



Ricardo Puziol de Oliveira

Aplicações de Alguns Métodos de Discretização nas Distribuições Lindley Contínuas

Maringá,
setembro, 2016

Ricardo Puziol de Oliveira

Aplicações de Alguns Métodos de Discretização nas Distribuições Lindley Contínuas

Dissertação apresentada à Universidade Estadual de Maringá, como parte das exigências do Programa de Pós-Graduação em Bioestatística, área de concentração em Estatística Aplicada, para a obtenção do título de Mestre.

Orientador: Josmar Mazucheli

Maringá,
setembro, 2016

**Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central - UEM, Maringá, PR, Brasil)**

O48a Oliveira, Ricardo Puziol de
Aplicações de alguns métodos de discretização nas distribuições Lindley contínuas / Ricardo Puziol de Oliveira. -- Maringá, 2016.
127 f. : il., figs., tabs.

Orientador: Prof. Dr. Josmar Mazucheli.
Dissertação (mestrado) - Universidade Estadual de Maringá, Departamento de Estatística, Programa de Pós-Graduação em Bioestatística, 2016.

1. Análise de sobrevivência. 2. Discretização (Estatística). 3. Variáveis aleatórias discretas. 4. Distribuição Lindley. 5. Função de probabilidade. 6. Monte Carlo, Método de. 7. Verossimilhança (Estatística). 8. Inferência estatística. I. Mazucheli, Josmar, orient. II. Universidade Estadual de Maringá. Departamento de Estatística. Programa de Pós-Graduação em Bioestatística. III. Título.

CDD 23.ed. 570.15195

AMMA-003358

RICARDO PUZIOL DE OLIVEIRA

**Aplicações de alguns métodos de Discretização nas Distribuições Lindley
Contínuas**

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

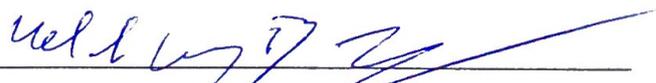
BANCA EXAMINADORA



Prof. Dr. Josmar Mazucheli
Universidade Estadual de Maringá - UEM



Prof. Dr. Jorge Alberto Achcar
Universidade de São Paulo - USP



Prof. Dr. Emílio Augusto Coelho Barros

Universidade Tecnológica Federal do Paraná - UFTPR

Maringá, 06 de setembro de 2016.

“Homens realmente grandes não nascem grandes, tornam-se grandes.”
Don Corleone - O Poderoso Chefão (1972)

AGRADECIMENTOS

Primeiramente agradeço a minha família, em especial meus pais, por todo apoio e por estarem a meu lado nessa jornada, por me darem todo o apoio nos momentos difíceis e por me incentivarem desde o princípio. E, principalmente por nunca terem desistido de mim e sempre acreditar em mim.

Ao meu orientador, professor Dr. Josmar Mazucheli por me ajudar em todo o processo de construção da dissertação, por todas as broncas e, principalmente, por dividir o seu conhecimento e propor temas de trabalho interessantes como o presente tema nessa dissertação.

A todos os professores do PBE que contribuíram para minha formação e também a CAPES por todo o apoio financeiro.

Aos meus amigos, que sempre estiveram ao meu lado tanto nos bons momentos quanto nos momentos ruins. Em especial a Márcia Lorena, que embarcou comigo no mestrado e por todos os anos de amizade. E também aos novos amigos que conheci durante o período de mestrado que fazem parte da minha vida. Agradeço a Isabela, Kelly, Omar, Emerson, Oilson, Marcos, Jardel, José André, Tiago, Beatriz, Jean, João, Guilherme, Aline e a todos que estiveram presente nessa caminhada.

RESUMO

Os métodos de construção de uma função de probabilidade a partir de uma função de densidade de probabilidade tem sido muito utilizado em anos recentes. No geral, o processo de discretização de uma variável aleatória contínua produz funções de probabilidade que podem ser concorrentes das distribuições tradicionais usadas na análise de dados de contagem — Poisson e Binomial negativa por exemplo. A discretização também evita o uso de uma distribuição contínua na análise de dados estritamente discretas. Pelo método baseado em uma série infinita, proposto por Good (1953), pelo método baseado na função de sobrevivência, proposto por Nakagawa e Osaki (1975), e pelo método baseado na função de risco, proposto por Roy e Ghosh (2009), o objetivo deste trabalho é propor novas distribuições discretas da distribuição Lindley (LINDLEY, 1958) e suas variantes Lindley potência (GHITANY et al., 2013) e Lindley ponderada (GHITANY et al., 2011) e também propor as versões truncada em zero, deslocada da origem e ponderada (apenas no caso da distribuição Lindley) para estas novas distribuições. Para cada distribuições propostas foram realizados estudos de simulação de Monte Carlo para avaliar o vício e o erro-quadrático-médio dos estimadores de máxima verossimilhança. Duas aplicações, em que não se observa resposta igual a zero, são consideradas para ilustrar a aplicabilidade das distribuições propostas. A primeira utiliza dados referentes ao comprimento da sequência de dias chuvosos e a segunda dados do comprimento da sequência de corridas positivas, em relação ao valor de retorno de ativos, de ações do mercado financeiro.

Palavras-chaves: Análise de sobrevivência, discretização, distribuição Lindley, função de probabilidade, Monte Carlo, verossimilhança.

ABSTRACT

The methods of constructing a probability function from a probability density function has long been used in recent years. In general the discretization process of a continuous variable produces probability functions that can be rivals to traditional distributions used in the count data analysis — Poisson and negative binomial for example. The discretization also avoids the use of a continuous distribution in the analysis of strictly discrete data. By the method based on an infinite series, proposed by Good (1953), by the method based on the survival function, proposed by Nakagawa e Osaki (1975), and by the method based on the hazard rate function, proposed by Roy e Ghosh (2009), the objective of this study is to propose new discrete analogues of Lindley (LINDLEY, 1958) distribution and the variants power Lindley distribution (GHITANY et al., 2013) and weighted Lindley distribution (GHITANY et al., 2011) and also propose zero-truncated, shifted and weighted (just in case of Lindley distribution) for these new distributions. For each proposed distributions were carried out Monte Carlo simulation studies to evaluate BIAS and the MSE of maximum likelihood estimators. Two applications where not observed response $x = 0$ are considered to illustrate the applicability of the proposed distributions. The first refers to the length of wet spells and the second to the length of positive runs in the financial market prices.

