior o tamanho da amostra) Amostra mantém família de distribuição desejada	Imprecisão no percentual de censura em amostras pequenas
<b>Dados Discretos</b>		
Tipo I - Algoritmo 7	Algoritmo simples para definir o indicador de censura	Amostra não segue a família de distribuição desejada
Tipo II - Algoritmo 8	Algoritmo simples para definir o indicador de censura Não apresenta erro na estimativa do percentual de censura Controle do percentual de censura (melhor controle quanto maior o tamanho da amostra)	Amostra não segue a família de distribuição desejada Necessita da solução analítica de $\bar{P}(\bar{C} < \bar{T})$ Imprecisão no percentual de censura em amostras pequenas
Aleatória Caso 1 - Algoritmo 9	Amostra mantém família de distribuição desejada Controle do percentual de censura (melhor controle quanto maior o tamanho da amostra)	Necessita da solução analítica de $\bar{P}(\bar{C} < \bar{T})$ Imprecisão no percentual de censura em amostras pequenas
Aleatória Caso 1 - Algoritmo 10	Controle do percentual de censura (melhor controle quanto maior o tamanho da amostra) Amostra mantém família de distribuição desejada	Necessita da solução analítica de $\bar{P}(\bar{C} < \bar{T})$ Imprecisão no percentual de censura em amostras pequenas
Aleatória Caso 4 - Algoritmo 11	Algoritmo simples para definição da censura Controle do percentual de censura (melhor controle quanto maior o tamanho da amostra) Amostra mantém família de distribuição desejada	Imprecisão no percentual de censura em amostras pequenas

# Capítulo 7

## Considerações finais

A simulação de dados de sobrevivência com censura é comum e necessária em diversas situações. Desta forma, responder a questão de qual método é mais adequado em uma situação/cenário era necessária. O objetivo desse trabalho foi produzir uma ferramenta para equipar o pesquisador e torná-lo capaz de decidir qual método de geração de dados com censura produz uma amostra simulada com as características a ele desejáveis.

Dado o exposto ao longo desse estudo, foi possível alcançar este objetivo proposto. Para tanto, inicialmente, foram consideradas a geração de dados de sobrevivência com censura à direita para censuras do Tipo I, Tipo II e Censura Aleatória, por meio de métodos usuais e dos métodos alternativos propostos.

Algoritmos foram apresentados para implementar os métodos propostos. Ao todo foram apresentados neste trabalho 11 algoritmos diferentes, sendo que 6 são para a geração de tempos de sobrevivência contínuos e 5 são destinados para a geração de tempos discretos. O Algoritmo 1 e 2 correspondem a censura Tipo I e Tipo II, respectivamente. Para Censura Aleatória, foram estudados três casos. O Algoritmo 3 corresponde ao caso 1, caso usual, em que a distribuição de tempo,  $T$ , é similar a distribuição de censura,  $C$  e os parâmetros da distribuição de  $C$  são obtidos analiticamente por meio do parâmetro de força-stress,  $R = P(C < T)$ . O Algoritmo 4 representa o caso 2, alternativa (ao caso 1) numérica, no qual os parâmetros da

distribuição de  $C$  são obtidos numericamente. O Algoritmo 5 é desenvolvido para o caso 3, que considera a distribuição de  $T$  diferente de  $C$ , tendo os parâmetros da distribuição de  $C$  obtidas numericamente. A outra proposta alternativa a usual é apresentada pelo Algoritmo 6, caso 4, em que o indicador de censura é obtido via distribuição Bernoulli.

Em seguida, foram realizadas 1000 simulações de cada um dos Algoritmos (1 a 6) para as distribuições Exponencial, Weibull, Log-Normal e Birnbaum-Saunders. Conclui-se, nessa parte, que todos os Algoritmos, independente da distribuição de tempo, conseguiram obter, em média, percentual de censura próximo ao percentual de censura desejado e a variância em torno dessa média diminui conforme a amostra aumenta.

Na sequência, foi apresentada a comparação entre os métodos de geração de dados de sobrevivência com censura por meio do vício dos parâmetros, vício da censura e ajuste do modelo de distribuição da amostra dos dados censurados. Constata-se que o Algoritmo 3 apresenta o menor erro na estimativa do percentual de censura e a amostra gerada mantém as características da família de distribuição desejada. Ainda, o Algoritmo 6 apresentou um dos menores erros na estimativa do percentual de censura e também manteve a família da distribuição desejada. Porém, tanto o Algoritmo 3 quanto 6 apresentaram imprecisão no percentual de censura em amostras pequenas. Verifica-se nos Algoritmos 1 e 2, que, apesar de possuírem algoritmo simples para definir o indicador de censura, a amostra gerada não segue a família da distribuição desejada, principalmente quando percentual de censura desejado é grande.

Por último, foi realizada a adaptação de alguns dos métodos estudados a dados discretos de sobrevivência censurados. A adaptação foi realizada discretizando a distribuição contínua com os métodos já propostos nesse estudo. Logo, os Algoritmos 7 e 8 foram obtidos adaptando os Algoritmos 1 e 2, respectivamente. O Algoritmo 9 é similar ao Algoritmo 3. O Algoritmo 10 também é similar ao Algoritmo 3, porém a decisão de censura é feita via Bernoulli. E, o Algoritmo 11 é o caso discretizado do Algoritmo 6. Assim como o observado no caso contínuo, nos Algoritmos 7 e 8 a amostra gerada não mantém a família de distribuição desejada. Destaca-

se que o Algoritmo 11 apresenta menor erro na estimativa do percentual de censura, a amostra segue a família de distribuição desejada e a construção do algoritmo é simples para definição de censura.

Os resultados e conclusões desse estudo, foram sintetizados na Tabela 6.1, de forma a fornecer ao pesquisador uma ferramenta de consulta rápida e objetiva das vantagens e desvantagens encontradas em cada um dos métodos de geração de dados (contínuos e discretos) de sobrevivência com censura à direita, permitindo-o decidir qual método cumpre os requisitos da amostra que deseja gerar.

Com base os resultados obtidos nesse observou-se que:

- o vício das estimativas dos parâmetros cresce quanto maior for o percentual de censura desejado;
- com exceção do mecanismo de censura do Tipo II, o controle (a precisão) do percentual de censura na amostra cresce, quanto maior for o tamanho da amostra;
- os mecanismos de censura do Tipo I e II se mostraram procedimentos simples de gerar dados censurados, porém, os mesmos têm a desvantagem da distribuição dos dados da amostra divergir da família de distribuição teórica adotada para a geração dos dados.

Desta forma, conclui-se que o mecanismo de censura que apresenta as melhores características desejadas (controle de percentual de censura e amostras seguindo a distribuição desejada) é o aleatório.

Ademais, visto que a maior dificuldade de implementar a geração de dados segundo um mecanismo de censura aleatório é a especificação da distribuição do tempo latente de censura (principalmente quando a distribuição de interesse é complexa), este trabalho também apresentou uma forma alternativa em que o indicador de censura é obtido por meio de uma distribuição de Bernoulli independente da distribuição do tempo de sobrevivência.

Os resultados obtidos nesse trabalho mostraram que o mecanismo de geração do indicador de censura por meio de uma distribuição Bernoulli é um método simples que também satisfaz as

características desejadas (controle de percentual de censura e amostras seguindo a distribuição desejada), sendo considerado, neste trabalho, a melhor e mais simples metodologia de geração de dados com censura à direita.

# **Apêndice A**

## **Simulações do Algoritmo 1**



Tabela A.1: Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Tipo I - Algoritmo 1 para as distribuições de tempo  $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

Distribuição	$n$	$E(\hat{p})$	$DP(\hat{p})$	$CV(\hat{p})$	Quantis			Média de tempo em segundos por simulação
					1°	Mediana	3°	
Exp(1)	10	0,1034	0,1006	0,9728	0,0000	0,1000	0,2000	0,00003
Exp(5)	10	0,1028	0,0975	0,9482	0,0000	0,1000	0,2000	0,00002
Exp(15)	10	0,0988	0,0941	0,9520	0,0000	0,1000	0,2000	0,00001
Exp(60)	10	0,0962	0,0922	0,9580	0,0000	0,1000	0,2000	0,00001
Exp(1)	100	0,0985	0,0311	0,3154	0,0800	0,1000	0,1200	0,00002
Exp(5)	100	0,0982	0,0295	0,2999	0,0800	0,1000	0,1200	0,00002
Exp(15)	100	0,0989	0,0307	0,3100	0,0800	0,1000	0,1200	0,00002
Exp(60)	100	0,1006	0,0297	0,2949	0,0800	0,1000	0,1200	0,00002
Exp(1)	1000	0,1002	0,0095	0,0950	0,0940	0,1000	0,1060	0,00006
Exp(5)	1000	0,1004	0,0095	0,0949	0,0930	0,1010	0,1070	0,00006
Exp(15)	1000	0,0995	0,0094	0,0947	0,0930	0,1000	0,1053	0,00006
Exp(60)	1000	0,1000	0,0091	0,0913	0,0940	0,1000	0,1060	0,00006
Weibull(1;1)	10	0,1030	0,0963	0,9353	0,0000	0,1000	0,2000	0,00004
Weibull(2;5)	10	0,0994	0,0947	0,9527	0,0000	0,1000	0,2000	0,00002
Weibull(7;15)	10	0,0970	0,0943	0,9726	0,0000	0,1000	0,1000	0,00002
Weibull(30;60)	10	0,1000	0,0913	0,9126	0,0000	0,1000	0,2000	0,00002
Weibull(1;1)	100	0,1007	0,0295	0,2933	0,0800	0,1000	0,1200	0,00002
Weibull(2;5)	100	0,1008	0,0302	0,2999	0,0800	0,1000	0,1200	0,00002
Weibull(7;15)	100	0,1001	0,0300	0,2999	0,0800	0,1000	0,1200	0,00003
Weibull(30;60)	100	0,1007	0,0300	0,2984	0,0800	0,1000	0,1200	0,00003
Weibull(1;1)	1000	0,0998	0,0095	0,0954	0,0938	0,1000	0,1060	0,00011
Weibull(2;5)	1000	0,0996	0,0096	0,0961	0,0930	0,1000	0,1060	0,00010
Weibull(7;15)	1000	0,1003	0,0097	0,0970	0,0940	0,1000	0,1070	0,00017
Weibull(30;60)	1000	0,1000	0,0096	0,0957	0,0930	0,1000	0,1060	0,00013
Lognormal(0,1;1)	10	0,1020	0,0988	0,9688	0,0000	0,1000	0,2000	0,00003
Lognormal(0,5;1,5)	10	0,1019	0,0957	0,9390	0,0000	0,1000	0,2000	0,00002
Lognormal(1;2)	10	0,1018	0,0987	0,9698	0,0000	0,1000	0,2000	0,00002
Lognormal(1,5;2,5)	10	0,0979	0,0928	0,9480	0,0000	0,1000	0,2000	0,00002
Lognormal(0,1;1)	100	0,1003	0,0303	0,3023	0,0800	0,1000	0,1200	0,00003
Lognormal(0,5;1,5)	100	0,0993	0,0293	0,2951	0,0800	0,1000	0,1200	0,00003
Lognormal(1;2)	100	0,1009	0,0314	0,3113	0,0800	0,1000	0,1200	0,00003
Lognormal(1,5;2,5)	100	0,1004	0,0301	0,2997	0,0800	0,1000	0,1200	0,00003
Lognormal(0,1;1)	1000	0,1004	0,0092	0,0917	0,0940	0,1010	0,1070	0,00011
Lognormal(0,5;1,5)	1000	0,0999	0,0096	0,0964	0,0940	0,1000	0,1060	0,00012
Lognormal(1;2)	1000	0,1001	0,0096	0,0963	0,0930	0,1000	0,1070	0,00011
Lognormal(1,5;2,5)	1000	0,1004	0,0095	0,0944	0,0940	0,1000	0,1060	0,00012
BS(0,5;1,5)	10	0,0966	0,0979	1,0131	0,0000	0,1000	0,2000	0,00029
BS(1,5;3,5)	10	0,0958	0,0913	0,9527	0,0000	0,1000	0,1000	0,00024
BS(1,5;5,5)	10	0,0976	0,0949	0,9722	0,0000	0,1000	0,2000	0,00026
BS(1,5;7,5)	10	0,0953	0,0922	0,9672	0,0000	0,1000	0,2000	0,00021
BS(0,5;1,5)	100	0,0989	0,0294	0,2974	0,0800	0,1000	0,1200	0,00021
BS(1,5;3,5)	100	0,0994	0,0303	0,3047	0,0800	0,1000	0,1200	0,00021
BS(1,5;5,5)	100	0,0996	0,0299	0,3002	0,0800	0,1000	0,1200	0,00020
BS(1,5;7,5)	100	0,1001	0,0295	0,2949	0,0800	0,1000	0,1200	0,00019
BS(0,5;1,5)	1000	0,1002	0,0091	0,0911	0,0940	0,1000	0,1070	0,00027
BS(1,5;3,5)	1000	0,1002	0,0094	0,0940	0,0940	0,1000	0,1063	0,00026
BS(1,5;5,5)	1000	0,1000	0,0095	0,0949	0,0940	0,0995	0,1060	0,00026
BS(1,5;7,5)	1000	0,0996	0,0094	0,0940	0,0930	0,0990	0,1060	0,00028

# **Apêndice B**

## **Simulações do Algoritmo 2**

Tabela B.1: Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Tipo II - Algoritmo 2 para as distribuições de tempo  $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

Distribuição	$n$	$E(\hat{p})$	$DP(\hat{p})$	$CV(\hat{p})$	Quantis			Média de tempo em segundos por simulação
					1°	Mediana	3°	
Exp(1)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00029
Exp(5)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00024
Exp(15)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00026
Exp(60)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Exp(1)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Exp(5)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Exp(15)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00020
Exp(60)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00019
Exp(1)	1000	0,1	0	0,1	0,1	0,1	0,0001	0,00027
Exp(5)	1000	0,1	0	0,1	0,1	0,1	0,0001	0,00026
Exp(15)	1000	0,1	0	0,1	0,1	0,1	0,0001	0,00026
Exp(60)	1000	0,1	0	0,1	0,1	0,1	0,0001	0,00028
Weibull(1;1)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00029
Weibull(2;5)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00024
Weibull(7;15)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00026
Weibull(30;60)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Weibull(1;1)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Weibull(2;5)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Weibull(7;15)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00020
Weibull(30;60)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00019
Weibull(1;1)	1000	0,1	0	0,1	0,1	0,1	0,0002	0,00027
Weibull(2;5)	1000	0,1	0	0,1	0,1	0,1	0,0002	0,00026
Weibull(7;15)	1000	0,1	0	0,1	0,1	0,1	0,0002	0,00026
Weibull(30;60)	1000	0,1	0	0,1	0,1	0,1	0,0002	0,00028
Lognormal(0,1;1)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00029
Lognormal(0,5;1,5)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00024
Lognormal(1;2)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00026
Lognormal(1,5;2,5)	10	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Lognormal(0,1;1)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Lognormal(0,5;1,5)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00021
Lognormal(1;2)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00020
Lognormal(1,5;2,5)	100	0,1	0	0,1	0,1	0,1	0,0001	0,00019
Lognormal(0,1;1)	1000	0,1	0	0,1	0,1	0,1	0,0002	0,00027
Lognormal(0,5;1,5)	1000	0,1	0	0,1	0,1	0,1	0,0002	0,00026
Lognormal(1;2)	1000	0,1	0	0,1	0,1	0,1	0,0002	0,00026
Lognormal(1,5;2,5)	1000	0,1	0	0,1	0,1	0,1	0,0002	0,00028
BS(0,5;1,5)	10	0,1	0	0,1	0,1	0,1	0,0004	0,00029
BS(1,5;3,5)	10	0,1	0	0,1	0,1	0,1	0,0003	0,00024
BS(1,5;5,5)	10	0,1	0	0,1	0,1	0,1	0,0003	0,00026
BS(1,5;7,5)	10	0,1	0	0,1	0,1	0,1	0,0003	0,00021
BS(0,5;1,5)	100	0,1	0	0,1	0,1	0,1	0,0003	0,00021
BS(1,5;3,5)	100	0,1	0	0,1	0,1	0,1	0,0002	0,00021
BS(1,5;5,5)	100	0,1	0	0,1	0,1	0,1	0,0002	0,00020
BS(1,5;7,5)	100	0,1	0	0,1	0,1	0,1	0,0003	0,00019
BS(0,5;1,5)	1000	0,1	0	0,1	0,1	0,1	0,0004	0,00027
BS(1,5;3,5)	1000	0,1	0	0,1	0,1	0,1	0,0003	0,00026
BS(1,5;5,5)	1000	0,1	0	0,1	0,1	0,1	0,0004	0,00026
BS(1,5;7,5)	1000	0,1	0	0,1	0,1	0,1	0,0004	0,00028

# **Apêndice C**

## **Simulações do Algoritmo 3**

Tabela C.1: Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Aleatória Caso 1; distribuição de  $C$  similar à  $T$  - Algoritmo 3 para as distribuições: Exponencial, Weibull e Log-Normal e , com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

Distribuição	$n$	$E(\hat{p})$	$DP(\hat{p})$	$CV(\hat{p})$	Quantis			Média de tempo em segundos por simulação
					1°	Mediana	3°	
Exp(1)	10	0,1046	0,0964	0,9214	0,0000	0,1000	0,2000	0,00003
Exp(5)	10	0,0996	0,0981	0,9852	0,0000	0,1000	0,2000	0,00002
Exp(15)	10	0,0965	0,0929	0,9625	0,0000	0,1000	0,1000	0,00002
Exp(60)	10	0,1028	0,0949	0,9229	0,0000	0,1000	0,2000	0,00002
Exp(1)	100	0,1002	0,0303	0,3022	0,0800	0,1000	0,1200	0,00004
Exp(5)	100	0,1010	0,0303	0,3000	0,0800	0,1000	0,1200	0,00003
Exp(15)	100	0,0988	0,0303	0,3070	0,0800	0,1000	0,1200	0,00003
Exp(60)	100	0,1008	0,0309	0,3065	0,0800	0,1000	0,1200	0,00003
Exp(1)	1000	0,1003	0,0098	0,0978	0,0940	0,1000	0,1063	0,00012
Exp(5)	1000	0,0999	0,0095	0,0951	0,0930	0,1000	0,1060	0,00012
Exp(15)	1000	0,1001	0,0093	0,0933	0,0940	0,1000	0,1060	0,00010
Exp(60)	1000	0,1000	0,0094	0,0939	0,0940	0,1000	0,1060	0,00011
Weibull(1;1)	10	0,0967	0,0943	0,9749	0,0000	0,1000	0,1250	0,00003
Weibull(2;5)	10	0,0979	0,0948	0,9687	0,0000	0,1000	0,1000	0,00002
Weibull(7;15)	10	0,0950	0,0926	0,9752	0,0000	0,1000	0,1000	0,00002
Weibull(30;60)	10	0,0993	0,0968	0,9753	0,0000	0,1000	0,2000	0,00002
Weibull(1;1)	100	0,1005	0,0303	0,3015	0,0800	0,1000	0,1200	0,00003
Weibull(2;5)	100	0,1001	0,0295	0,2945	0,0800	0,1000	0,1200	0,00003
Weibull(7;15)	100	0,1000	0,0308	0,3078	0,0800	0,1000	0,1200	0,00005
Weibull(30;60)	100	0,0997	0,0293	0,2939	0,0800	0,1000	0,1200	0,00005
Weibull(1;1)	1000	0,1001	0,0096	0,0963	0,0940	0,1000	0,1060	0,00018
Weibull(2;5)	1000	0,0997	0,0097	0,0972	0,0930	0,0990	0,1060	0,00019
Weibull(7;15)	1000	0,1003	0,0096	0,0955	0,0940	0,1000	0,1070	0,00031
Weibull(30;60)	1000	0,1000	0,0094	0,0943	0,0930	0,0995	0,1060	0,00030
Lognormal(0,1;1)	10	0,1051	0,0991	0,9426	0,0000	0,1000	0,2000	0,00004
Lognormal(0,5;1,5)	10	0,0997	0,0965	0,9683	0,0000	0,1000	0,2000	0,00002
Lognormal(1;2)	10	0,0982	0,0932	0,9490	0,0000	0,1000	0,2000	0,00002
Lognormal(1,5;2,5)	10	0,1046	0,0954	0,9124	0,0000	0,1000	0,2000	0,00002
Lognormal(0,1;1)	100	0,0993	0,0303	0,3056	0,0800	0,1000	0,1200	0,00004
Lognormal(0,5;1,5)	100	0,1016	0,0306	0,3008	0,0800	0,1000	0,1200	0,00004
Lognormal(1;2)	100	0,0994	0,0294	0,2959	0,0800	0,1000	0,1200	0,00004
Lognormal(1,5;2,5)	100	0,1008	0,0298	0,2956	0,0800	0,1000	0,1200	0,00004
Lognormal(0,1;1)	1000	0,1003	0,0096	0,0958	0,0940	0,1000	0,1070	0,00022
Lognormal(0,5;1,5)	1000	0,0996	0,0096	0,0962	0,0930	0,1000	0,1060	0,00020
Lognormal(1;2)	1000	0,0998	0,0093	0,0934	0,0940	0,1000	0,1060	0,00023
Lognormal(1,5;2,5)	1000	0,0995	0,0097	0,0978	0,0930	0,0990	0,1060	0,00023

## **Apêndice D**

### **Algoritmo 4 - Etas escolhidos**

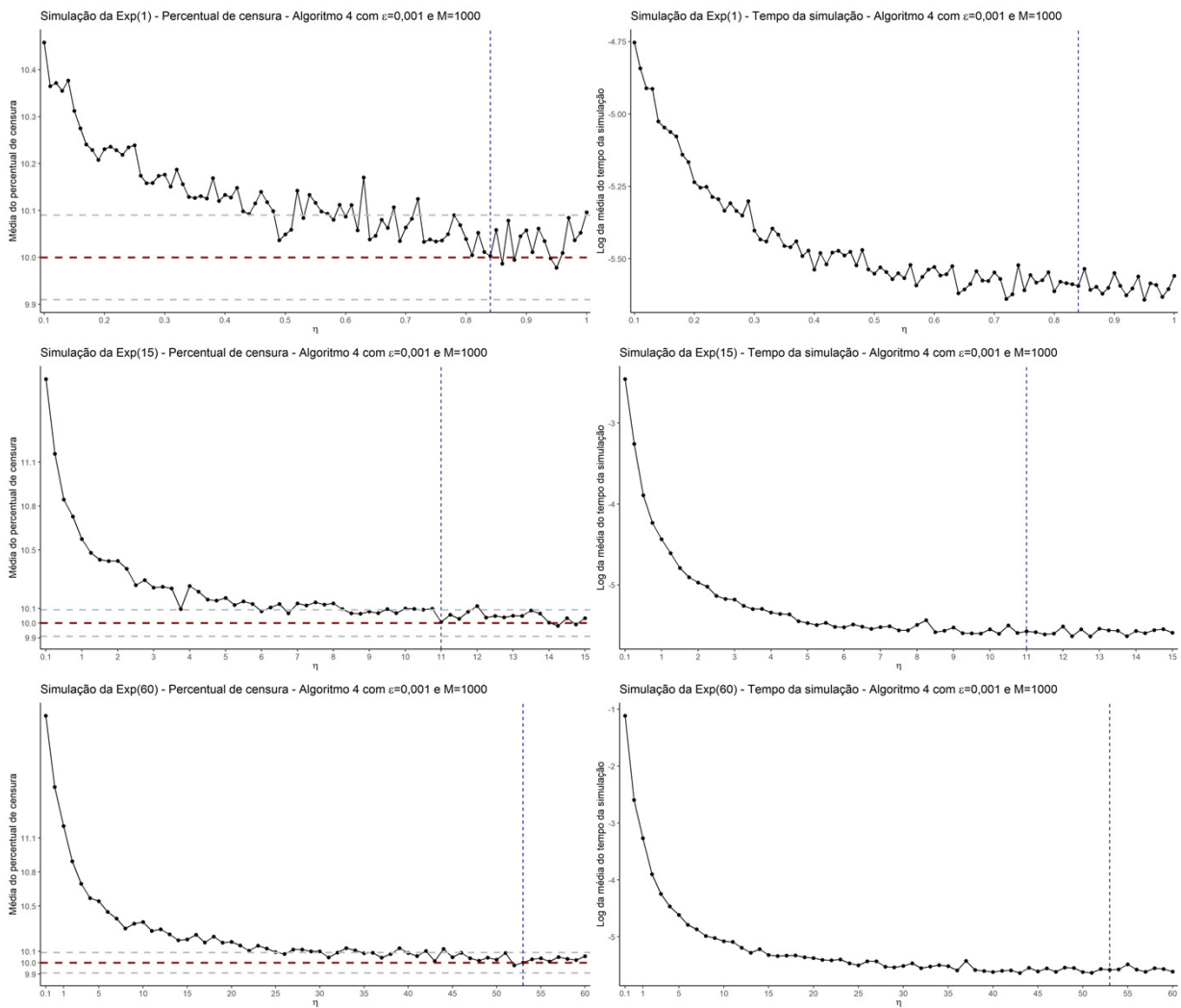


Figura D.1: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim Exp$  para diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

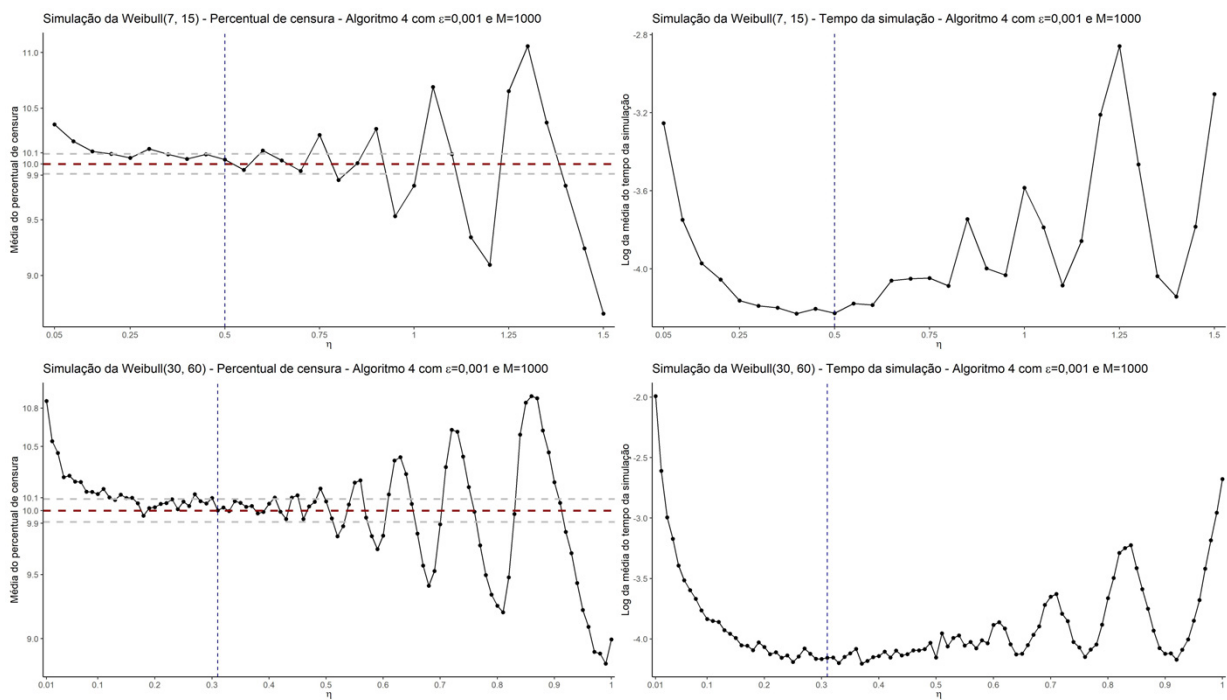


Figura D.2: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim Weibull$  para diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$



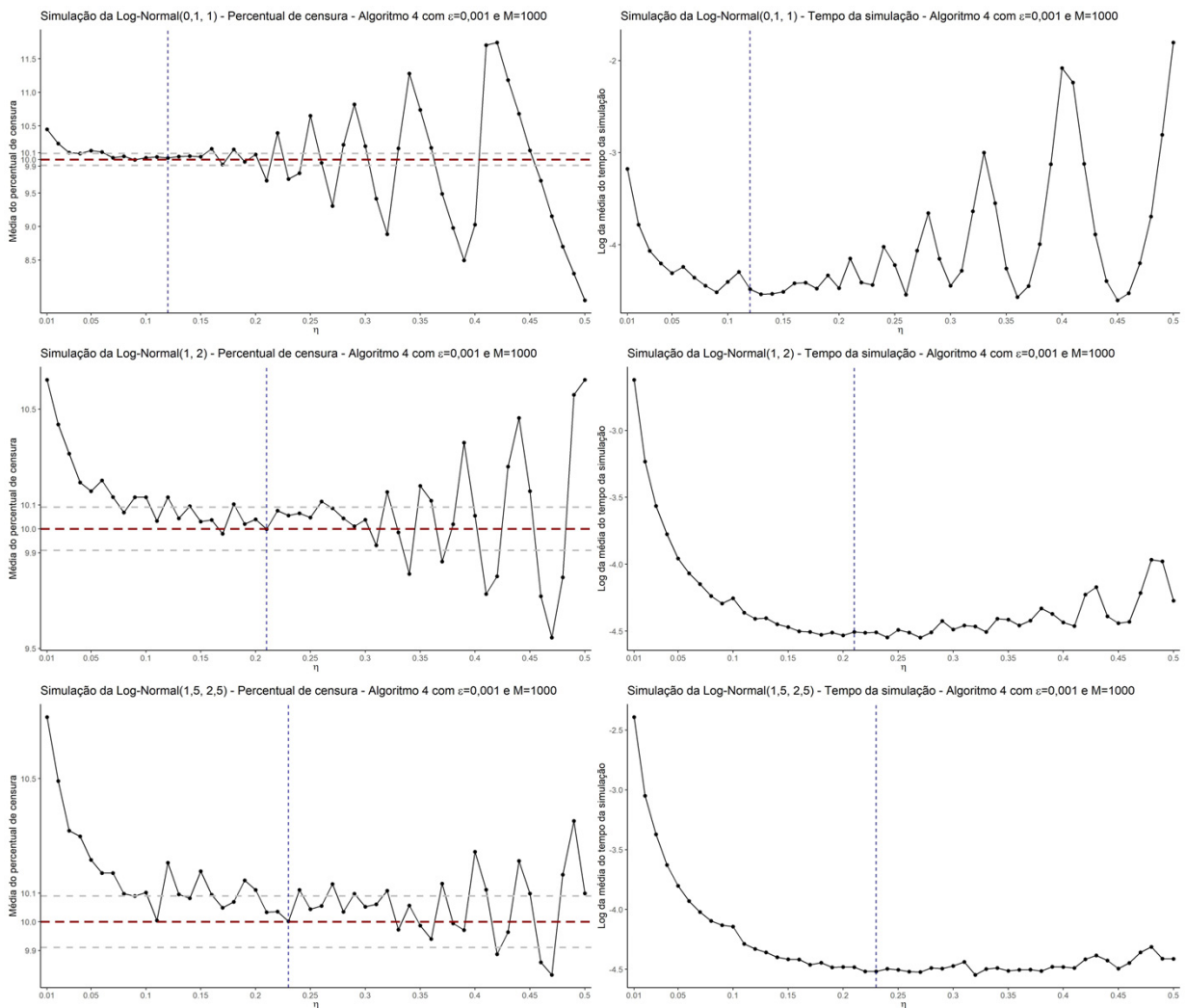


Figura D.3: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim Lognormal$  para diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

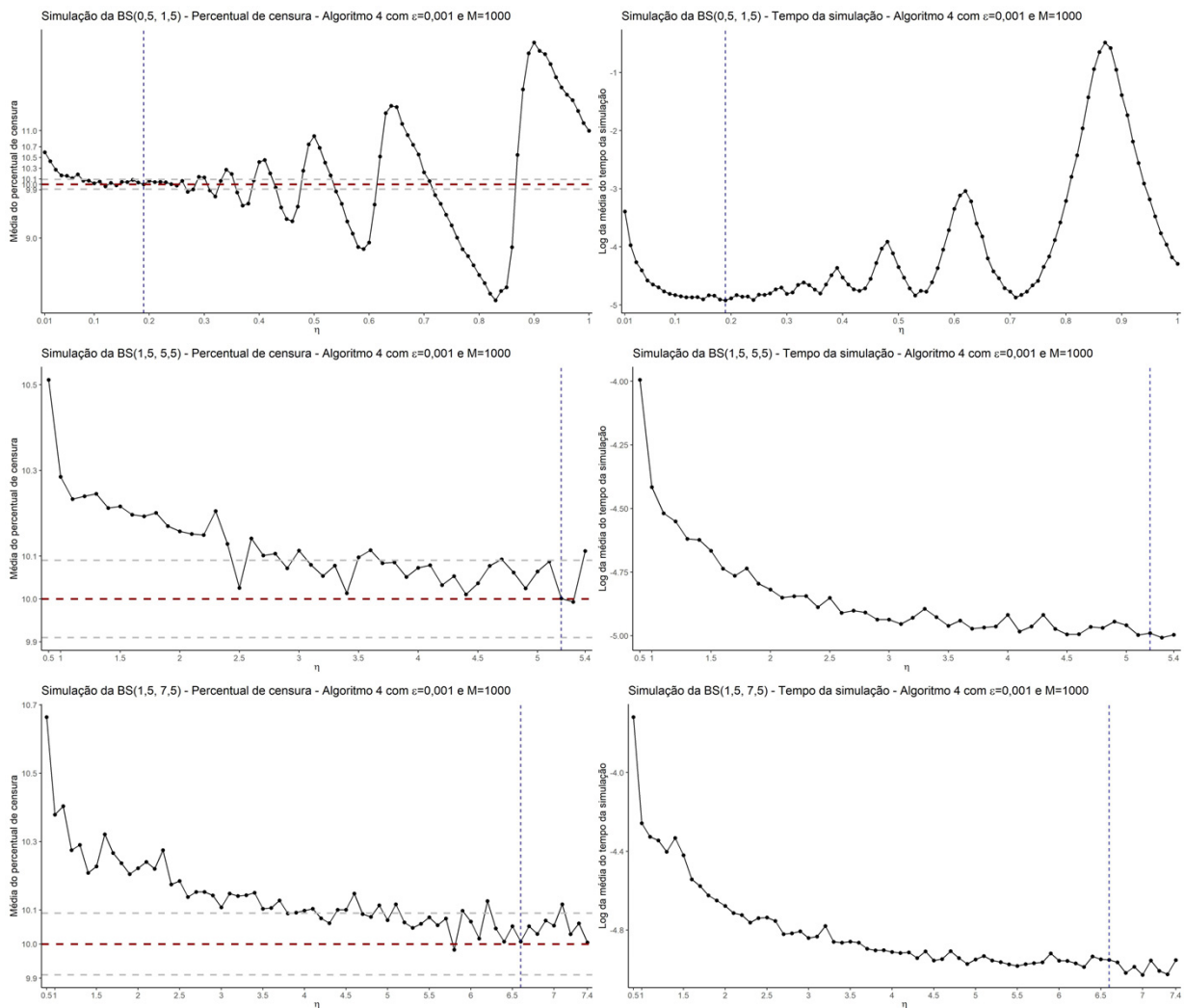


Figura D.4: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 2, alternativo numérico - Algoritmo 4 da distribuição de tempo  $T \sim BS$  para diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

# **Apêndice E**

## **Simulações do Algoritmo 4**

Tabela E.1: Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Aleatória Caso 2, alternativo numérico (tentativa e erro), - Algoritmo 4 para as distribuições: Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