Key-words: Survival analysis, discretization, Lindley distribution, probability function, Monte Carlo simulation, likelihood.

LISTA DE ILUSTRAÇÕES

Figura 1 – Comportamento da função de probabilidade da distribuição Lindley discreta (via função de sobrevivência) para alguns valores de β	33
Figura 2 – Comportamento da função de risco da distribuição Lindley discreta (via função de sobrevivência) para alguns valores de β	34
Figura 3 – Esperança e variância distribuição Lindley discreta (via função de sobrevivência) para alguns valores de β	35
Figura 4 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta (via função de sobrevivência).	36
Figura 5 – Comportamento da função de probabilidade da distribuição Lindley discreta (via série infinita) para alguns valores de β	37
Figura 6 – Comportamento da função de risco da distribuição Lindley discreta (via série infinita) para alguns valores de β	38
Figura 7 – Esperança e variância distribuição Lindley discreta (via série infinita) para alguns valores de β	39
Figura 8 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta (via série infinita).	40
Figura 9 – Comportamento da função de risco da distribuição Lindley discreta (via função de risco) para alguns valores de β	41
Figura 10 – Comportamento da função de probabilidade da distribuição Lindley potência discreta para diferentes valores de α e β	43
Figura 11 – Comportamento da função de risco da distribuição Lindley potência discreta para alguns valores de α e β	44
Figura 12 – Esperança e variância distribuição Lindley potência discreta para alguns valores de α e $\beta = 3$	45
Figura 13 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta.	47
Figura 14 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta.	47
Figura 15 – Comportamento da função de probabilidade da distribuição Lindley ponderada discreta para diferentes valores de θ e β	49

Figura 16 – Comportamento da função de risco da distribuição Lindley ponderada discreta para diferentes valores de θ e β	50
Figura 17 – Esperança e variância distribuição Lindley ponderada discreta para alguns valores de θ e $\beta = 3$	51
Figura 18 – Vício e EQM de $\hat{\theta}$ da distribuição Lindley ponderada discreta.	53
Figura 19 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley ponderada discreta.	53
Figura 20 – Comportamento da função de probabilidade da distribuição Lindley discreta truncada em zero para alguns valores de β	55
Figura 21 – Comportamento da função de risco da distribuição Lindley discreta truncada em zero (via função de sobrevivência) para diferentes valores de β	56
Figura 22 – Esperança e variância distribuição Lindley discreta truncada em zero (via função de sobrevivência) para alguns valores de β	56
Figura 23 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta truncada em zero (via função de sobrevivência).	57
Figura 24 – Comportamento da função de probabilidade da distribuição Lindley discreta truncada em zero (via série infinita) para alguns valores de β	59
Figura 25 – Comportamento da função de risco da distribuição Lindley discreta truncada em zero (via série infinita) para alguns valores de β	59
Figura 26 – Esperança e variância distribuição Lindley discreta truncada em zero (via série infinita) para alguns valores de β	59
Figura 27 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta truncada em zero (via série infinita).	60
Figura 28 – Comportamento da função de probabilidade da distribuição Lindley potência discreta truncada em zero para diferentes valores de α e β	61
Figura 29 – Comportamento da função de probabilidade da distribuição Lindley potência discreta truncada em zero para diferentes valores de α e β	62
Figura 30 – Esperança e variância distribuição Lindley potência discreta truncada em zero para alguns valores de α e $\beta = 1$	63
Figura 31 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta truncada em zero.	65
Figura 32 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta truncada em zero.	65
Figura 33 – Comportamento da função de probabilidade da distribuição Lindley ponderada discreta truncada em zero para diferentes valores de θ e β	66
Figura 34 – Comportamento da função de risco da distribuição Lindley ponderada discreta truncada em zero para diferentes valores de θ e β	67
Figura 35 – Comportamento da função de probabilidade da distribuição Lindley discreta deslocada da origem (via função de sobrevivência) para alguns valores de β	68

Figura 36 – Comportamento da função de risco da distribuição Lindley discreta deslocada da origem (via função de sobrevivência) para alguns valores de β	69
Figura 37 – Esperança e variância distribuição Lindley discreta deslocada da origem (via função de sobrevivência) para alguns valores de β	69
Figura 38 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta deslocada da origem (via função de sobrevivência).	70
Figura 39 – Comportamento da função de probabilidade da distribuição Lindley discreta deslocada da origem (via série infinita) para diferentes valores de β	71
Figura 40 – Comportamento da função de risco da distribuição Lindley discreta deslocada da origem (via série infinita) para alguns valores de β	72
Figura 41 – Esperança e variância distribuição Lindley discreta deslocada da origem (via série infinita) para alguns valores de β	72
Figura 42 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta deslocada da origem (via série infinita).	73
Figura 43 – Comportamento da função de probabilidade da distribuição Lindley potência discreta deslocada da origem para diferentes valores de α e β	74
Figura 44 – Comportamento da função de risco da distribuição Lindley potência discreta deslocada da origem para diferentes valores de α e β	75
Figura 45 – Esperança e variância distribuição Lindley potência discreta deslocada da origem para alguns valores de α e $\beta = 1$	75
Figura 46 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta deslocada da origem.	76
Figura 47 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta deslocada da origem.	76
Figura 48 – Comportamento da função de probabilidade da distribuição Lindley ponderada discreta deslocada da origem para diferentes valores de θ e β	78
Figura 49 – Comportamento da função de risco da distribuição Lindley ponderada discreta deslocada da origem para diferentes valores de θ e β	79
Figura 50 – Vício e EQM de $\hat{\theta}$ da distribuição Lindley ponderada discreta deslocada da origem.	79
Figura 51 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley ponderada discreta deslocada da origem.	80
Figura 52 – Comportamento da função de probabilidade da distribuição Lindley discreta ponderada (via função de sobrevivência) para alguns valores de β	82
Figura 53 – Comportamento da função de risco da distribuição Lindley discreta ponderada (via função de sobrevivência) para alguns valores de β	83
Figura 54 – Esperança e variância distribuição Lindley discreta ponderada (via função de sobrevivência) para alguns valores de β	83

Figura 55 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta ponderada (via função de sobrevivência).	84
Figura 56 – Comportamento da função de probabilidade da distribuição Lindley discreta ponderada (via série infinita) para alguns valores de β	85
Figura 57 – Comportamento da função de risco da distribuição Lindley discreta ponderada (via série infinita) para alguns valores de β	86
Figura 58 – Esperança e variância distribuição Lindley discreta ponderada (via série infinita) para alguns valores de β	86
Figura 59 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta ponderada (via série infinita).	87
Figura 60 – CDF empírica vs CDF ajustada das distribuições LD, LDS, LPD, LPonD, POIS, GEO, B, DBURR e NB em relação ao número de molares cariados.	90
Figura 61 – CDF empírica vs CDF ajustada das distribuições LD, LDS, LPD, LPonD, POIS, GEO, B, DBURR e NB em relação ao número de vezes que o computador falhou.	92
Figura 62 – Função de probabilidade empírica vs função de probabilidade ajustada das distribuições LDTZ ¹ , LDD ¹ , LPondD ¹ e ZTNB para as estações de Santa Catarina.	109
Figura 63 – Função de probabilidade empírica vs função de probabilidade ajustada das distribuições LDTZ ¹ , LDD ¹ , LPondD ¹ e ZTNB para as estações do Paraná.	110
Figura 64 – Função de probabilidade empírica vs função de probabilidade ajustada das distribuições LDTZ ¹ , LDD ¹ , LPondD ¹ e ZTNB para as estações do Rio Grande do Sul.	111
Figura 65 – Função de probabilidade empírica vs função de probabilidade ajustada das distribuições LDTZ ¹ , LDD ¹ , LPondD ¹ e ZTNB para as estações do Rio Grande do Sul.	112

LISTA DE TABELAS

Tabela 1 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta (via função de sobrevivência).	36
Tabela 2 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta (via série infinita).	40
Tabela 3 – Índice de dispersão da distribuição Lindley potência discreta.	45
Tabela 4 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta.	48
Tabela 5 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta.	48
Tabela 6 – Índice de dispersão da distribuição Lindley ponderada discreta.	52
Tabela 7 – Vício e EQM de $\hat{\theta}$ da distribuição Lindley ponderada discreta.	54
Tabela 8 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley ponderada discreta.	54
Tabela 9 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta truncada em zero (via função de sobrevivência).	58
Tabela 10 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta truncada em zero (via série infinita).	60
Tabela 11 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta truncada em zero.	64
Tabela 12 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta truncada em zero.	64
Tabela 13 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta deslocada da origem (via função de sobrevivência).	70
Tabela 14 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta deslocada da origem (via série infinita).	73
Tabela 15 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta deslocada da origem.	77
Tabela 16 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta deslocada da origem.	77
Tabela 17 – Vício e EQM de $\hat{\theta}$ da distribuição Lindley ponderada discreta deslocada da origem.	80

Tabela 18 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley ponderada discreta deslocada da origem.	81
Tabela 19 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta ponderada (via função de sobrevivência).	84
Tabela 20 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta ponderada (via série infinita).	87
Tabela 21 – Resumo das distribuições propostas.	88
Tabela 22 – N° de molares cariados observados e esperados pelas distribuições LPD, LPonD, LDS, LD, DBURR, NB, B, POIS e GEO	89
Tabela 23 – Estimativas dos parâmetros e critérios de comparação das distribuições LD, LDS, LPD, LPonD, POIS, WEI, GD, BURR e NB.	89
Tabela 24 – N° de vezes que o computador falha observadas e esperados pelas distribuições LPD, LPonD, LDS, LD, DBURR, NB, B, POIS e GEO	91
Tabela 25 – Estimativas dos parâmetros e critérios de comparação das distribuições LD, LDS, LPD, LPonD, POIS, WEI, GD, BURR e NB.	91
Tabela 26 – Características do banco de dados para cada estação.	97
Tabela 27 – Número de séries ajustadas para cada distribuição em relação ao comprimento da sequência de dias chuvosos.	99
Tabela 28 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83923.	99
Tabela 29 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83766.	100
Tabela 30 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83767.	100
Tabela 31 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83783.	100
Tabela 32 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83811.	100
Tabela 33 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83813.	101
Tabela 34 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83836.	101
Tabela 35 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83842.	101
Tabela 36 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83844.	101
Tabela 37 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83881.	102

Tabela 38 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83907.	102
Tabela 39 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83912.	102
Tabela 40 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83914.	102
Tabela 41 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83916.	103
Tabela 42 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83919.	103
Tabela 43 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83927.	103
Tabela 44 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83936.	103
Tabela 45 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83941.	104
Tabela 46 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83942.	104
Tabela 47 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83948.	104
Tabela 48 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83953.	104
Tabela 49 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83964.	105
Tabela 50 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83967.	105
Tabela 51 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83980.	105
Tabela 52 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83985.	105
Tabela 53 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83995.	106
Tabela 54 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83997.	106
Tabela 55 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83872.	106
Tabela 56 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83883.	106