Distribuição	$n$	$E(\hat{p})$	$DP(\hat{p})$	$CV(\hat{p})$	Quantis			Média de tempo em segundos por simulação
					1°	Mediana	3°	
Exp(1)	10	0,0979	0,0955	0,9752	0,0000	0,1000	0,2000	0,00352
Exp(5)	10	0,0948	0,0919	0,9692	0,0000	0,1000	0,1000	0,00353
Exp(15)	10	0,0991	0,0986	0,9948	0,0000	0,1000	0,2000	0,00351
Exp(60)	10	0,1041	0,0985	0,9462	0,0000	0,1000	0,2000	0,00354
Exp(1)	100	0,1012	0,0313	0,3093	0,0800	0,1000	0,1200	0,00355
Exp(5)	100	0,1014	0,0315	0,3102	0,0800	0,1000	0,1200	0,00349
Exp(15)	100	0,1001	0,0309	0,3085	0,0800	0,1000	0,1200	0,00366
Exp(60)	100	0,0979	0,0294	0,2997	0,0800	0,1000	0,1200	0,00348
Exp(1)	1000	0,1000	0,0113	0,1129	0,0920	0,0990	0,1080	0,00373
Exp(5)	1000	0,1004	0,0111	0,1105	0,0930	0,1000	0,1080	0,00363
Exp(15)	1000	0,1008	0,0112	0,1111	0,0940	0,1010	0,1080	0,00359
Exp(60)	1000	0,1009	0,0116	0,1154	0,0930	0,1010	0,1090	0,00386
Weibull(1;1)	10	0,0961	0,0929	0,9663	0,0000	0,1000	0,2000	0,00709
Weibull(2;5)	10	0,1033	0,0961	0,9300	0,0000	0,1000	0,2000	0,00753
Weibull(7;15)	10	0,1007	0,0963	0,9566	0,0000	0,1000	0,2000	0,01483
Weibull(30;60)	10	0,0990	0,0925	0,9339	0,0000	0,1000	0,2000	0,01496
Weibull(1;1)	100	0,0996	0,0306	0,3077	0,0800	0,1000	0,1200	0,00703
Weibull(2;5)	100	0,1002	0,0300	0,2995	0,0800	0,1000	0,1200	0,00738
Weibull(7;15)	100	0,0996	0,0313	0,3143	0,0800	0,1000	0,1200	0,01450
Weibull(30;60)	100	0,1009	0,0303	0,3003	0,0800	0,1000	0,1200	0,01544
Weibull(1;1)	1000	0,1007	0,0117	0,1156	0,0920	0,1000	0,1090	0,00732
Weibull(2;5)	1000	0,1005	0,0117	0,1168	0,0930	0,1000	0,1080	0,00792
Weibull(7;15)	1000	0,1008	0,0123	0,1219	0,0930	0,1000	0,1090	0,01547
Weibull(30;60)	1000	0,1004	0,0121	0,1204	0,0920	0,1000	0,1080	0,01597
Lognormal(0,1;1)	10	0,1000	0,0944	0,9439	0,0000	0,1000	0,2000	0,01115
Lognormal(0,5;1,5)	10	0,1033	0,0992	0,9607	0,0000	0,1000	0,2000	0,01214
Lognormal(1;2)	10	0,1004	0,0973	0,9692	0,0000	0,1000	0,2000	0,01110
Lognormal(1,5;2,5)	10	0,0969	0,0957	0,9871	0,0000	0,1000	0,2000	0,01108
Lognormal(0,1;1)	100	0,1007	0,0305	0,3031	0,0800	0,1000	0,1200	0,01151
Lognormal(0,5;1,5)	100	0,1017	0,0299	0,2944	0,0800	0,1000	0,1200	0,01205
Lognormal(1;2)	100	0,1013	0,0293	0,2899	0,0800	0,1000	0,1200	0,01095
Lognormal(1,5;2,5)	100	0,1012	0,0321	0,3176	0,0800	0,1000	0,1200	0,01176
Lognormal(0,1;1)	1000	0,1002	0,0121	0,1206	0,0920	0,1000	0,1090	0,01105
Lognormal(0,5;1,5)	1000	0,1009	0,0117	0,1156	0,0930	0,1000	0,1080	0,01156
Lognormal(1;2)	1000	0,1009	0,0120	0,1191	0,0930	0,1010	0,1090	0,01097
Lognormal(1,5;2,5)	1000	0,1011	0,0121	0,1200	0,0930	0,1010	0,1090	0,01119
BS(0,5;1,5)	10	0,1005	0,0945	0,9407	0,0000	0,1000	0,2000	0,00825
BS(1,5;3,5)	10	0,0983	0,0953	0,9692	0,0000	0,1000	0,1000	0,00693
BS(1,5;5,5)	10	0,0980	0,0911	0,9299	0,0000	0,1000	0,2000	0,00708
BS(1,5;7,5)	10	0,0964	0,0930	0,9651	0,0000	0,1000	0,2000	0,00677
BS(0,5;1,5)	100	0,0997	0,0298	0,2990	0,0800	0,1000	0,1200	0,00778
BS(1,5;3,5)	100	0,0989	0,0301	0,3047	0,0800	0,1000	0,1200	0,00692
BS(1,5;5,5)	100	0,1006	0,0303	0,3014	0,0800	0,1000	0,1200	0,00687
BS(1,5;7,5)	100	0,1014	0,0315	0,3106	0,0800	0,1000	0,1200	0,00675
BS(0,5;1,5)	1000	0,1011	0,0124	0,1229	0,0920	0,1010	0,1100	0,00800
BS(1,5;3,5)	1000	0,1011	0,0115	0,1133	0,0920	0,1010	0,1090	0,00707
BS(1,5;5,5)	1000	0,1003	0,0117	0,1163	0,0920	0,0990	0,1090	0,00702
BS(1,5;7,5)	1000	0,1012	0,0116	0,1143	0,0940	0,1010	0,1090	0,00679

## **Apêndice F**

### **Algoritmo 5 - Etas escolhidos**

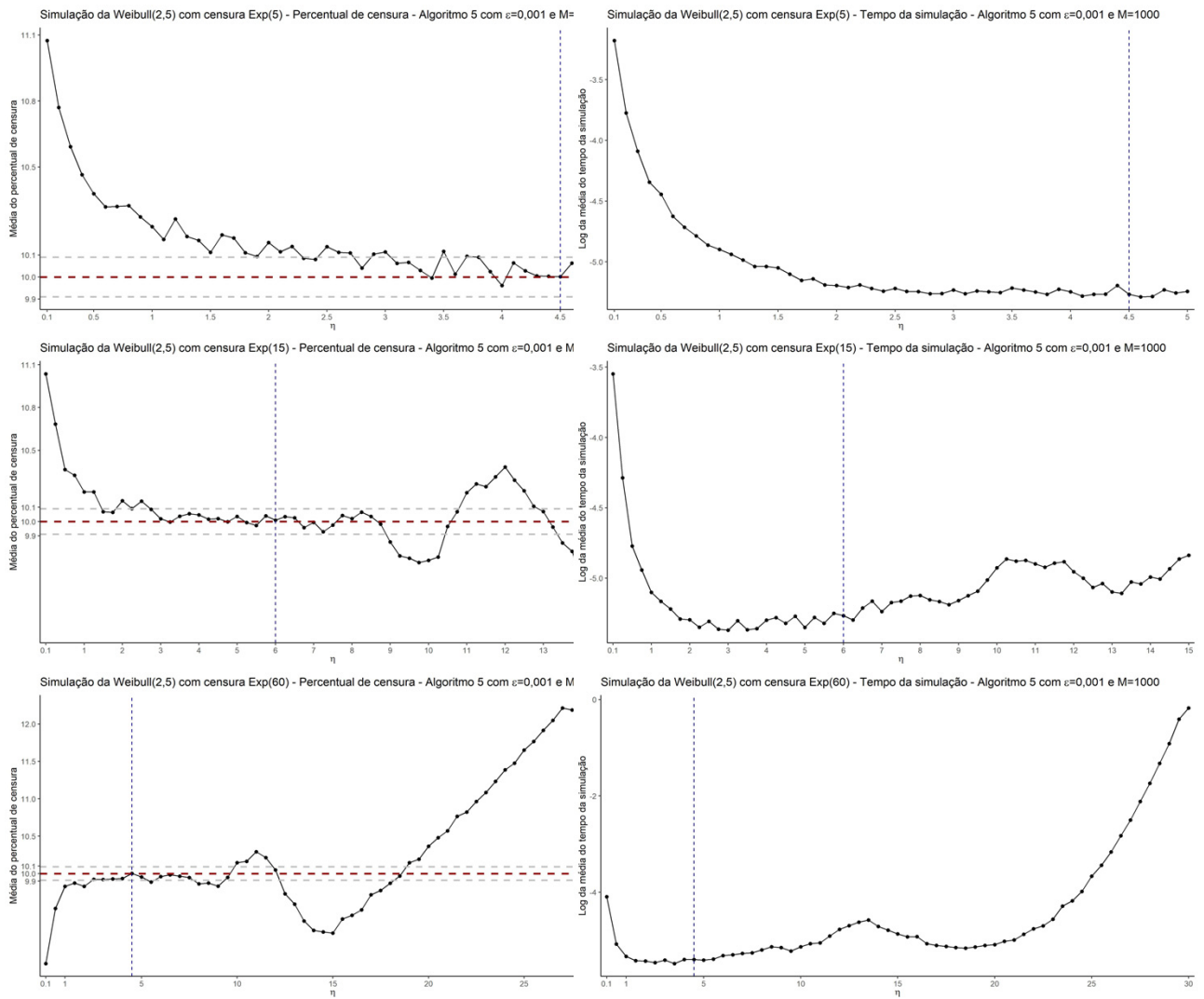


Figura F.1: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 da distribuição de tempo  $T \sim Weibull(2; 5)$  para distribuição de censura,  $C$ , Exponencial, com diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\epsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

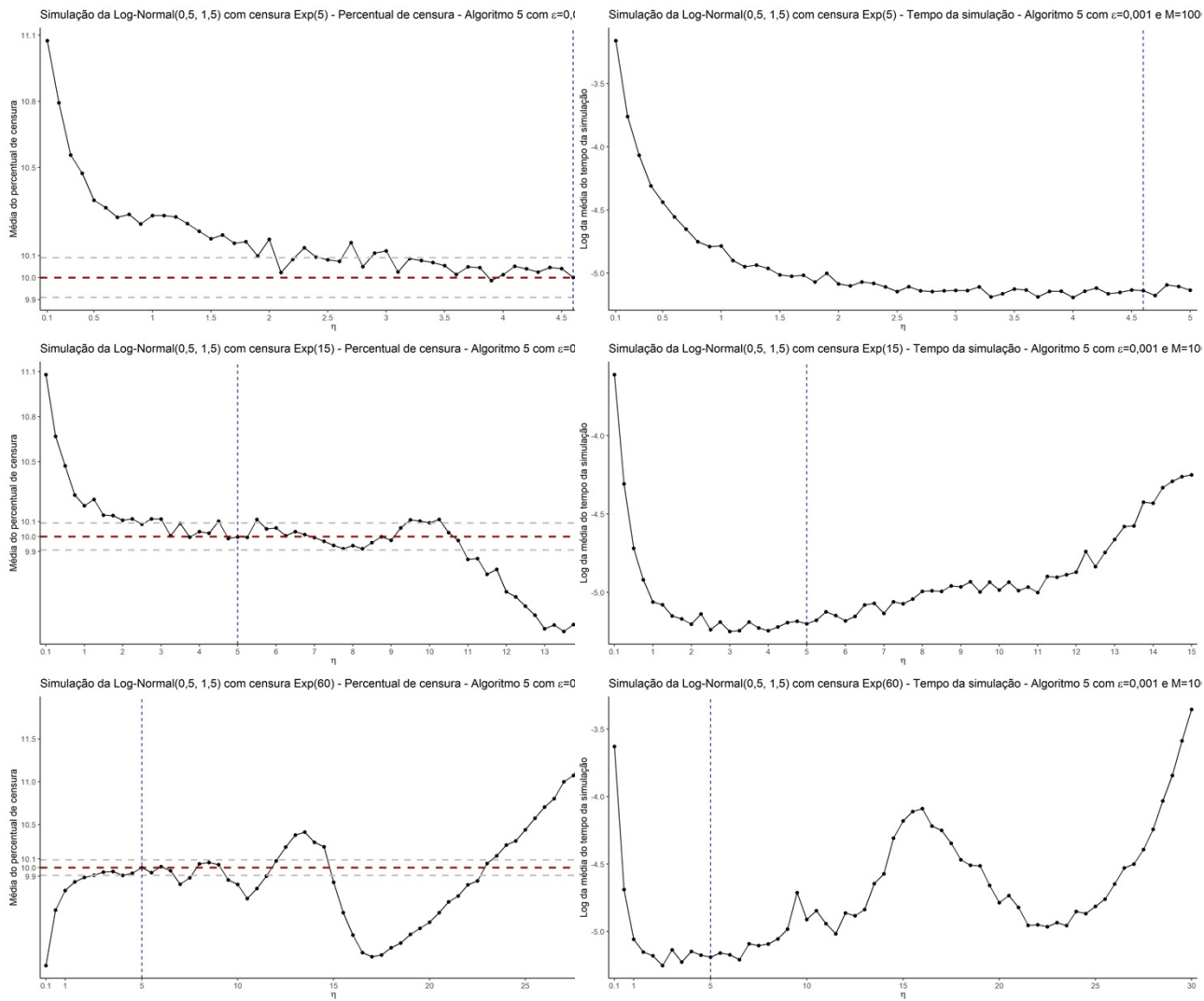


Figura F.2: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 da distribuição de tempo  $T \sim Lognormal(0, 5; 1, 5)$  para distribuição de censura,  $C$ , Exponencial, com diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

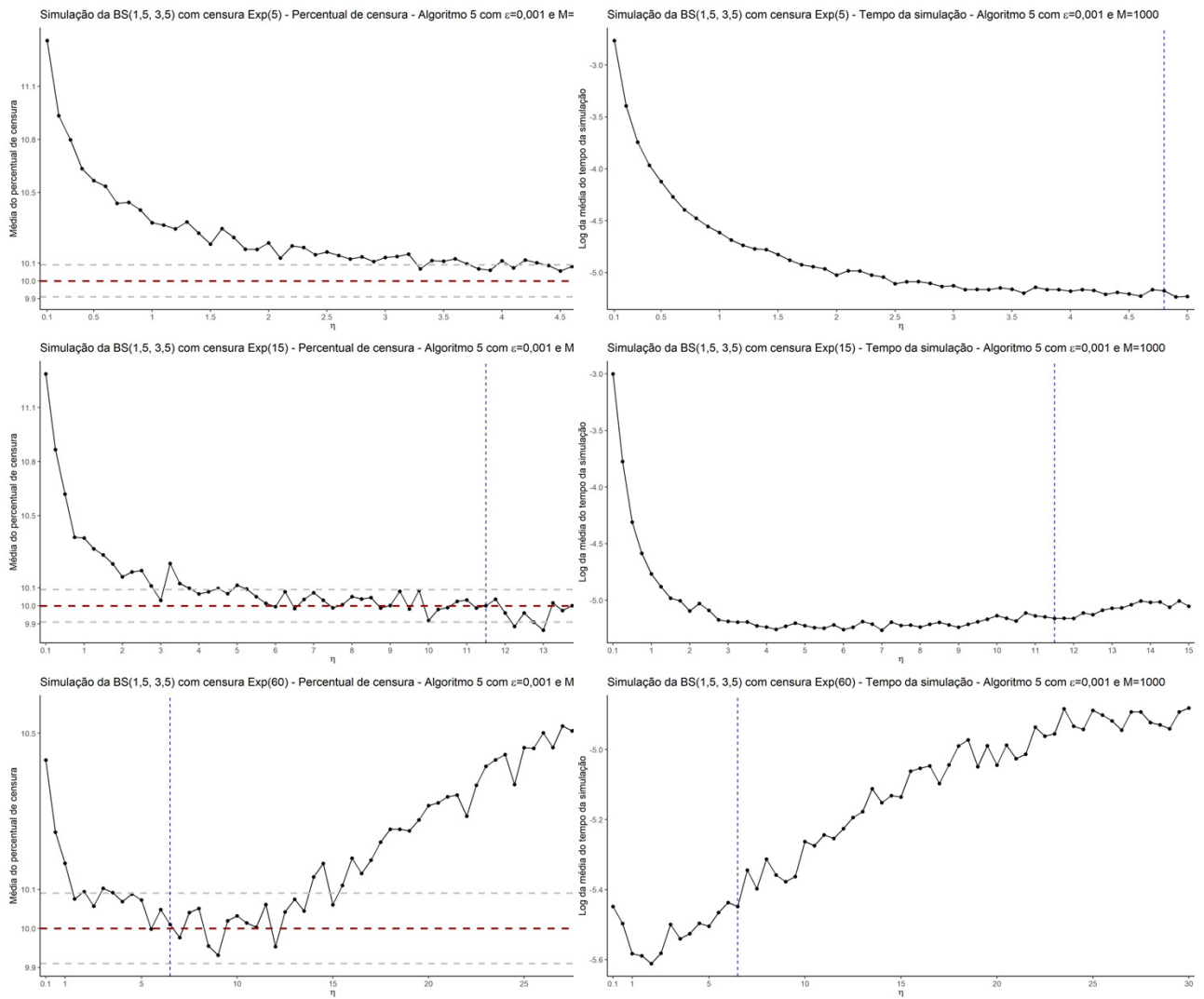


Figura F.3: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 da distribuição de tempo  $T \sim BS(1, 5; 3, 5)$  para distribuição de censura,  $C$ , Exponencial, com diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$