Tabela 57 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83887.	107
Tabela 58 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83891.	107
Tabela 59 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83897.	107
Tabela 60 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83920.	107
Tabela 61 – Características do banco de dados para cada empresa.	115
Tabela 62 – Número de séries ajustadas para cada distribuição em relação ao comprimento da sequência de corridas positivas.	117

LISTA DE ABREVIATURAS E SIGLAS

G - Geométrica

B - Binomial

DBURR - Burr XII discreta

NB - Binomial Negativa

LD - Lindley discreta via função de sobrevivência

LDTZ - Lindley discreta truncada em zero via função de sobrevivência

SLD - Lindley discreta deslocada da origem via função de sobrevivência

LDP - Lindley discreta ponderada via função de sobrevivência

LDS - Lindley discreta via série infinita

LDSTZ - Lindley discreta truncada em zero via série infinita

SLDS - Lindley discreta deslocada da origem via série infinita

LDSP - Lindley discreta ponderada via série infinita

LDR - Lindley discreta via função de risco

LPD - Lindley potência discreta

LPDTZ - Lindley potência discreta truncada em zero

SLPD - Lindley potência discreta deslocada da origem

LPonD - Lindley ponderada discreta

LPonDTZ - Lindley ponderada discreta truncada em zero

SLPonD - Lindley ponderada discreta deslocada da origem

SUMÁRIO

1	Visão Geral	19
1.1	Introdução	19
1.2	Justificativa	19
1.3	Motivação	20
1.4	Objetivos	20
1.4.1	Objetivo Geral	20
1.4.2	Objetivos Específicos	20
1.5	Organização do Trabalho	21
1.6	Conclusões	21
1.7	Propostas Futuras	22
1.8	Artigos Encaminhados E Pacote do R	22
2	Métodos de Discretização de Variáveis Aleatórias Contínuas	23
2.1	Introdução	23
2.2	Métodos de Discretização	26
2.2.1	Discretização via Função de Sobrevivência	26
2.2.2	Discretização via uma Série Infinita	27
2.2.3	Discretização via Função de Risco	29
3	Discretização da Distribuição Lindley e Algumas de Suas Variações	31
3.1	Introdução	31
3.2	Discretizações da Distribuição Lindley	32
3.2.1	Distribuição Lindley Discreta (via Função de Sobrevivência)	32
3.2.2	Distribuição Lindley Discreta (via Série Infinita)	36
3.2.3	Distribuição Lindley Discreta (via Função de Risco)	40
3.2.4	Distribuição Lindley Potência Discreta	42
3.2.5	Distribuição Lindley Ponderada Discreta	49
3.3	Modelos Discretos Truncados em Zero	54
3.3.1	Distribuição Lindley Discreta Truncada em Zero (via Função de Sobrevivência)	55
3.3.2	Distribuição Lindley Discreta Truncada em Zero (via Série Infinita)	58
3.3.3	Distribuição Lindley Potência Discreta Truncada em Zero	61
3.3.4	Distribuição Lindley Ponderada Discreta Truncada em Zero	65
3.4	Modelos Discretos Deslocados da Origem	67

3.4.1	Distribuição Lindley Discreta Deslocada da Origem (via Função de Sobrevivência)	68
3.4.2	Distribuição Lindley Discreta Deslocada da Origem (via Série Infinita)	71
3.4.3	Distribuição Lindley Potência Discreta Deslocada da Origem	73
3.4.4	Distribuição Lindley Ponderada Deslocada da Origem	78
3.5	Modelos Ponderados	81
3.5.1	Distribuição Lindley Discreta Ponderada (via Função de Sobrevivência)	81
3.5.2	Distribuição Lindley Discreta Ponderada (via Série Infinita)	85
3.6	Resumo das Distribuições Propostas	88
3.7	Aplicação em Dados da Literatura	89
3.7.1	Aplicação 1	89
3.7.2	Aplicação 2	91
4	Modelando o Comprimento das Sequências de Dias Chuvosos	93
4.1	Introdução	93
4.2	Distribuições de Probabilidade	94
4.2.1	Binomial Negativa Truncada em Zero	95
4.2.2	Poisson Truncada em Zero	96
4.3	Os Dados	96
4.4	Resultados e Discussões	98
4.5	Conclusões	108
5	Modelando o Comprimento das Sequências Positivas	113
5.1	Introdução	113
5.2	Distribuições de Probabilidade	114
5.3	Os Dados	115
5.4	Resultados e Discussões	116
6	Conclusões e Propostas Futuras	118
6.1	Conclusões	118
6.2	Propostas Futuras	119
	Referências	121

VISÃO GERAL

1.1 Introdução

O processo de discretização tem sido estudado por muitos autores, como Nakagawa e Osaki (1975), Good (1953), Roy e Ghosh (2009), nas últimas décadas. Basicamente, este processo produz funções de probabilidade que podem ser concorrentes às distribuições tradicionais usadas nas análises de dados de contagem — Poisson e Binomial Negativa por exemplo. A discretização também evita o uso de uma distribuição contínua em dados de natureza estritamente discreta.

Na literatura existem diversas formas de se obter uma distribuição discreta a partir de uma distribuição contínua. Todavia, neste trabalho, será abordado apenas três métodos: o método de discretização baseado na função de sobrevivência (NAKAGAWA; OSAKI, 1975), o método de discretização baseado em uma série infinita (GOOD, 1953; KULASEKERA; TONKYN, 1992; KEMP, 1997; SATO et al., 1999) e o método de discretização baseado na função de risco (ROY; GHOSH, 2009).

1.2 Justificativa

Os modelos de probabilidade tradicionais como o de Poisson e o Geométrico, em geral, tem aplicabilidade limitada quando se trata de tempo de falha, confiabilidade, ou até mesmo contagem. De fato, no caso do modelo geométrico, a função de risco possui comportamento constante e, no caso do modelo de Poisson, a função de risco tem comportamento crescente não-dependente do parâmetro do modelo. Além disso, a pouca variedade de modelos em que o suporte é discreto e definido para $x > 0$, é uma motivação para propor novos modelos para a análise de dados desse tipo.

Logo, propomos neste trabalho novos modelos discretos resultantes da discretização da distribuição Lindley e algumas de suas variantes e as versões truncada em zero, deslocada da origem e ponderada dos mesmos como uma alternativa aos modelos tradicionais.

1.3 Motivação

Alguns trabalhos que serviram de motivação para o presente estudo sobre discretização de uma variável aleatória contínua:

- A distribuição discreta de Valores Extremos (NAKAGAWA, 1978);
- A distribuição Gama discreta (YANG, 1994);
- As distribuições Normal e de Laplace discretas (LISMAN; ZUYLEN, 1972; INUSAH; KOZUBOWSKI, 2006);
- A distribuição de Skew-Laplace discreta (KOZUBOWSKI; INUSAH, 2006);
- As distribuições Burr e Pareto discretas (KRISHNA; PUNDIR, 2009);
- A distribuição Maxwell (KRISHNA; PUNDIR, 2007);
- A distribuição Exponencial discreta Marshall-Olkin (MARSHALL; OLKIN, 1997);
- A distribuição Weibull inversa discreta (JAZI; LAI; ALAMATSAZ, 2010);
- A distribuição Lindley discreta (GÓMEZ-DÉNIZ; CALDERÍN-OJEDA, 2011; BAKOUCH; JAZI; NADARAJAH, 2014);
- A distribuição Beta Exponencial discreta (NEKOUKHOU et al., 2015);
- A distribuição Rayleigh inversa discreta (HUSSAIN; AHMAD, 2014);
- A distribuição de Séries de Dirichlet (SIROMONEY, 1964).

1.4 Objetivos

1.4.1 Objetivo Geral

Propor novas distribuições discretas resultantes das discretizações das distribuições Lindley (LINDLEY, 1958), Lindley Potência (GHITANY et al., 2013) e Lindley Ponderada (GHITANY et al., 2011) a fim ilustrar novas alternativas para a modelagem de dados de natureza estritamente discreta.

1.4.2 Objetivos Específicos

- Por meio dos métodos de discretização de Nakagawa e Osaki (1975), de Good (1953), Sato et al. (1999) e de Roy e Ghosh (2009), propor novas distribuições discretas baseadas nas distribuições Lindley e suas variantes;

- Caracterizar e estudar algumas propriedades destas novas distribuições;
- Estudar, via estudos de simulação Monte Carlo, algumas propriedades dos estimadores de máxima verossimilhança;
- Propor versões truncadas em zero, deslocadas da origem e ponderada dessas novas distribuições discretas para a análise de dados em que a resposta $x = 0$ não é observada;
- Apresentar aplicações em dados reais para verificar a qualidade do ajuste destas novas distribuições em comparação aos modelos discretos tradicionais.
- Modelar o comprimento das sequências de dias chuvosos na região sul do Brasil;
- Modelar o comprimento das sequências de corridas positivas no mercado de retorno do preço de abertura das ações da bolsa de valores.

1.5 Organização do Trabalho

O trabalho está organizado como segue: no Capítulo 2 são apresentados em detalhes os métodos de discretização utilizados na discretização da distribuição Lindley e as variantes Lindley potência (GHITANY et al., 2013) e Lindley ponderada (GHITANY et al., 2011). No Capítulo 3, são apresentadas as versões discretas das distribuições supracitadas juntamente as respectivas versões truncada em zero e deslocada da origem. E, no caso da discretização da distribuição Lindley considerada por Bakouch, Jazi e Nadarajah (2014) e da discretização da mesma feita pelo método de Good (1953), são apresentadas as versões truncada em zero, deslocada da origem e ponderada.

No Capítulo 4, as distribuições propostas são aplicadas na análise de dados que descrevem o comprimento das sequências de dias chuvosos da região sul do Brasil e, no Capítulo 5, as mesmas são aplicadas na modelagem do comprimento das sequências de dias em que houve corrida positiva no mercado de retornos no preço de abertura das ações da bolsa de valores. Na literatura, dados dessa natureza são geralmente analisados considerando as distribuições Poisson truncada em zero e Binomial Negativa truncada em zero (DENI; JEMAIN, 2008a). Por fim, o Capítulo 6 finaliza este trabalho com algumas conclusões e propostas futuras.

1.6 Conclusões

Em suma, conclui-se que as discretizações das distribuições Lindley e suas variantes Lindley potência e Lindley ponderada oferecem um ajuste significativo para análise de dados climatológicos e financeiro consistentes, respectivamente, em sequência de dias chuvosos e em sequência de dias em que se houve corridas positivas quando comparadas aos modelos tradicionais utilizados em análise de dados dessa natureza. Além disso, os métodos de

discretização utilizados se mostraram como alternativa flexível para o desenvolvimento de distribuições discretas a partir de distribuições contínuas com objetivo de se analisar dados de natureza discreta.

1.7 Propostas Futuras

Com objetivo de continuar o estudo de discretização de uma variável aleatória contínua, estabelecemos como propostas para um trabalho futuro:

- A aplicação de outros métodos de discretização presentes na literatura;
- A discretização de outras variantes da distribuição Lindley;
- Modelagem de dados de sobrevivência;
- Investigar novas propriedades dos modelos propostos;
- Realizar uma comparação entre os principais métodos de discretização;
- Trabalhar com a inflação de zeros para os modelos propostos.