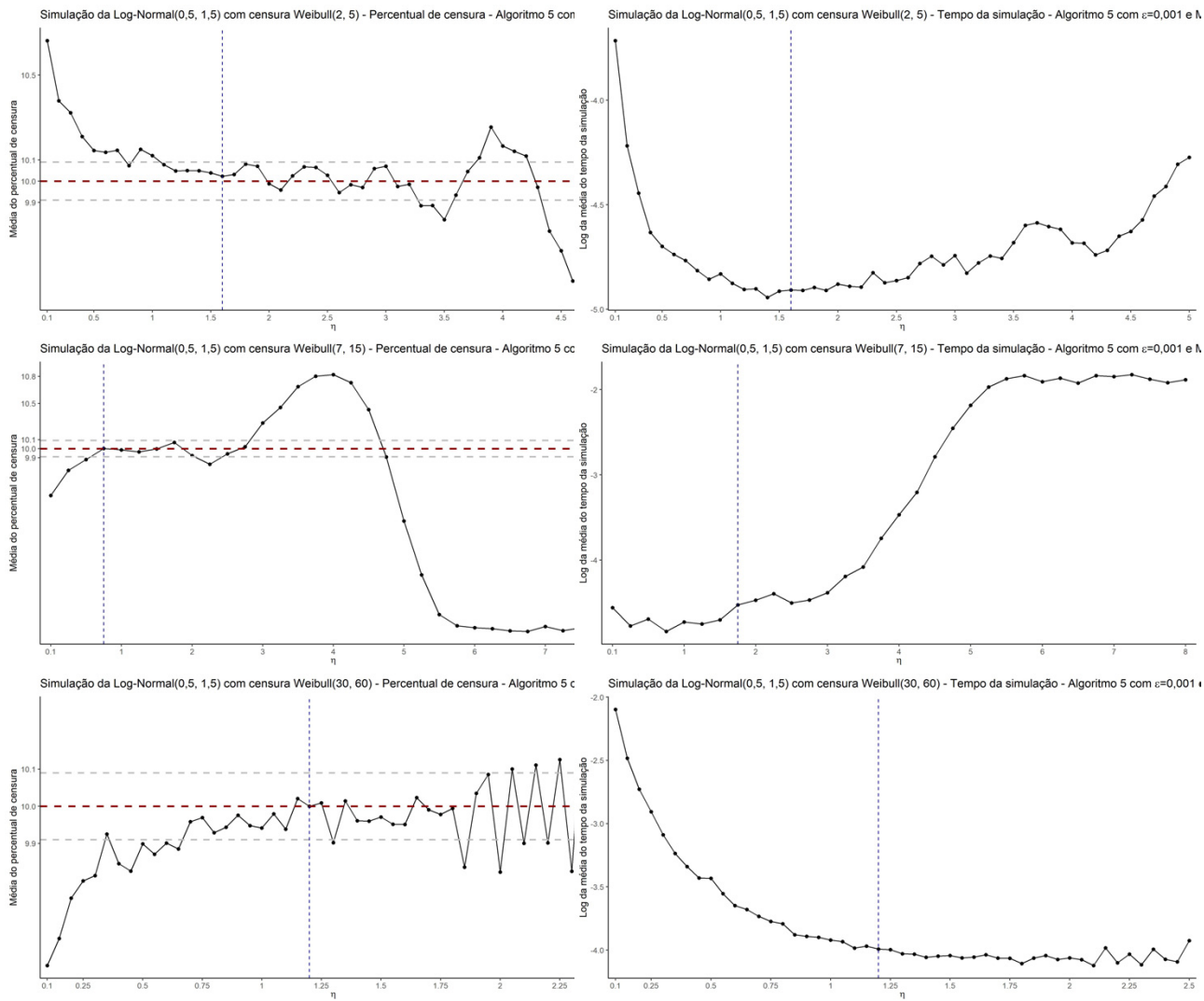


Figura F.4: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 da distribuição de tempo  $T \sim Lognormal(0, 5; 1, 5)$  para distribuição de censura,  $C$ , Weibull, com diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

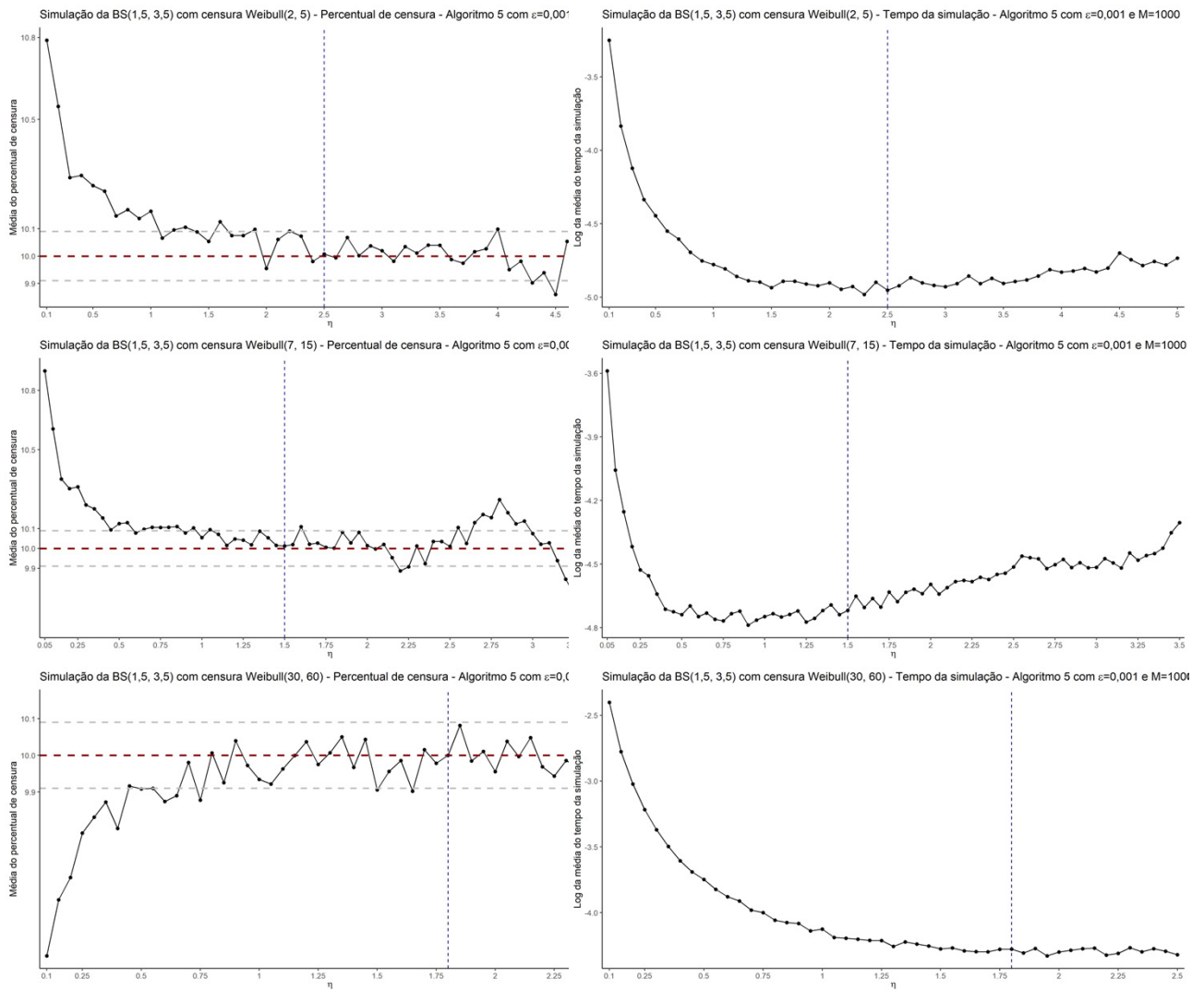


Figura F.5: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 da distribuição de tempo  $T \sim BS(1, 5; 3, 5)$  para distribuição de censura,  $C$ , Weibull, com diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

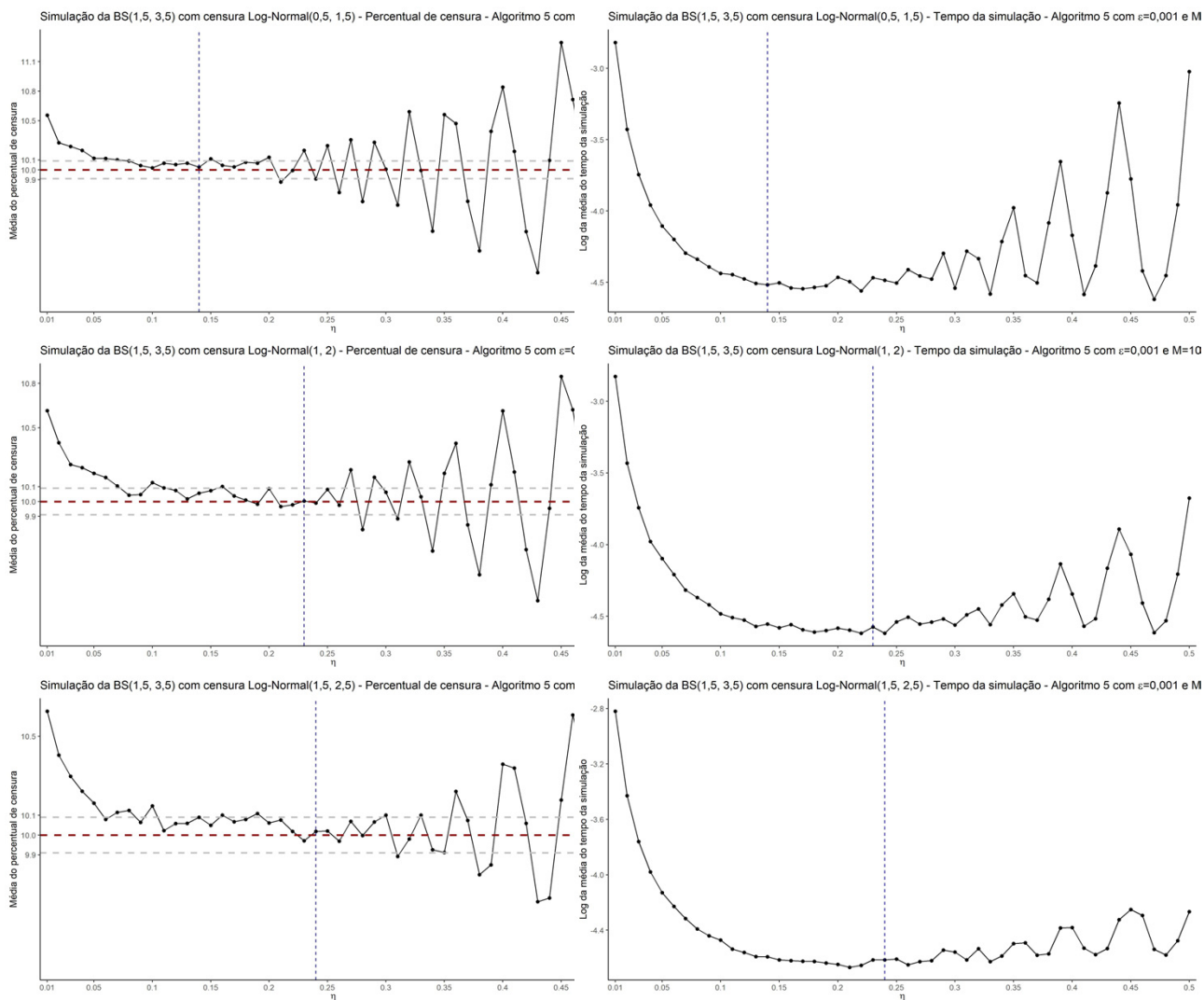


Figura F.6: Média do percentual de censura e Log da média do tempo da simulação. De 1000 simulações de Censura Aleatória Caso 3, distribuição de  $C$  diferente de  $T$  - Algoritmo 5 da distribuição de tempo  $T \sim BS(1, 5; 3, 5)$  para distribuição de censura,  $C$ , Log-Normal, com diferentes parâmetros e  $\eta$  com  $M = 1000$ ,  $\varepsilon = 0,001$ ,  $p = 0,1$  e  $n = 1000$

# **Apêndice G**

## **Simulações do Algoritmo 6**

Tabela G.1: Resultado da média do percentual de censura ( $E(\hat{p})$ ), desvio-padrão ( $DP(\hat{p})$ ), coeficiente de variação ( $CV(\hat{p})$ ) e média de tempo em segundos por simulação em 1000 simulações de Censura Aleatória Caso 4, indicador de Censura via Bernoulli - Algoritmo 6 para as distribuições de tempo  $T$ : Exponencial, Weibull, Log-Normal e Birnbaum-Saunders, com diferentes parâmetros e amostras de tamanho 10, 100 e 1000

Distribuição	$n$	$E(\hat{p})$	$DP(\hat{p})$	$CV(\hat{p})$	Quantis			Média de tempo em segundos por simulação
					1°	Mediana	3°	
Exp(1)	10	0,0976	0,0937	0,9602	0,0000	0,1000	0,2000	0,00004
Exp(5)	10	0,1050	0,0973	0,9265	0,0000	0,1000	0,2000	0,00001
Exp(15)	10	0,1020	0,0977	0,9579	0,0000	0,1000	0,2000	0,00002
Exp(60)	10	0,0982	0,0947	0,9642	0,0000	0,1000	0,2000	0,00002
Exp(1)	100	0,1013	0,0293	0,2889	0,0800	0,1000	0,1200	0,00002
Exp(5)	100	0,1004	0,0295	0,2936	0,0800	0,1000	0,1200	0,00002
Exp(15)	100	0,1007	0,0297	0,2955	0,0800	0,1000	0,1200	0,00003
Exp(60)	100	0,0990	0,0305	0,3080	0,0800	0,1000	0,1200	0,00002
Exp(1)	1000	0,1001	0,0097	0,0971	0,0930	0,1000	0,1070	0,00009
Exp(5)	1000	0,0995	0,0093	0,0937	0,0930	0,0990	0,1060	0,00009
Exp(15)	1000	0,1003	0,0092	0,0916	0,0940	0,1000	0,1060	0,00010
Exp(60)	1000	0,1007	0,0093	0,0919	0,0940	0,1000	0,1070	0,00009
Weibull(1;1)	10	0,0986	0,0945	0,9582	0,0000	0,1000	0,1000	0,00003
Weibull(2;5)	10	0,1001	0,0919	0,9177	0,0000	0,1000	0,2000	0,00002
Weibull(7;15)	10	0,0956	0,0927	0,9694	0,0000	0,1000	0,2000	0,00002
Weibull(30;60)	10	0,1005	0,0952	0,9470	0,0000	0,1000	0,2000	0,00003
Weibull(1;1)	100	0,1008	0,0296	0,2936	0,0800	0,1000	0,1200	0,00003
Weibull(2;5)	100	0,1001	0,0304	0,3040	0,0800	0,1000	0,1200	0,00003
Weibull(7;15)	100	0,0962	0,0305	0,3172	0,0700	0,0900	0,1200	0,00003
Weibull(30;60)	100	0,1016	0,0302	0,2975	0,0800	0,1000	0,1200	0,00003
Weibull(1;1)	1000	0,1003	0,0093	0,0929	0,0940	0,1000	0,1060	0,00013
Weibull(2;5)	1000	0,0997	0,0093	0,0929	0,0930	0,1000	0,1060	0,00013
Weibull(7;15)	1000	0,0994	0,0094	0,0946	0,0930	0,0990	0,1060	0,00021
Weibull(30;60)	1000	0,0996	0,0096	0,0963	0,0930	0,0990	0,1060	0,00020
Lognormal(0,1;1)	10	0,1005	0,0958	0,9533	0,0000	0,1000	0,2000	0,00003
Lognormal(0,5;1,5)	10	0,0936	0,0890	0,9512	0,0000	0,1000	0,1000	0,00002
Lognormal(1;2)	10	0,0993	0,0928	0,9349	0,0000	0,1000	0,2000	0,00002
Lognormal(1,5;2,5)	10	0,0970	0,0960	0,9899	0,0000	0,1000	0,2000	0,00002
Lognormal(0,1;1)	100	0,1009	0,0310	0,3071	0,0800	0,1000	0,1200	0,00002
Lognormal(0,5;1,5)	100	0,0984	0,0305	0,3095	0,0800	0,1000	0,1200	0,00003
Lognormal(1;2)	100	0,1001	0,0297	0,2968	0,0800	0,1000	0,1200	0,00003
Lognormal(1,5;2,5)	100	0,1013	0,0308	0,3045	0,0800	0,1000	0,1200	0,00002
Lognormal(0,1;1)	1000	0,1000	0,0093	0,0935	0,0940	0,1000	0,1060	0,00016
Lognormal(0,5;1,5)	1000	0,0996	0,0095	0,0951	0,0928	0,1000	0,1060	0,00016
Lognormal(1;2)	1000	0,1002	0,0101	0,1003	0,0930	0,1000	0,1070	0,00016
Lognormal(1,5;2,5)	1000	0,1003	0,0094	0,0935	0,0940	0,1000	0,1060	0,00015
BS(0,5;1,5)	10	0,1012	0,0941	0,9295	0,0000	0,1000	0,2000	0,00024
BS(1,5;3,5)	10	0,0995	0,0951	0,9555	0,0000	0,1000	0,2000	0,00019
BS(1,5;5,5)	10	0,1007	0,0954	0,9472	0,0000	0,1000	0,2000	0,00019
BS(1,5;7,5)	10	0,1077	0,1005	0,9332	0,0000	0,1000	0,2000	0,00020
BS(0,5;1,5)	100	0,1007	0,0308	0,3054	0,0800	0,1000	0,1200	0,00020
BS(1,5;3,5)	100	0,1001	0,0298	0,2977	0,0800	0,1000	0,1200	0,00021
BS(1,5;5,5)	100	0,0995	0,0302	0,3032	0,0800	0,1000	0,1200	0,00020
BS(1,5;7,5)	100	0,0995	0,0299	0,3005	0,0800	0,1000	0,1200	0,00019
BS(0,5;1,5)	1000	0,1002	0,0092	0,0920	0,0940	0,1000	0,1070	0,00031
BS(1,5;3,5)	1000	0,1000	0,0093	0,0929	0,0940	0,1000	0,1060	0,00031
BS(1,5;5,5)	1000	0,1006	0,0095	0,0945	0,0940	0,1010	0,1070	0,00032
BS(1,5;7,5)	1000	0,0999	0,0094	0,0942	0,0940	0,1000	0,1060	0,00030