1.8 Artigos Encaminhados E Pacote do R

- *A Discrete Analogue of the Power Lindley Distribution and its Applications* by Josmar Mazucheli, Ricardo P. Oliveira, Marcia L. Alves and Kelly V. Barco. - Em avaliação pela revista: Applied Mathematics and Computation.
- *A Comparative Study Between Two Discrete Lindley Distributions* by Josmar Mazucheli and Ricardo P. Oliveira. - Em processo de submissão.
- *A New Generalized Discrete Lindley Distribution* by Josmar Mazucheli, M. E. Ghitany and Ricardo P. Oliveira. - Em processo de submissão.
- *The LindleyR Package* by Josmar Mazucheli, Larissa B. and Ricardo P. Oliveira.

MÉTODOS DE DISCRETIZAÇÃO DE VARIÁVEIS ALEATÓRIAS CONTÍNUAS

2.1 Introdução

Em algumas situações, é difícil ou até mesmo inconveniente obter amostras de uma distribuição contínua. Segundo Chakraborty (2015), isso se dá ao fato de que quase sempre os dados observados são de natureza discreta e, mesmo que as medidas forem tomadas em uma escala contínua, as observações podem ser registradas de forma em que o uso de um modelo discreto seja mais apropriado.

Por exemplo, na análise de dados de sobrevivência ou confiabilidade, não é incomum usar a distribuição Weibull e a distribuição Geométrica para descrever o número de ciclos que um equipamento suporta antes de vir a falhar. Uma infinidade de outros exemplos podem ser encontrados nos principais livros da área cujo autores são: Rupert (2011), Hamada et al. (2008), Collett (2003), Lee e Wang (2003), Lawless (2003), Kalbfleisch e Prentice (2002), Meeker e Escobar (1998), Klein e Moeschberger (1997).

Já em estudos de estresse-força, de acordo com Roy e Ghosh (2009), se são conhecidas as distribuições da força e do estresse, a confiabilidade pode ser estimada utilizando técnicas de transformação. No entanto, quando as relações funcionais de força e estresse são complexas, tais técnicas analíticas são intratáveis e, neste caso, as técnicas de discretização são uma alternativa.

Diante as situações supracitadas, nas últimas décadas, a construção de uma função de probabilidade a partir da discretização de uma variável aleatória contínua tem sido estudado por diversos autores como, por exemplo, Good (1953), Nakagawa e Osaki (1975), Roy e Ghosh (2009) e Ghosh, Roy e Chandra (2013). De acordo com Boulle (2004), o processo de discretização de uma variável aleatória contínua consiste em transformar seus atributos contínuos em atributos discretos e, principalmente, redefinir seu suporte (domínio) de \mathbb{R} para um suporte (domínio) definido em \mathbb{Z} . Basicamente, este processo produz funções de probabi-

lidade que podem ser concorrentes às distribuições tradicionais usadas nas análises de dados de contagem — Poisson e Binomial Negativa por exemplo. A discretização também evita o uso de uma distribuição contínua em dados de natureza estritamente discreta.

Na literatura existem diversos métodos de se obter uma distribuição discreta a partir de uma distribuição contínua, dentre eles destacamos: o método de discretização baseado na função de sobrevivência (NAKAGAWA; OSAKI, 1975), o método de discretização baseado em uma série infinita (GOOD, 1953; KULASEKERA; TONKYN, 1992; KEMP, 1997; SATO et al., 1999), o método via função de risco (STEIN; DATTERO, 1984), o método de duas fases compostas (CHAKRABORTY, 2015), o método via função de risco reversa (GHOSH; ROY; CHANDRA, 2013), o método via função de distribuição acumulada (ROY; DASGUPTA, 2001) e o método de equações diferenciais, análogo as equações diferenciais de Pearson (PEARSON, 1895). Todavia, neste trabalho, abordamos apenas os três primeiros métodos citados. Uma recente revisão de todos esses métodos de discretização e das distribuições discretizadas por eles pode ser encontrada em (CHAKRABORTY, 2015).

Uma das primeiras distribuições contínuas discretizadas a ser introduzida na literatura foi a distribuição Weibull discreta proposta por Nakagawa e Osaki (1975) em estudos de tempo de falha. Considerando a distribuição Weibull com função densidade de probabilidade:

$$f(x | \mu, \beta) = \frac{\beta}{\mu^\beta} x^{\beta-1} \exp \left[- \left(\frac{x}{\mu} \right)^\beta \right] \quad (2.1)$$

e a função de sobrevivência:

$$S(x | \mu, \beta) = \exp \left[- \left(\frac{x}{\mu} \right)^\beta \right] \quad (2.2)$$

A esperança e a variância de uma variável aleatória contínua com distribuição Weibull são dadas, respectivamente, por:

$$\mathbb{E}(X) = \mu \Gamma \left(1 + \frac{1}{\beta} \right) \quad (2.3)$$

e,

$$\text{Var}(X) = \mu^2 \left[\Gamma \left(1 + \frac{2}{\beta} \right) - \left[\Gamma \left(1 + \frac{1}{\beta} \right) \right]^2 \right] \quad (2.4)$$

Nakagawa e Osaki (1975), a partir da função de sobrevivência descrita em (2.2), propuseram a distribuição Weibull discreta com função de probabilidade escrita na forma:

$$P(X = x | \mu, \beta) = \exp \left[- \left(\frac{x}{\mu} \right)^\beta \right] - \exp \left[- \left(\frac{x+1}{\mu} \right)^\beta \right] \quad (2.5)$$

em que $x = 0, 1, 2, \dots$, $\mu > 0$ e $\beta > 0$.

É fácil verificar que a expressão (2.5) é, de fato, uma função de probabilidade. Além disso, quando $\beta = 1$, a função de probabilidade descrita pela equação (2.5) corresponde a função de probabilidade da distribuição geométrica com parametrização $P(X = x | \beta) = (1 - e^{-\frac{1}{\beta}})e^{-\frac{1}{\beta}x}$ e quando $\beta = 2$ corresponde a função de probabilidade da distribuição Rayleigh discreta (RD) apresentada e estudada por Roy (2004).