## **Apêndice H**

### **Distribuição da amostra dos dados censurados**

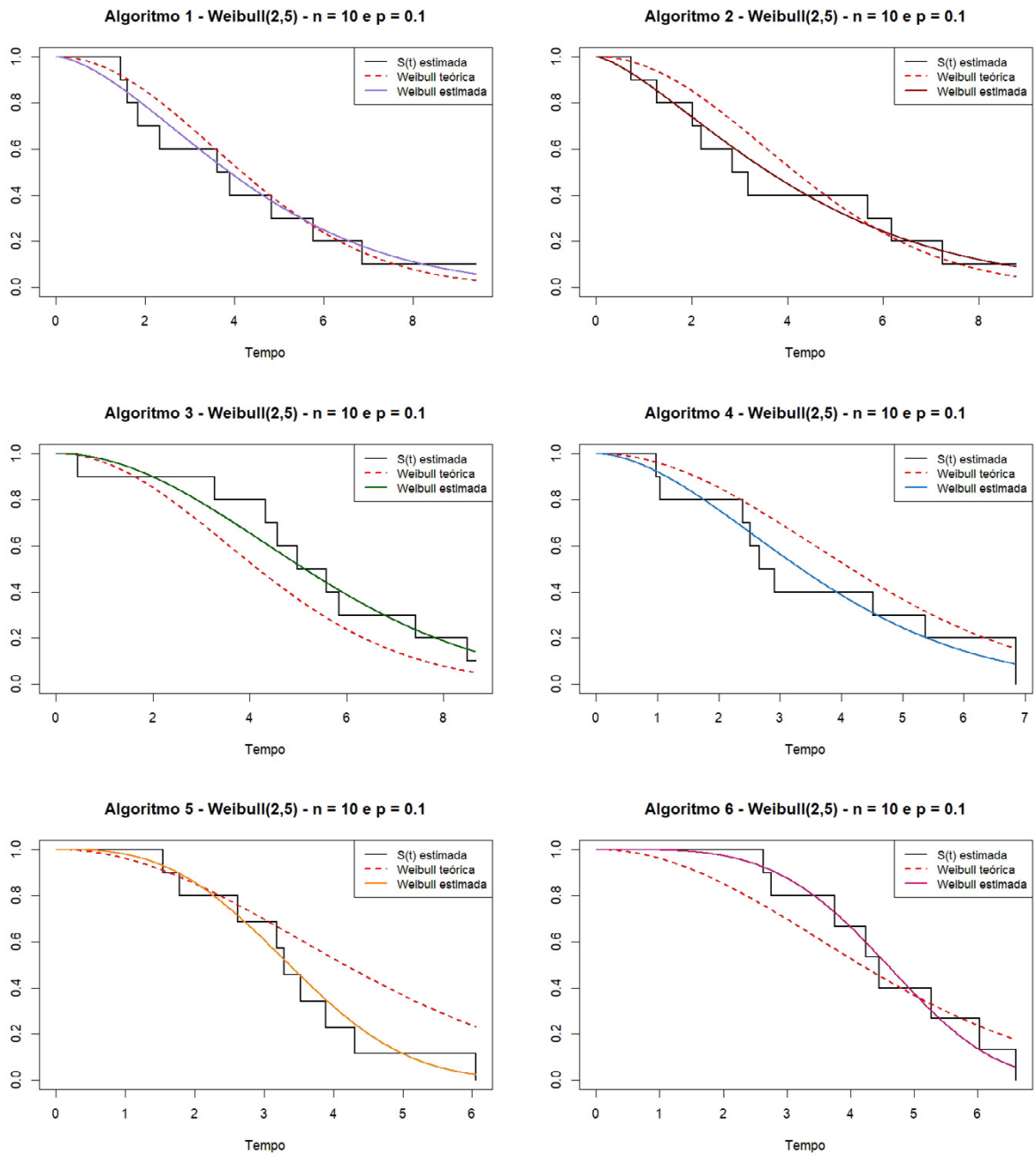


Figura H.1: Ajuste do modelo  $Weibull(2;5)$  a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 10 e percentual de censura de 10%

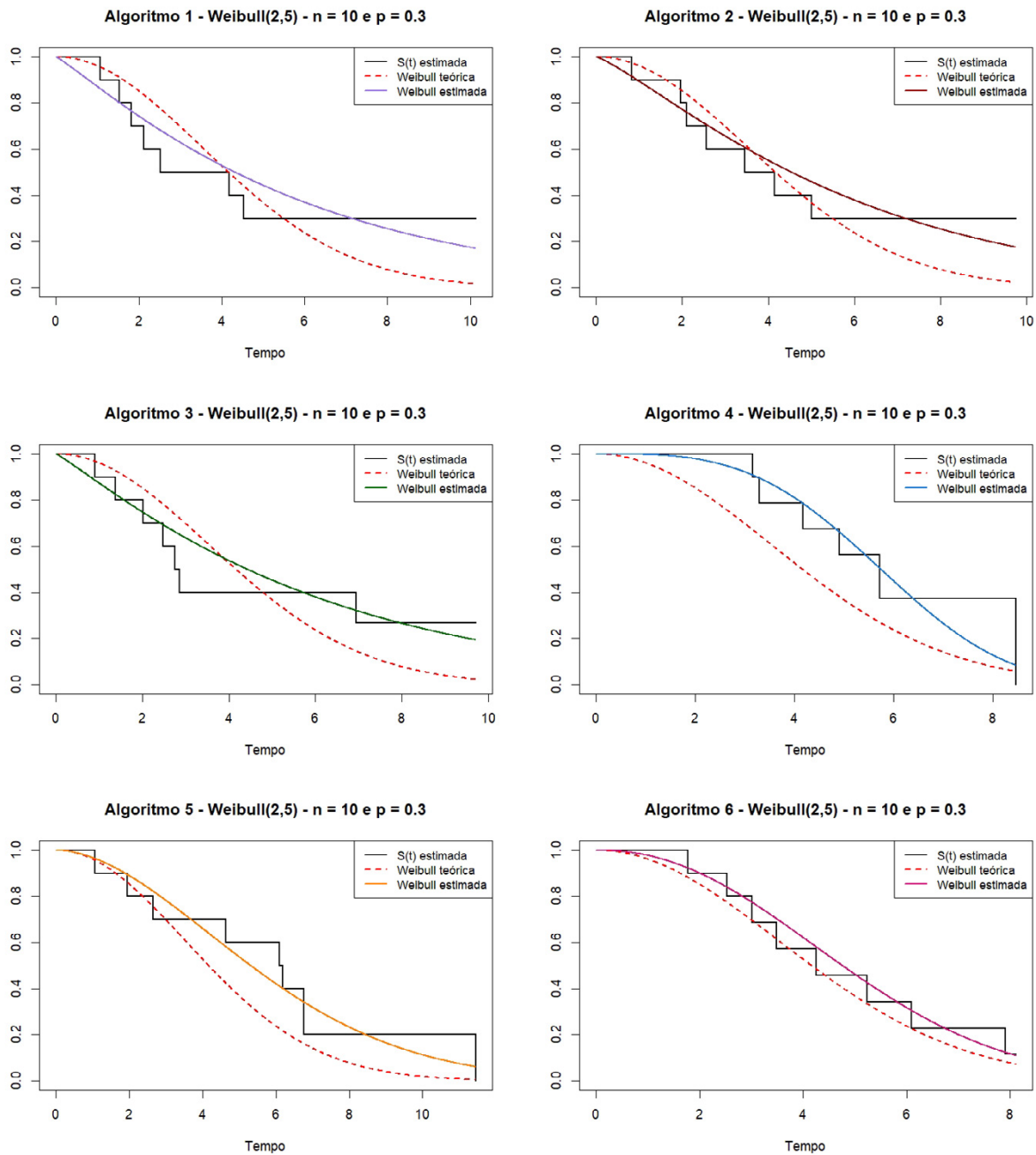


Figura H.2: Ajuste do modelo  $Weibull(2;5)$  a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 10 e percentual de censura de 30%



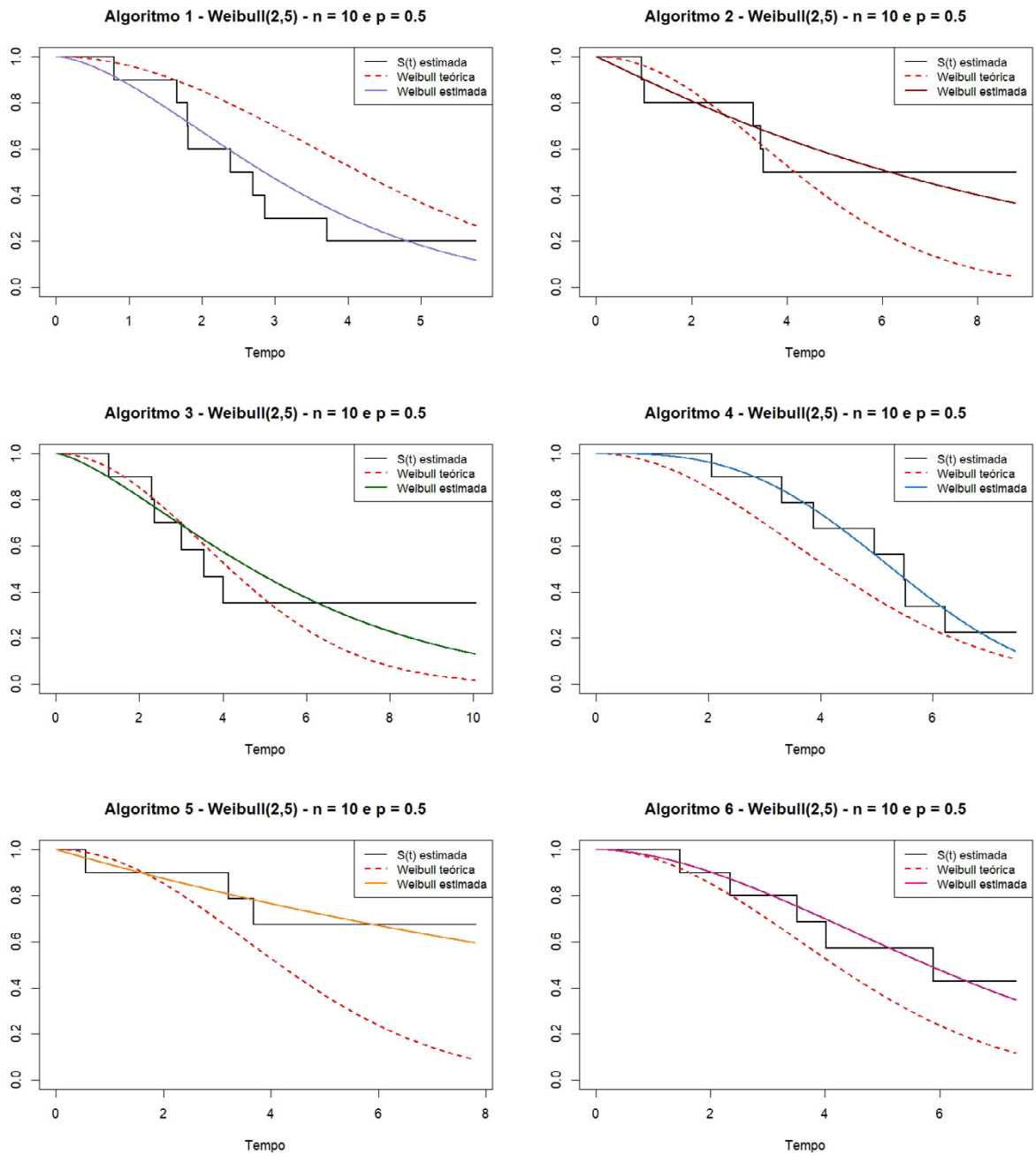


Figura H.3: Ajuste do modelo  $Weibull(2;5)$  a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 10 e percentual de censura de 50%

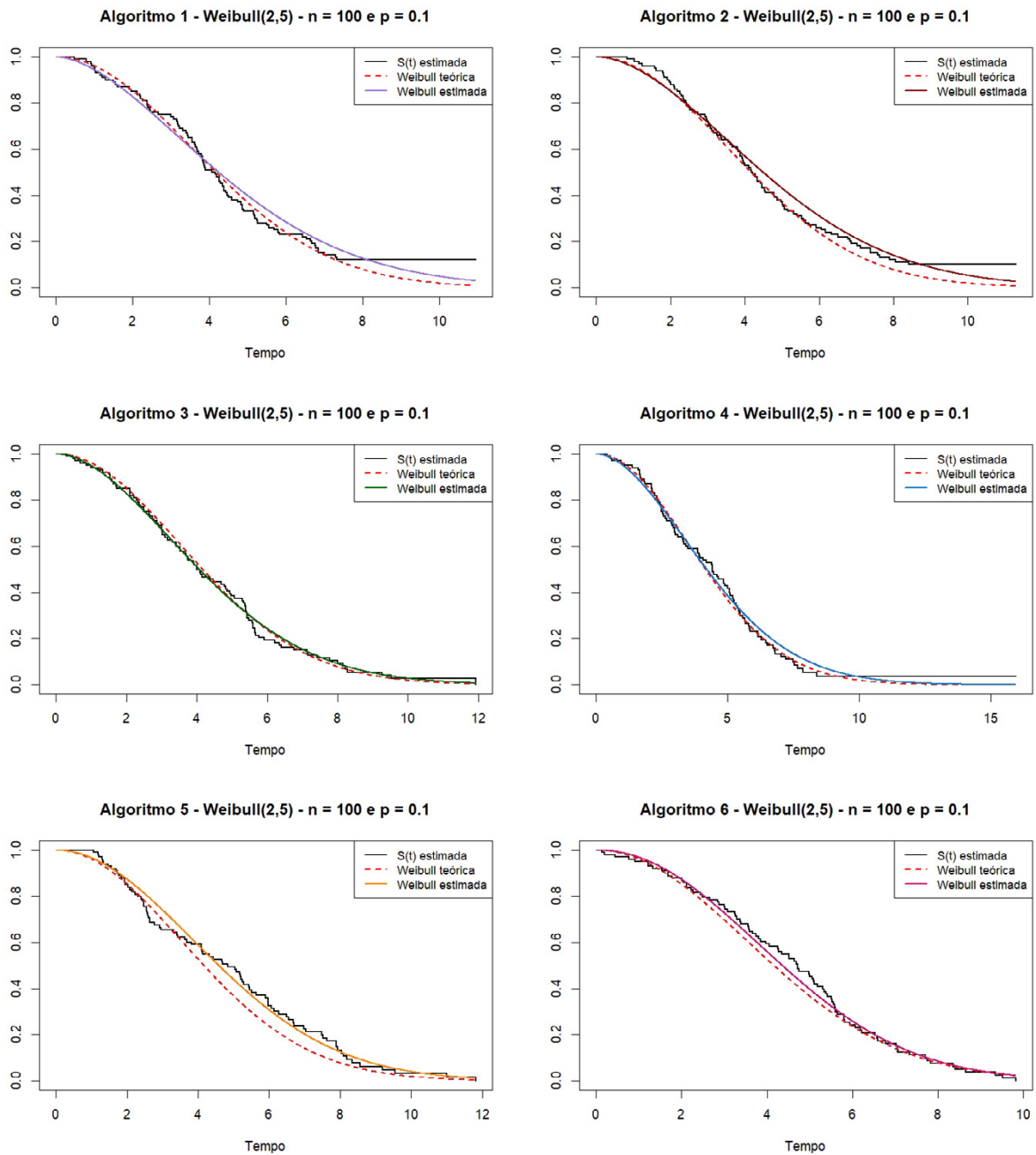


Figura H.4: Ajuste do modelo  $Weibull(2;5)$  a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 100 e percentual de censura de 10%

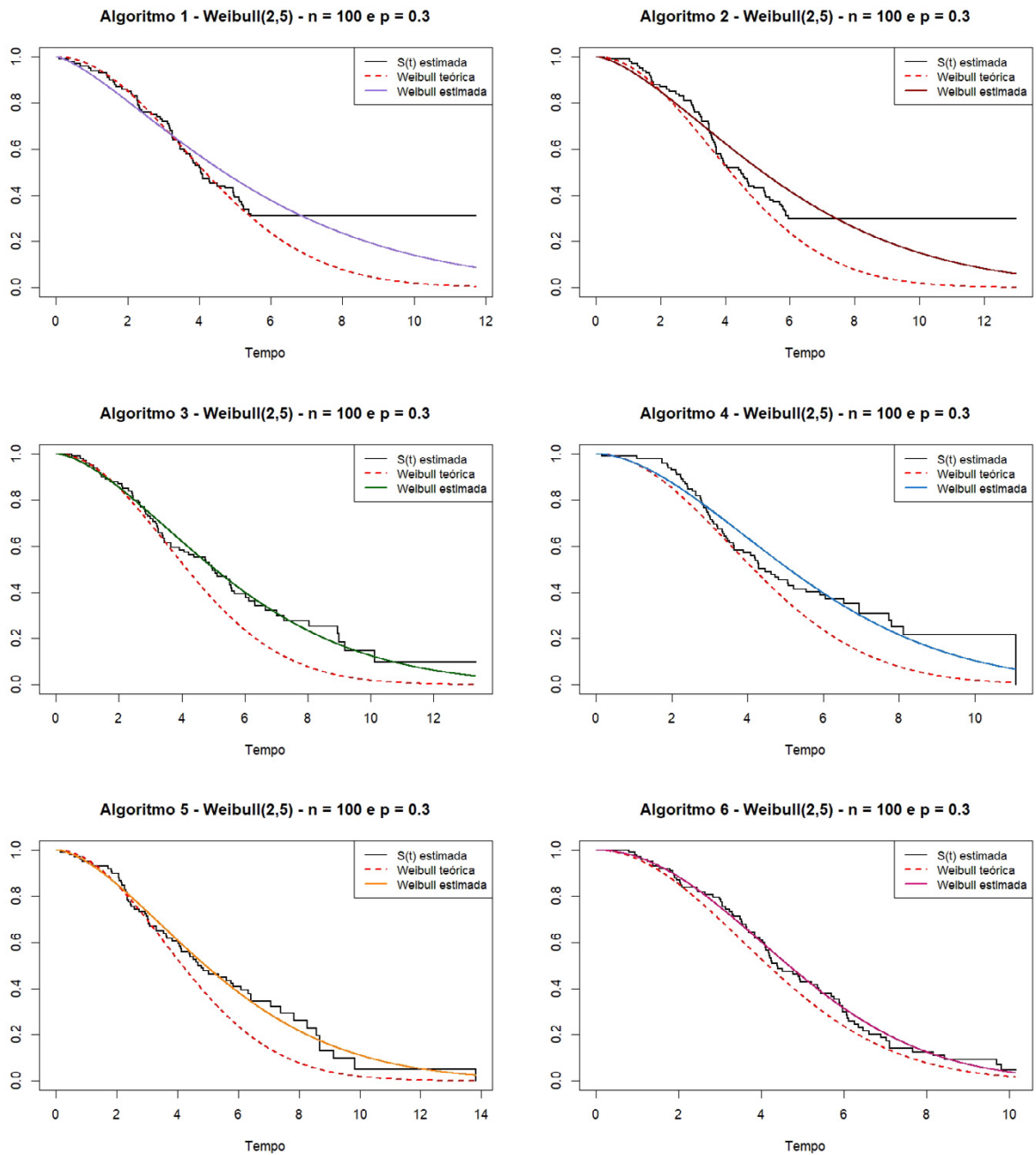


Figura H.5: Ajuste do modelo *Weibull*(2;5) a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 100 e percentual de censura de 30%