Diversas aplicações, propriedades e métodos de estimação da distribuição Weibull discreta foram apresentadas e discutidas por vários autores como Khan, Khaliq e Abouammoh (1989), Kulasekera (1994), Roy (2002), Murthy, Xie e Jiang (2004), Englehardt e Li (2011), Almalki e Nadarajah (2014a) e Brunello e Nakano (2015).

Recentemente, Nekoukhou, Alamatsaz e Bidram (2012), introduziram uma nova distribuição discreta, com dois parâmetros, conhecida como distribuição beta-exponencial discreta. Tal distribuição foi obtida por um método de discretização diferente do método considerado por Nakagawa e Osaki (1975) e sua função de probabilidade é escrita na forma:

$$P(X = x | \alpha, \lambda) = \lambda^{x-1}(1 - \lambda^x)^{\alpha-1} \left[\sum_{j=1}^{\infty} \binom{\alpha-1}{j} \frac{(-1)^j \lambda^j}{1 - \lambda^{1+j}} \right]^{-1} \quad (2.6)$$

em que $x = 1, 2, \dots$, $\binom{\alpha-1}{j} = \frac{(\alpha-1)\dots(\alpha-j)}{j!}$, $0 < \lambda < 1$ e $\alpha > 0$. Neste caso, o método utilizado para a construção dessa distribuição foi baseado em uma série infinita que nos diz que se uma variável aleatória contínua tem função de densidade de probabilidade $f(x)$ e suporte $-\infty < x < \infty$, então a variável aleatória discreta Y correspondente, tem função de probabilidade definida por:

$$P(Y = k) = \frac{f(k)}{\sum_{j=-\infty}^{\infty} f(j)}. \quad (2.7)$$

Assim, considerando a função densidade da distribuição beta-exponencial, $f(x) = \frac{\lambda}{B(\alpha, \beta)} e^{-\beta\lambda x} (1 - e^{-\lambda x})^{\alpha-1}$, e a equação (2.7), a função de probabilidade expressa em (2.6) é facilmente obtida após algumas manipulações algébricas.

Além da distribuição Weibull discreta e Exponencial Generalizada discreta, nas últimas décadas, diversas outras distribuições, obtidas por processos de discretização, foram apresentadas e descritas na literatura. Por exemplo, a distribuição discreta de Valores Extremos foi usada por Nakagawa (1978) na análise de resistência de materiais; a distribuição Gama discreta foi usada por Yang (1994) no estudo de sequências de DNA e suas propriedades foram estudadas por Chakraborty e Chakravarty (2012); a distribuição Normal discreta foi proposta por Lisman e Zuylen (1972) e estudada posteriormente por Kemp (1997); a distribuição de Laplace discreta foi apresentada por Inusah e Kozubowski (2006) em um estudo similar ao de Kemp (1997); a distribuição de Skew-Laplace discreta foi proposta por Kozubowski e Inusah (2006); as distribuições de Maxwell, Burr e Pareto discretas foram apresentadas, respectivamente, por Krishna e Pundir (2007) e Krishna e Pundir (2009); Gómez-Déniz (2010) obteve a

distribuição Exponencial discreta Marshall-Olkin (MARSHALL; OLKIN, 1997); a distribuição Weibull inversa discreta foi proposta por Jazi, Lai e Alamatsaz (2010); a distribuição Lindley discreta foi recentemente considerada por Gómez-Déniz e Calderín-Ojeda (2011), Bakouch, Jazi e Nadarajah (2014); Nekoukhou et al. (2015) estudaram a distribuição Beta-Exponencial discreta; Hussain e Ahmad (2014) considerou a distribuição Rayleigh inversa discreta. Uma série de outras distribuições discretizadas foram apresentadas e discutidas em Chakraborty (2015).

2.2 Métodos de Discretização

Nesta seção, apresentamos os métodos de discretização utilizados no decorrer deste trabalho. Três métodos de discretização são considerados: o método de discretização via função de sobrevivência (NAKAGAWA; OSAKI, 1975) e o método de discretização baseado em uma série infinita (GOOD, 1953).

2.2.1 Discretização via Função de Sobrevivência

Este método de discretização de uma variável aleatória contínua foi proposto por Nakagawa e Osaki (1975) em estudos de confiabilidade. Eles discretizaram a distribuição Weibull por meio de sua função de sobrevivência e discutiram algumas propriedades da função de risco. Outros autores como Kemp (2004), Bracquemond e Gaudoin (2003), Roy (2003), Chakraborty (2015), estudaram este método de discretização e apresentaram algumas propriedades do método como por exemplo a preservação da função de sobrevivência, função de risco, entre outras.

Seguindo a notação de Kemp (2004), podemos definir uma variável aleatória discreta a partir de uma variável aleatória contínua conforme descrito na Definição 1 abaixo.

Definição 1: *Seja X uma variável aleatória contínua. Se X tem função de sobrevivência $S_X(x) = P[X \geq x]$, então a variável aleatória $Y = \lfloor X \rfloor$, em que $\lfloor X \rfloor$ indica a parte inteira menor ou igual a X , tem função de probabilidade definida por:*

$$\begin{aligned}
 P(Y = k) &= P(k < Y < k + 1) \\
 &= P(Y \geq k) - P(Y \geq k + 1) \\
 &= S_X(k) - S_X(k + 1) \\
 &= \sum_{i=0}^1 (-1)^i S_X(k + i), \quad k \in \mathbb{Z}_+. \tag{2.8}
 \end{aligned}$$

É fácil de se verificar que a expressão definida em (2.8) satisfaz as condições de uma função de probabilidade para $k \in \mathbb{Z}_+$. De fato, sabendo que $0 \leq S_X(k) \leq 1$ para $k \in \mathbb{Z}_+$, obtém-se que:

$$I. 0 \leq P(Y = k) \leq 1, k \in \mathbb{Z}_+;$$

$$II. \sum_{j=0}^{\infty} P(Y = j) = S_X(0) - S_X(1) + S_X(1) - S_X(2) + \dots + S_X(j-1) - S_X(j) + \dots = 1$$

isto é, por (I) e (II), a expressão definida em (2.8) caracteriza uma função de probabilidade para $k \in \mathbb{Z}_+$.