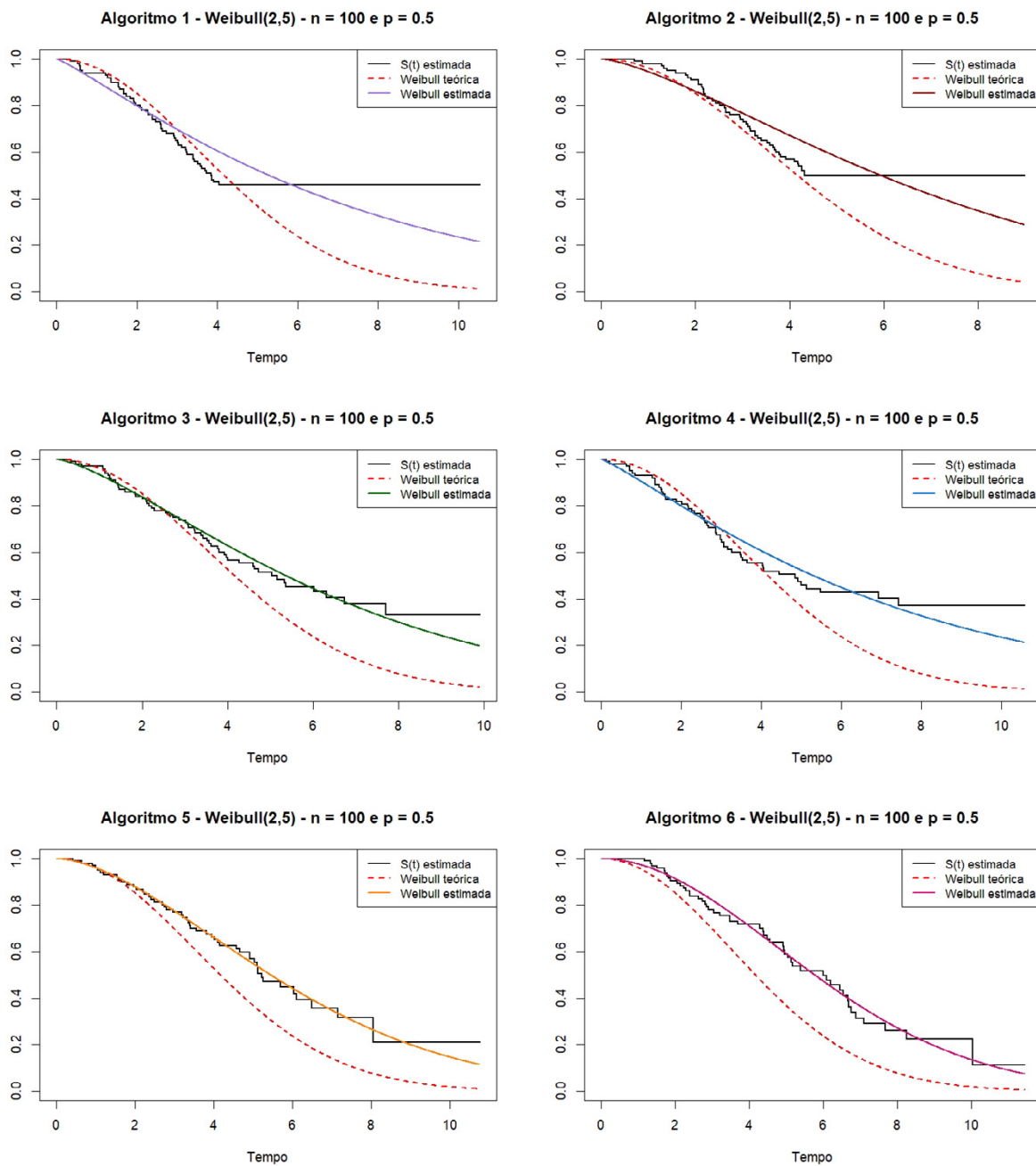


Figura H.6: Ajuste do modelo  $Weibull(2;5)$  a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 100 e percentual de censura de 50%

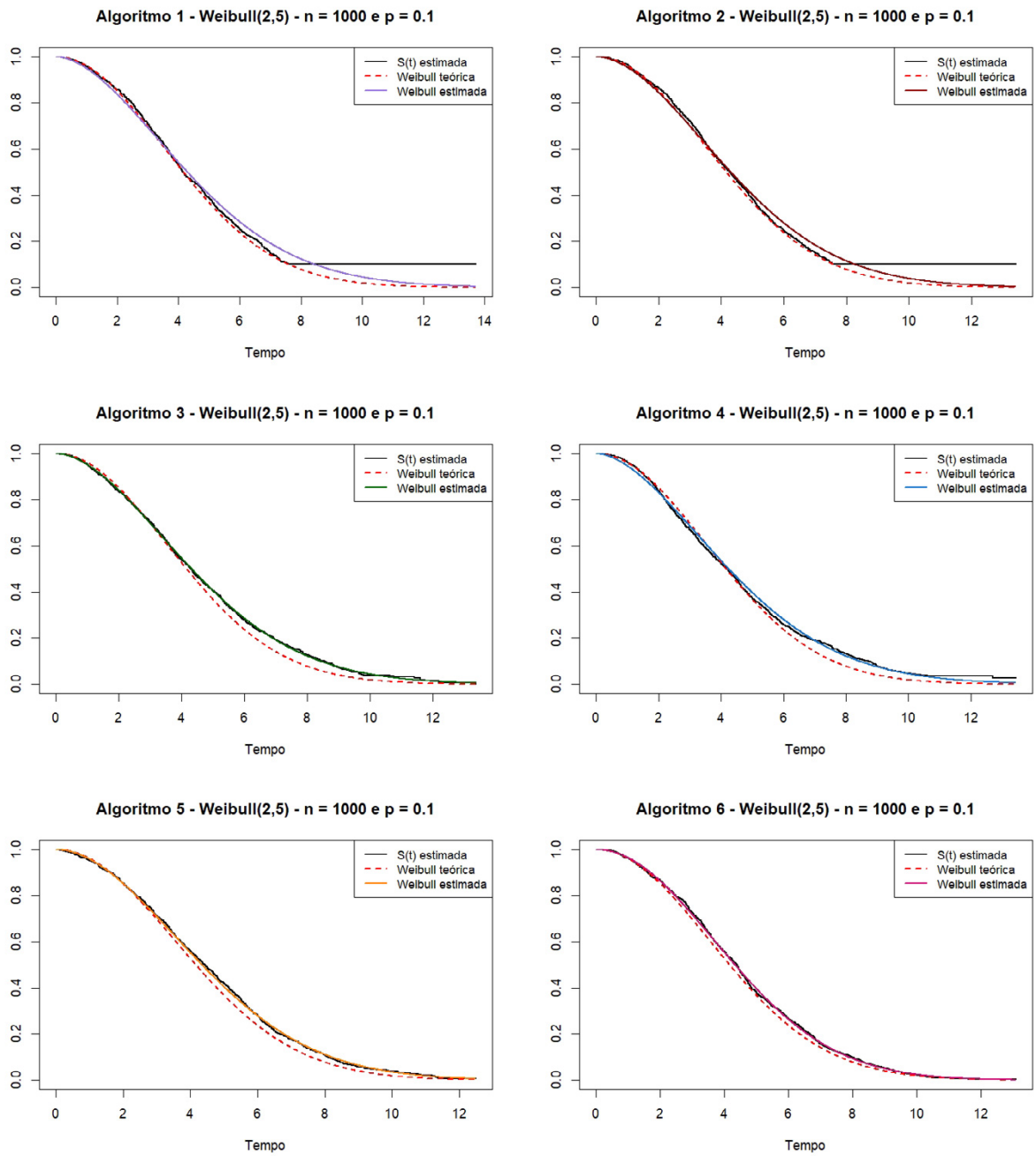


Figura H.7: Ajuste do modelo *Weibull*(2;5) a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 1000 e percentual de censura de 10%

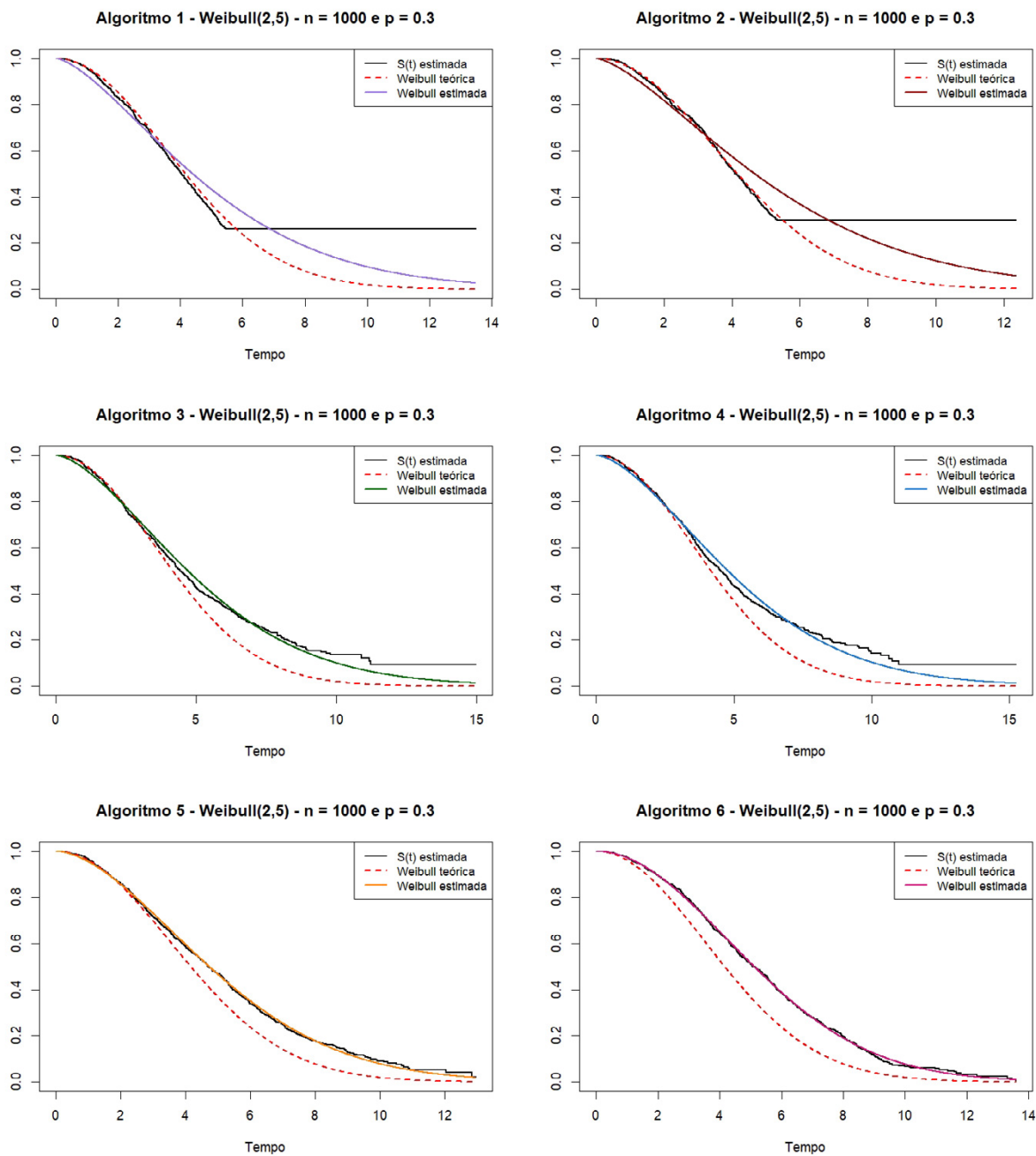


Figura H.8: Ajuste do modelo  $Weibull(2;5)$  a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 1000 e percentual de censura de 30%

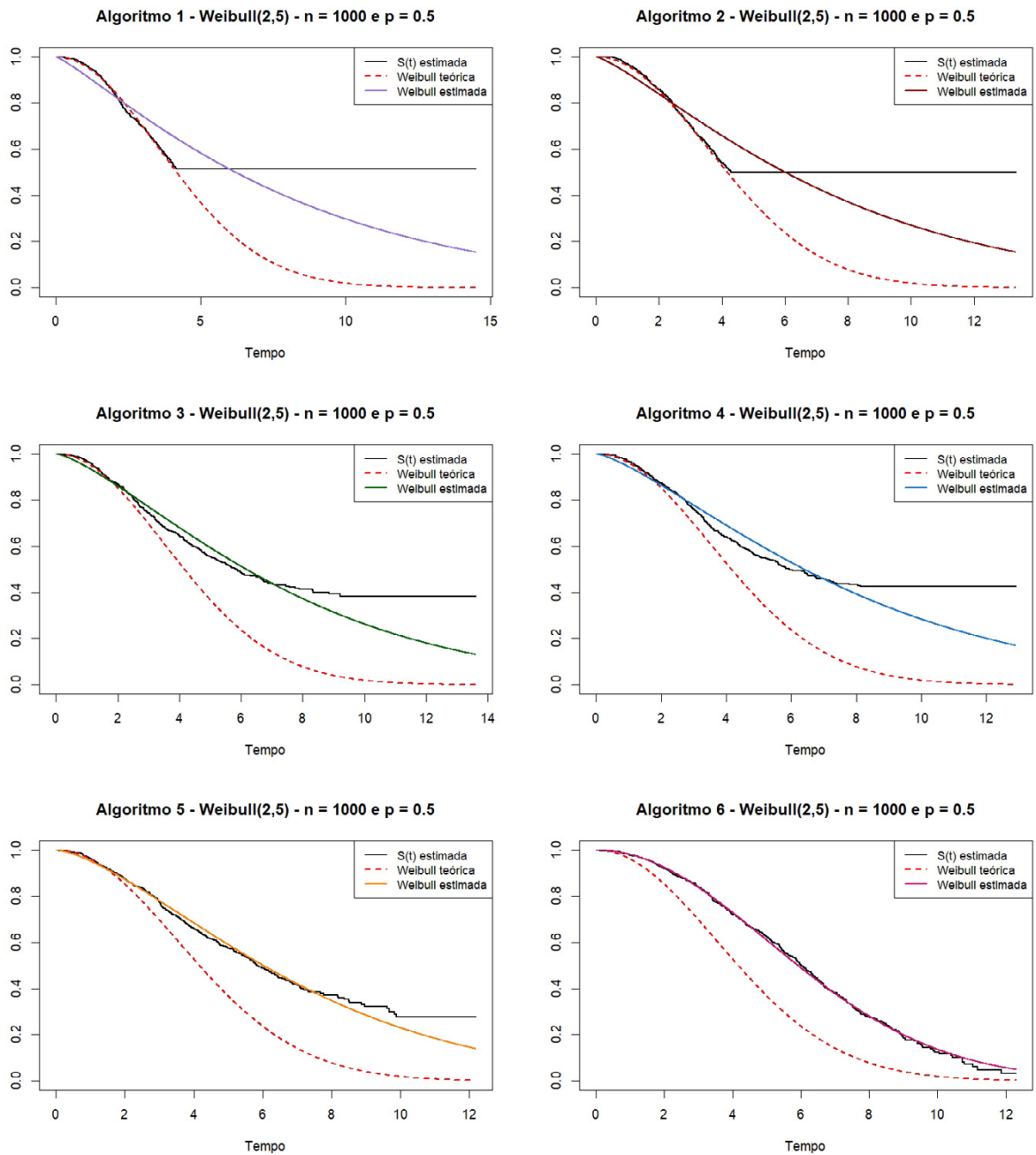


Figura H.9: Ajuste do modelo  $Weibull(2;5)$  a Função de Sobrevivência estimada para os Algoritmos 1 a 6, amostra de tamanho 1000 e percentual de censura de 50%

# **Apêndice I**

## **Distribuição da amostra dos dados discretos censurados**



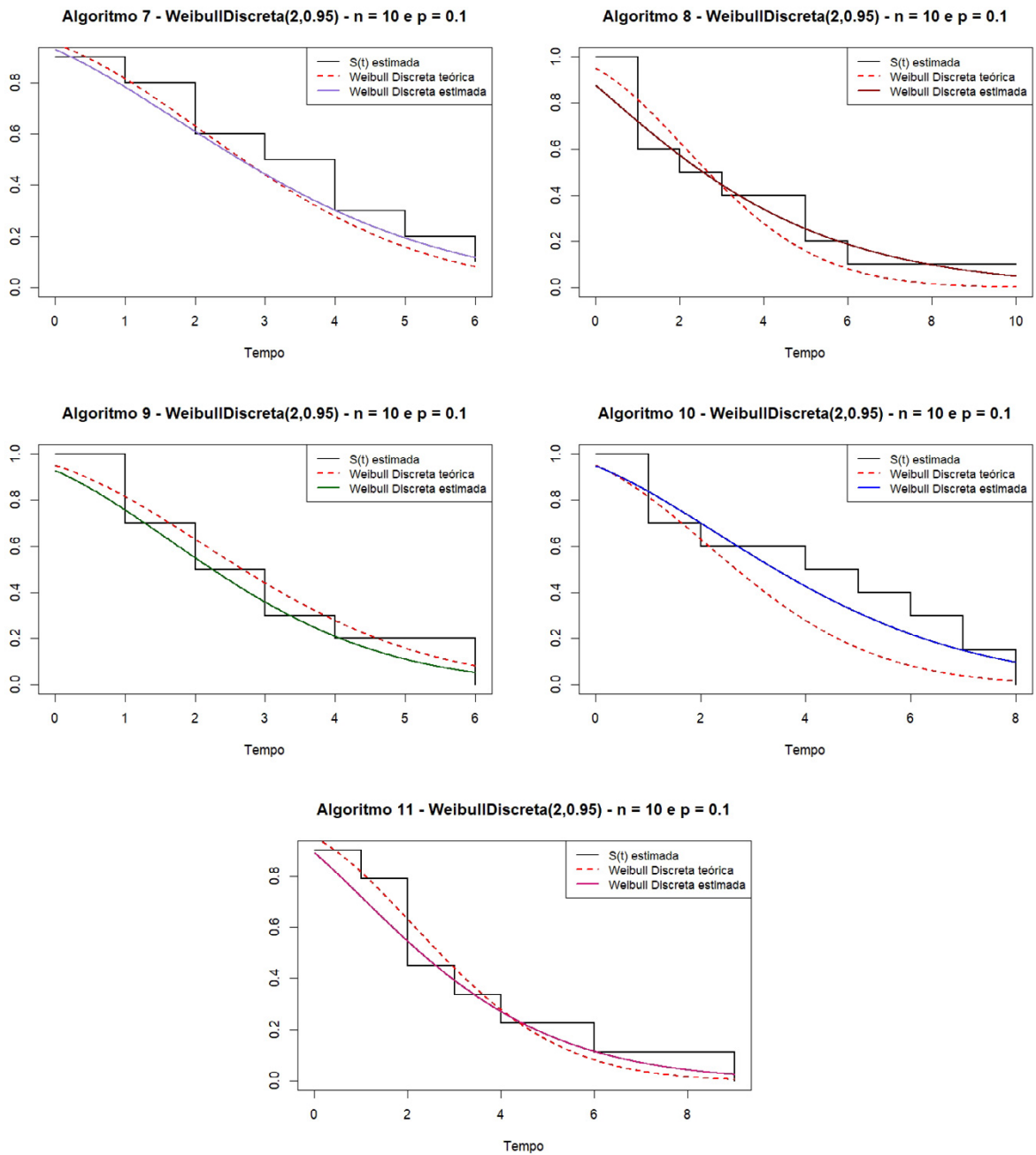


Figura I.1: Ajuste do modelo  $WeibullDiscreta(2;0,95)$  a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 10 e percentual de censura de 10%

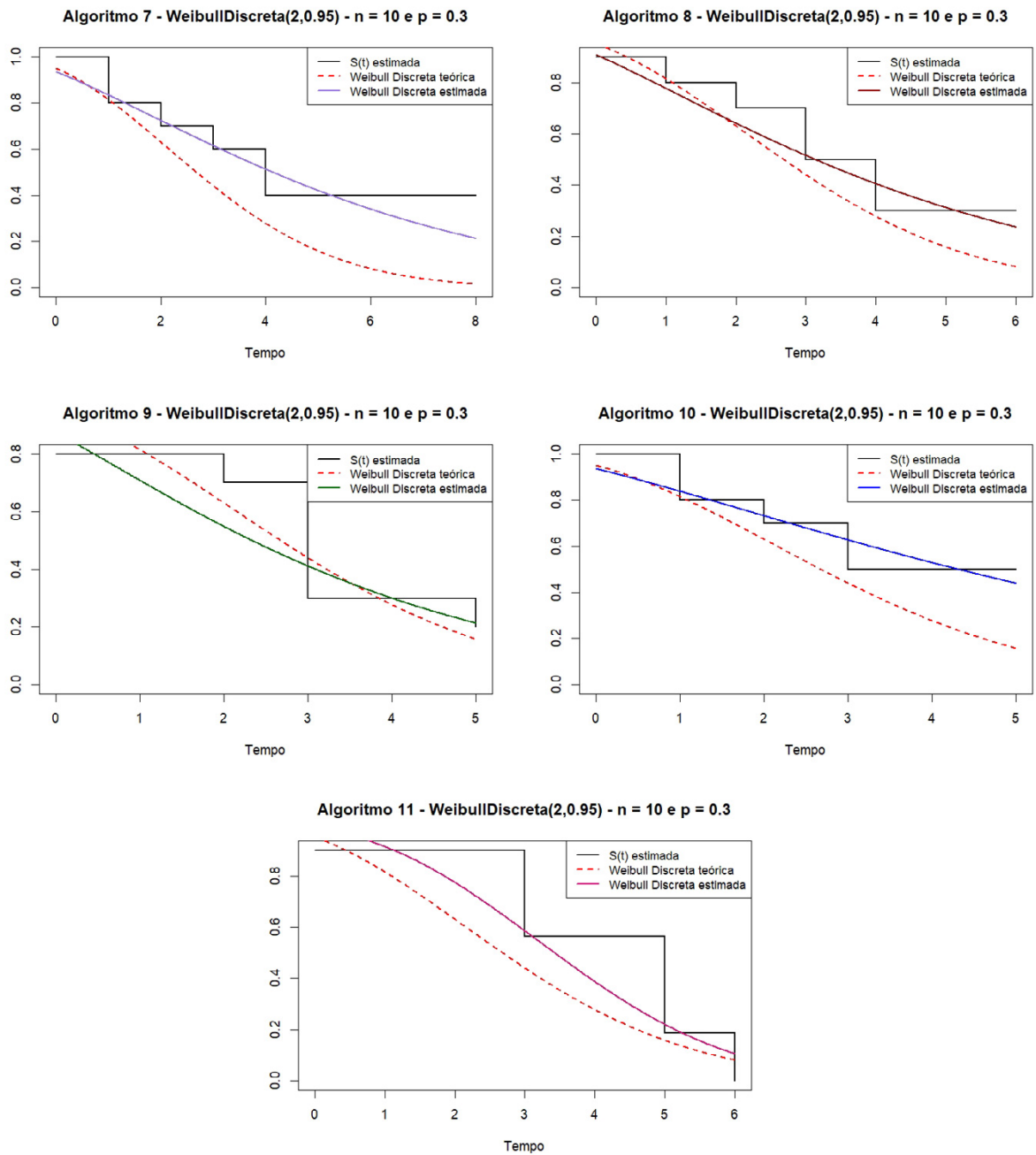


Figura I.2: Ajuste do modelo  $WeibullDiscreta(2;0,95)$  a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 10 e percentual de censura de 30%

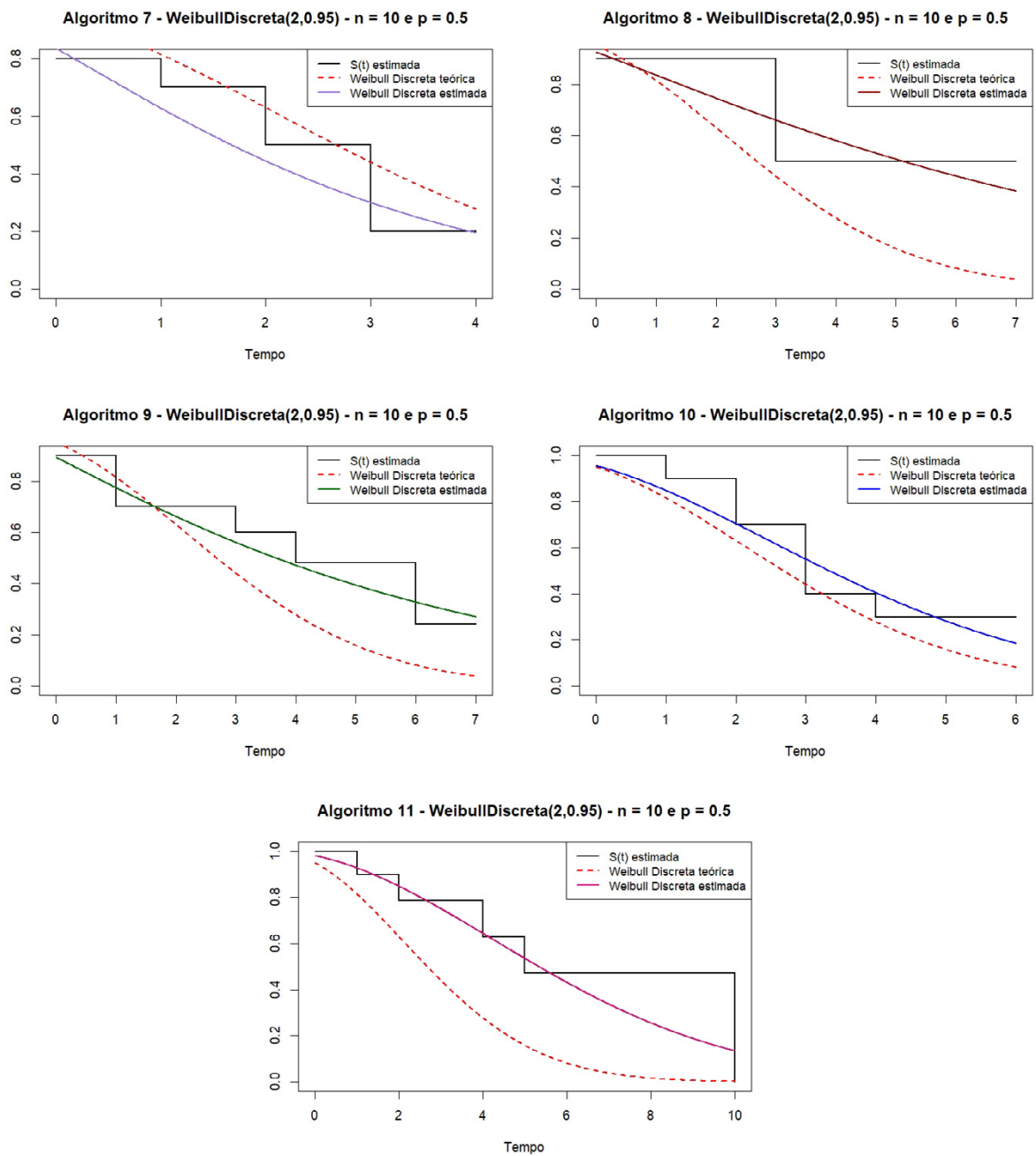


Figura I.3: Ajuste do modelo *WeibullDiscreta*(2;0,95) a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 10 e percentual de censura de 50%

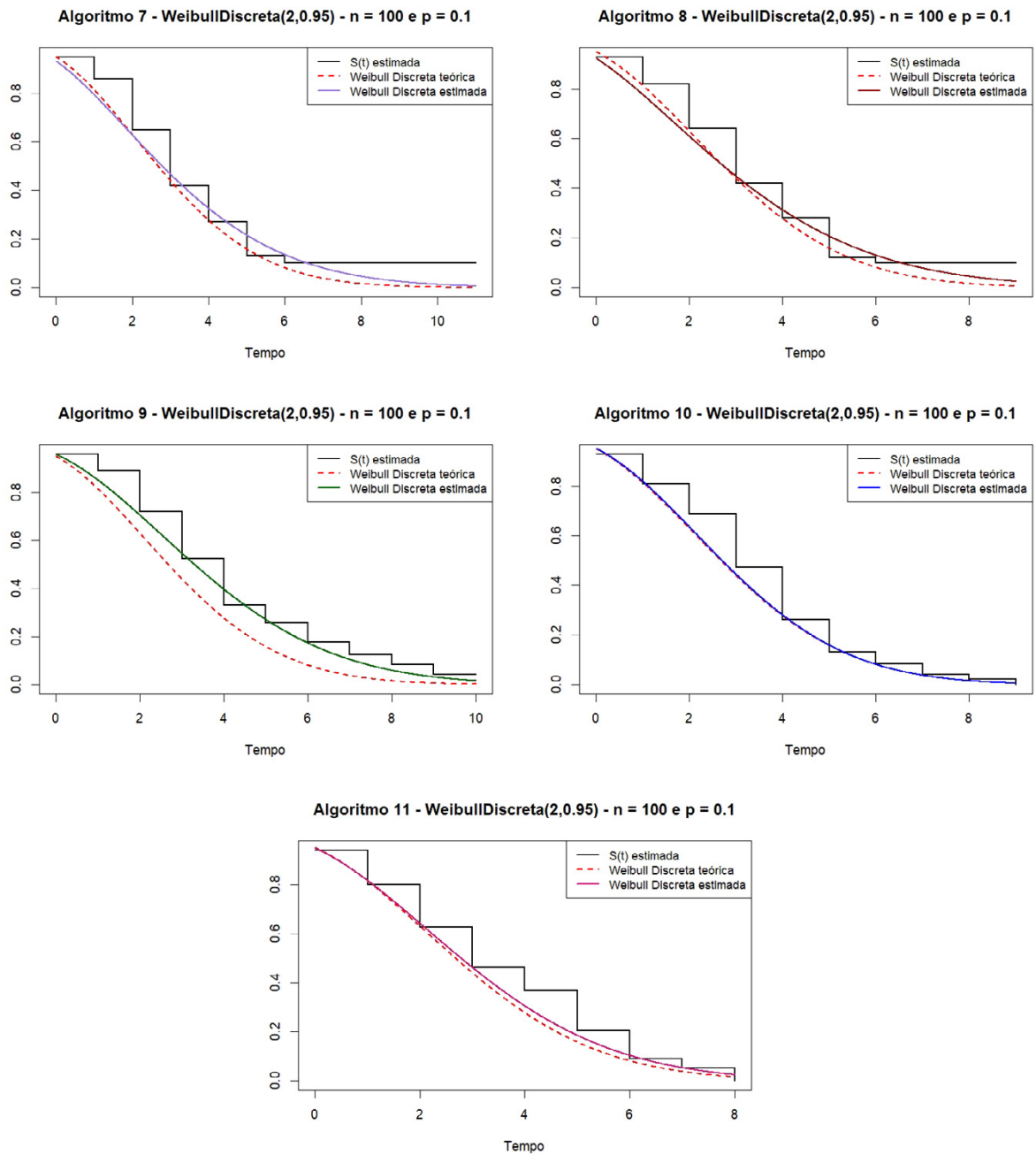


Figura I.4: Ajuste do modelo  $WeibullDiscreta(2; 0,95)$  a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 100 e percentual de censura de 10%

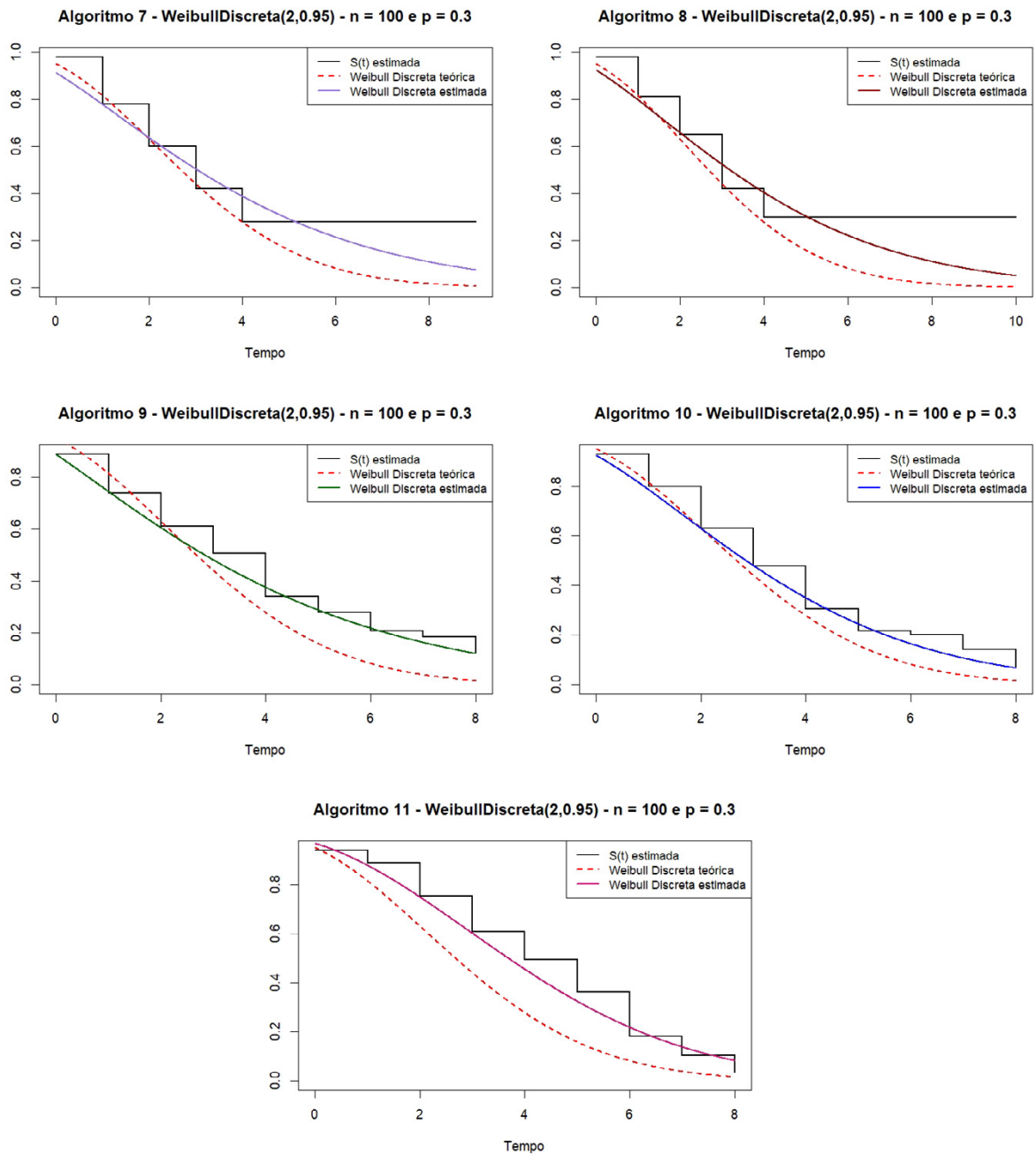


Figura I.5: Ajuste do modelo  $WeibullDiscreta(2; 0,95)$  a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 100 e percentual de censura de 30%

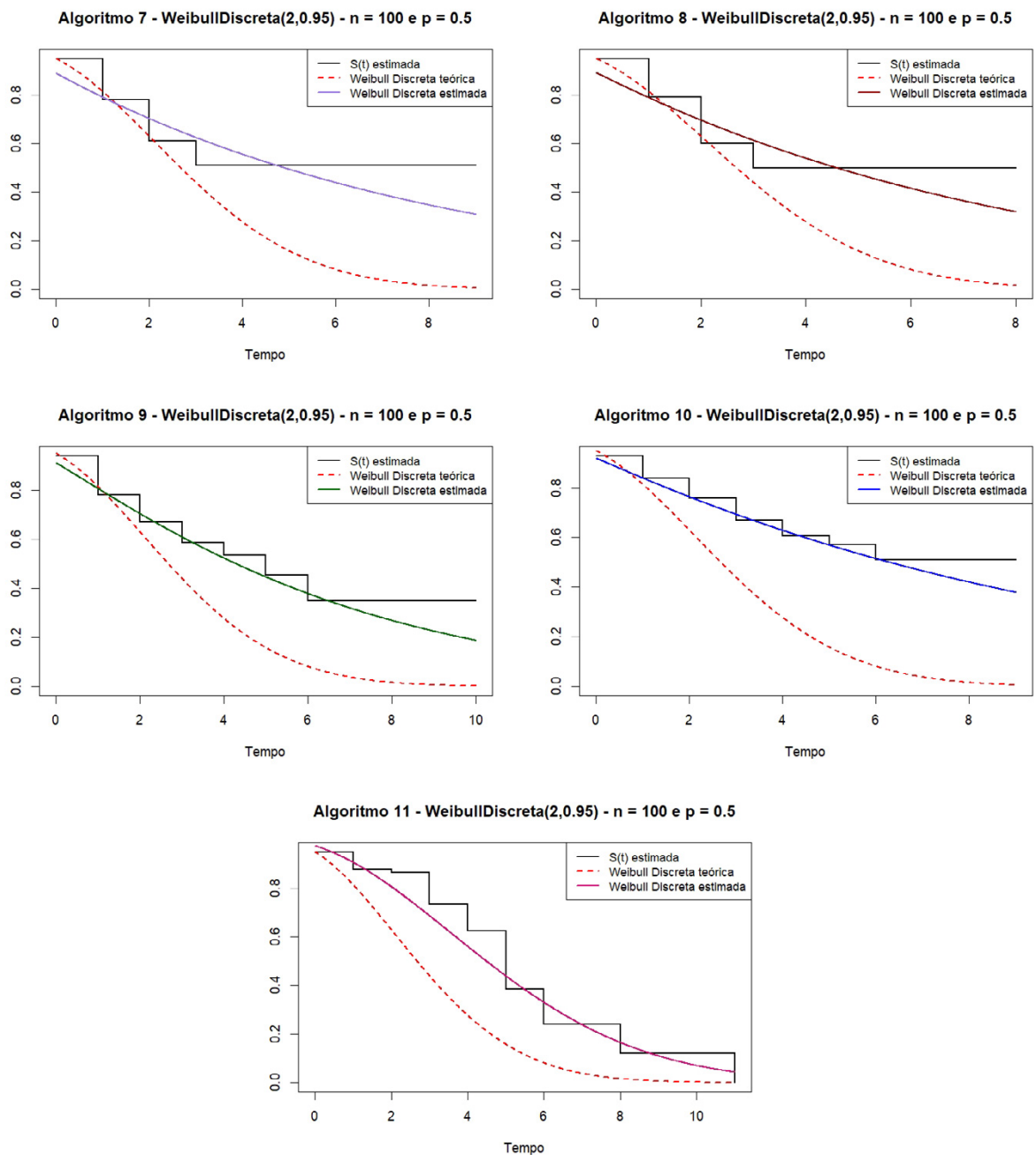


Figura I.6: Ajuste do modelo  $WeibullDiscreta(2;0,95)$  a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 100 e percentual de censura de 50%

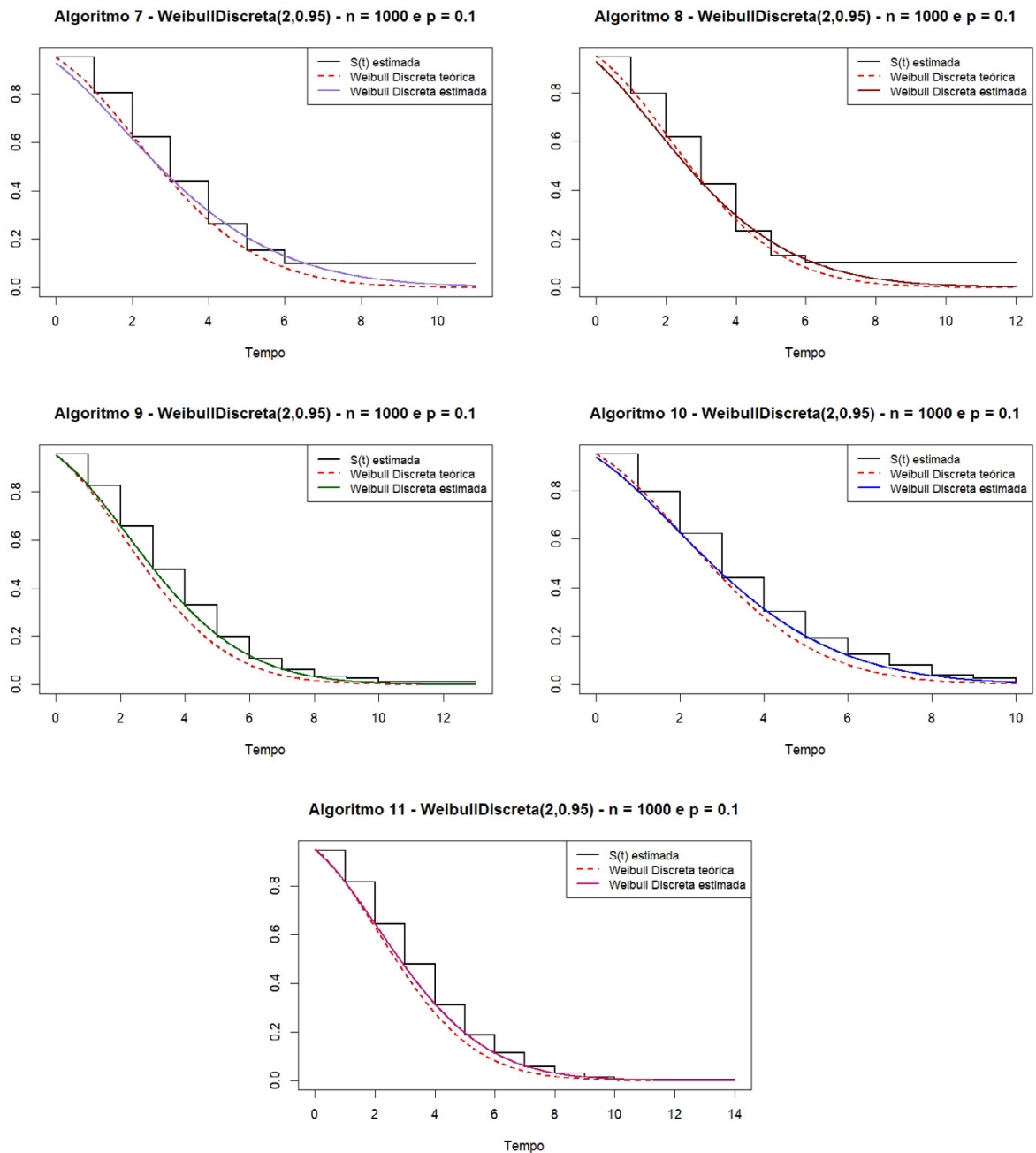


Figura I.7: Ajuste do modelo  $WeibullDiscreta(2;0,95)$  a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 1000 e percentual de censura de 10%

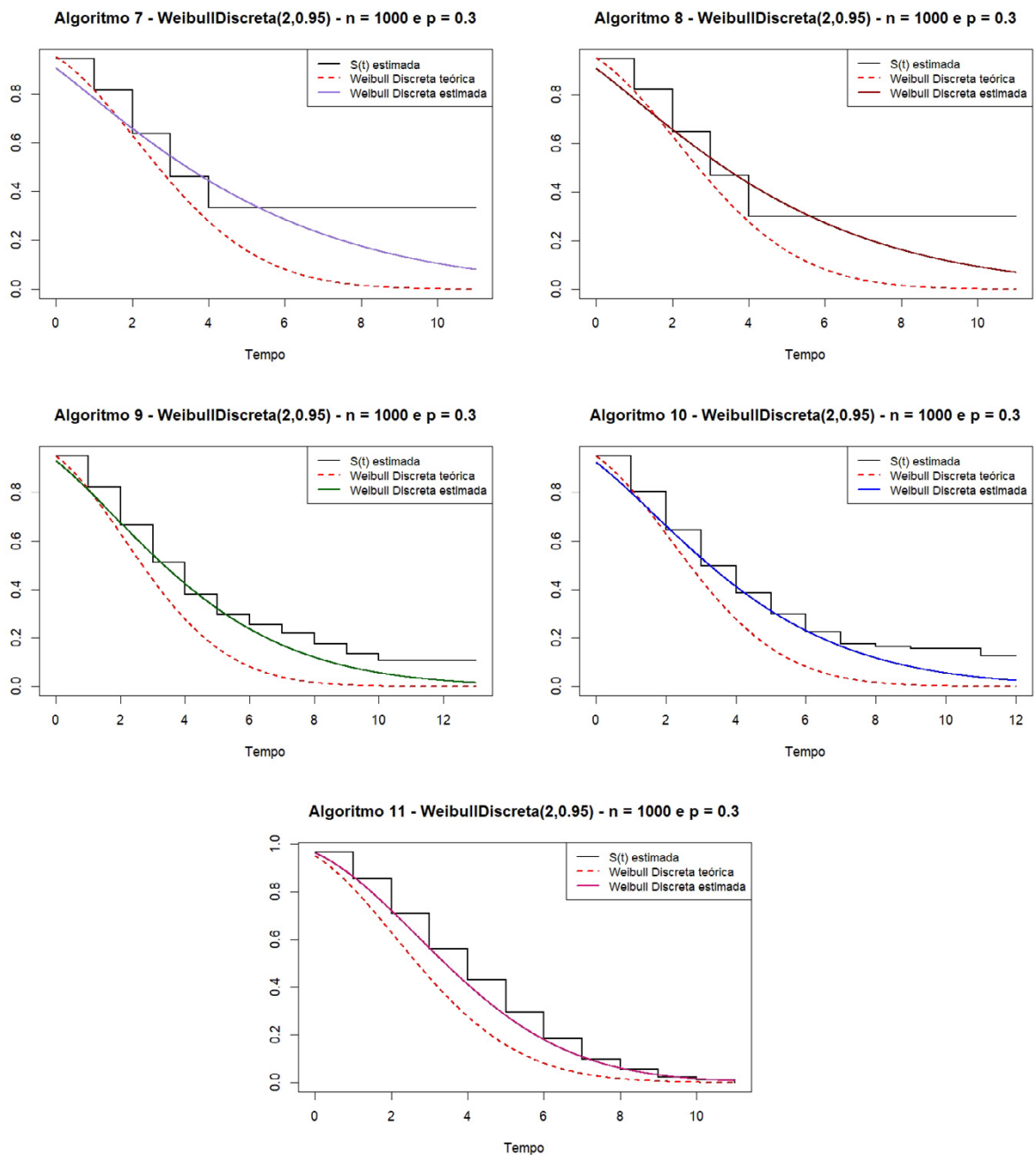


Figura I.8: Ajuste do modelo  $WeibullDiscreta(2; 0,95)$  a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 1000 e percentual de censura de 30%



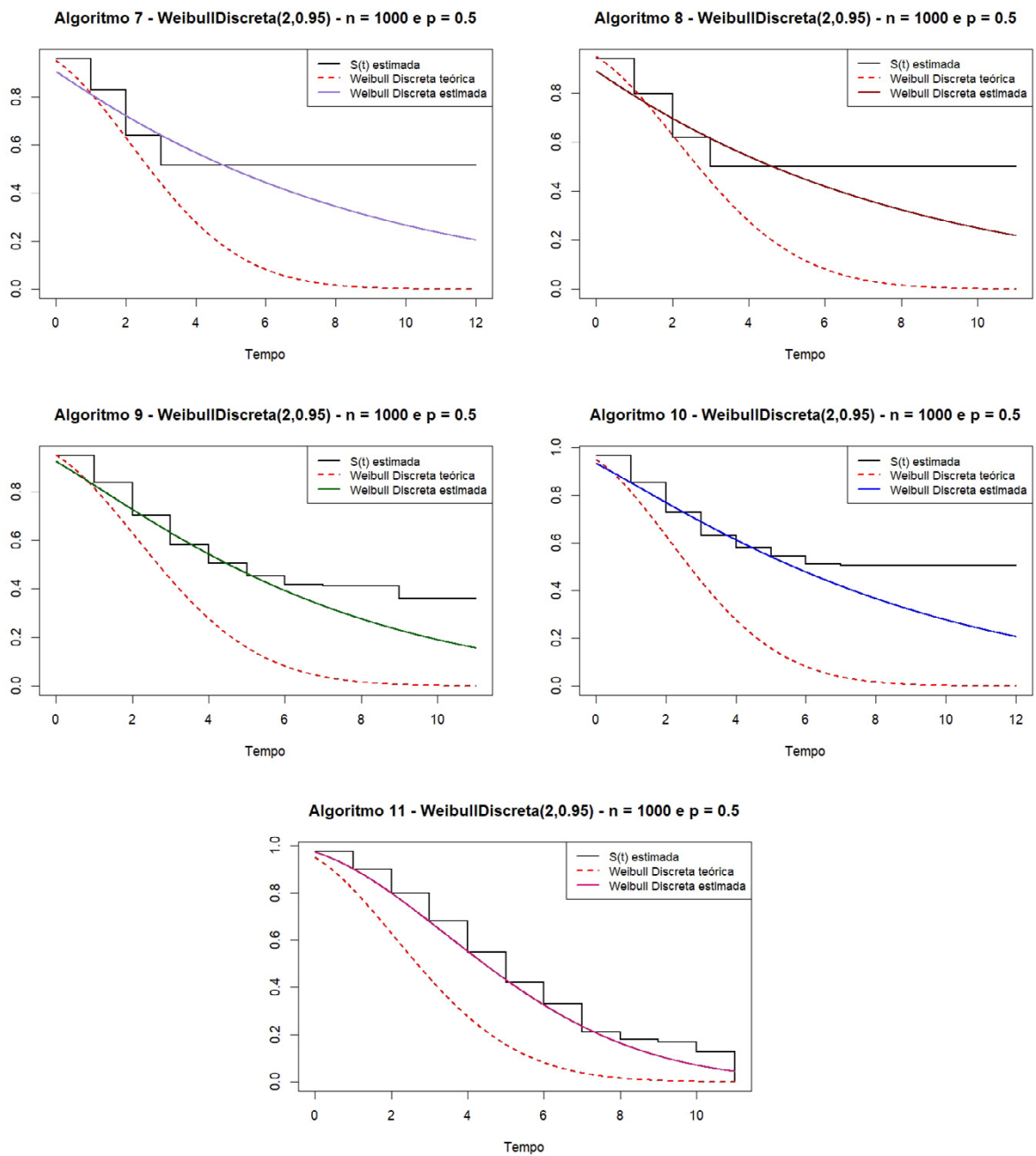


Figura I.9: Ajuste do modelo  $WeibullDiscreta(2;0,95)$  a Função de Sobrevivência estimada para os Algoritmos 7 a 11, amostra de tamanho 1000 e percentual de censura de 50%

# Referências Bibliográficas

- Amiri, N. et al. (2013). “Estimation of stress-strength parameter for two-parameter weibull distribution”. *International Journal of Advanced Statistics and Probability* 1.1, pp. 4–8.
- Balakrishnan, N. e Kundu, D. (2018). “Birnbbaum-Saunders distribution: A review of models, analysis, and applications”. *Appl Stochastic Models Bus Ind.*, pp. 1–46.
- Biazatti, E. C. e Nakano, E. Y. (2020). “Uma proposta de orientação para o uso de modelos contínuos em dados de sobrevivência discretos”. *REMAT: Revista Eletrônica da Matemática* 6.2, e4002.
- Bolfarine, H. e Sandoval, M. C. (2001). *Introdução à Inferência estatística*. 1ª ed. Sociedade Brasileira de Matemática.
- Breslow, N. E. e Crowley, J. (1974). “A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship”. *Annals of Statistics* 2.3, pp. 437–453.
- Brunello, G. H. V. e Nakano, E. Y. (2015). “Inferência bayesiana no modelo Weibull discreto em dados com presença de censura”. *TEMA - Tend. Mat. Apl. Comput.* 16.2, pp. 97–110.
- Cardial, M. R. P., Fachini-Gomes, J. B. e Nakano, E. Y. (2020). “Exponentiated discrete Weibull distribution for censored data”. *Rev. Bras. Biom.* 38.1, pp. 35–56.
- Casella, G. e Berger, R. L. (2010). *Inferência estatística*. 2ª ed. São Paulo: Cengage Learning.
- Colosimo, E. A. e Giolo, S. R. (2006). *Análise de Sobrevivência Aplicada*. São Paulo: Blucher. ISBN: 9788521215042.
- Cox, D. R. e Oakes, D. (1984). *Analysis of Survival Data*. New York: Chapman e Hall.

- Kaplan, E.L. e Meier, P. (1958). “Nonparametric Estimation from Incomplete Observations”. *Journal of the American Statistical Association* 53.282, pp. 457–481.
- Kim, Joongdae, Moon, Yeung-Gil e Kang, Junho (2003). “Inference on  $P(Y < X)$  in an Exponential Distribution”. *Journal of Korean* 14.4, pp. 989–995.
- Lawless, J. F. (2003). *Statistical models and methods for lifetime data*. 2ª ed. New York: John Wiley & Sons.
- Lindsey, J. C. e Ryan, L. M. (1998). “Tutorial in biostatistics methods for interval-censored data”. *Statistics in medicine* 17.2, pp. 219–238.
- Meier, P. (1975). “Estimation of a Distribution Function from Incomplete Observations”. 12, pp. 67–87.
- Meyer, P. L. (1983). *Probabilidade - Aplicações à Estatística*. 2ª ed. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A., p. 444.
- Nakano, E.Y. e Carrasco, C.G. (2006). “Uma avaliação do uso de um modelo contínuo na análise de dados discretos de sobrevivência”. *TEMA - Tend. Mat. Apl. Comput.* 7.1, pp. 91–100.
- Naznin, S. e Rahman, Md. S. (2015). “On Estimation of  $R=P(Y < X)$  for Birnbaum-Saunders Distribution”. *Research & Reviews: Journal of Statistics* 4.1, pp. 1–7.
- Nkemnole, E. B. e Samiyu, M. A. (2017). “Inference on Stress-Strength Reliability for Log-Normal Distribution based on Lower Record Values”. *AMSE JOURNALS-AMSE IIETA* 22.1, pp. 77–97.
- Vila, R., Nakano, E.Y. e Saulo, H. (2019). “Theoretical results on the discrete Weibull distribution of Nakagawa and Osaki”. *Statistics* 53.2, pp. 339–363.