De acordo com Chakraborty (2015), se a função de sobrevivência da variável aleatória X possuir forma fechada, então a função de probabilidade construída em (2.8) também terá forma fechada. Em adição, conforme Kemp (2004) e Chakraborty (2015), a função de sobrevivência preserva sua forma para o caso discreto, isto é, $S_Y(k) = S_X(k), k \in \mathbb{Z}_+$. Uma vez que a função de sobrevivência mantém sua forma no caso discreto, a função de distribuição acumulada definida como $F_Y(k) = 1 - S_Y(k)$ também mantém sua forma para o caso discreto, ou seja, $F_Y(k) = F_X(k), k \in \mathbb{Z}_+$.

Por outro lado, usando a propriedade de que a função de risco pode ser escrita como a razão entre a função de probabilidade e a função de sobrevivência, tem-se que a função de risco da variável aleatória Y com função de probabilidade dada pela equação (2.8) é definida como:

$$h_Y(k) = \frac{S_X(k) - S_X(k+1)}{S_X(k)} = 1 - \frac{S_X(k+1)}{S_X(k)}. \quad (2.9)$$

Alguns exemplos de distribuições discretizadas por meio deste método são: a distribuição Rayleigh inversa discreta (HUSSAIN; AHMAD, 2014), a distribuição Lindley discreta (GÓMEZ-DÉNIZ; CALDERÍN-OJEDA, 2011; BAKOUCH; JAZI; NADARAJAH, 2014), a distribuição Exponencial Generalizada discreta Tipo II (NEKOUKHOUS; ALAMATSAZ; BIDRAM, 2013), a distribuição Gama discreta (CHAKRABORTY; CHAKRAVARTY, 2012), a distribuição Weibull inversa discreta (JAZI; LAI; ALAMATSAZ, 2010), as distribuições Burr e Pareto discretas (KRISHNA; PUNDIR, 2009), a distribuição Rayleigh discreta (ROY, 2004), a distribuição Weibull Geométrica discreta (BRACQUEMOND; GAUDOIN, 2003), a distribuição Weibull modificada discreta (ALMALKI; NADARAJAH, 2014b), entre outras.

2.2.2 Discretização via uma Série Infinita

Este método teve seus primeiros traços na literatura com Good (1953) em um estudo sobre modelagem da frequência populacional de espécies de animais. Recentemente, Chakraborty (2015) mostrou que a distribuição utilizada por Good (1953) tem função de probabilidade escrita na forma:

$$P(Y = k) = \frac{q^k k^\alpha}{\sum_{j=1}^{\infty} q^j j^\alpha} \quad (2.10)$$

em que $0 < q < 1$, $\alpha \in \mathbb{R}$ e $k \in \mathbb{Z}_+$. E também que, quando $\alpha > -1$, a expressão (2.10) corresponde a distribuição Gama discreta considerando a densidade da distribuição Gama contínua como $f(x) = \frac{1}{\theta^\beta \Gamma(\beta)} x^{\beta-1} e^{-x/\theta}$ e fazendo $e^{-1/\theta} = q$ e $\beta - 1 = \alpha$.

Décadas após Good (1953) ter introduzido tal distribuição, autores como Kulasekera e Tonkyn (1992), Doray e Luong (1997), Kemp (1997), Sato et al. (1999) voltaram a investigá-la e concluíram que a construção da mesma pode ser utilizada como uma alternativa para a discretização de uma variável aleatória contínua.

Em outras palavras, esses autores, apresentaram essa alternativa como sendo o método de discretização baseado em uma série infinita para os casos em que o suporte da variável aleatória contínua é definido em $(-\infty, \infty)$ ou em $(0, \infty)$. O método é descrito na Definição 2 abaixo.

Definição 2: *Seja X uma variável aleatória contínua. Se X tem função de densidade de probabilidade $f(x)$ e suporte $-\infty < x < \infty$, então a variável aleatória discreta Y correspondente, tem função de probabilidade definida por:*

$$P(Y = k) = \frac{f(k)}{\sum_{j=-\infty}^{\infty} f(j)}. \quad (2.11)$$

No caso em que o suporte de X é definido em $(0, \infty)$, de acordo com Sato et al. (1999), a função de probabilidade de Y é definida como:

$$P(Y = k) = \frac{f(k)}{\sum_{j=0}^{\infty} f(j)}. \quad (2.12)$$

Facilmente verifica-se que as equações (2.11) e (2.12) são, de fato, funções de probabilidade, respectivamente, para $k \in \mathbb{Z}$ e $k \in \mathbb{Z}_+$. De fato, sendo $0 < M = \frac{1}{\sum_{j=-\infty}^{\infty} f(j)} \leq 1$, obtém-se, no caso em que o suporte é definido em $(-\infty, \infty)$, que:

- I. $0 \leq P(Y = k) \leq 1, k \in \mathbb{Z}$;
- II. $\sum_{k=0}^{\infty} P(Y = k) = M \cdot \sum_{k=0}^{\infty} f(k) = 1$

isto é, por (I) e (II), a expressão definida em (2.11) caracteriza uma função de probabilidade para $k \in \mathbb{Z}$. Para o caso em que o suporte é definido em $(0, \infty)$, o resultado é análogo.

Diferente do método proposto por Nakagawa e Osaki (1975), a discretização via uma série infinita não preserva a forma da função de sobrevivência na parte discreta o que torna este método menos vantajoso nesse aspecto. A forma da função de sobrevivência, neste caso, pode ser escrita, respectivamente, para $k \in \mathbb{Z}$ e $k \in \mathbb{Z}_+$, como:

$$S_Y(k) = \sum_{i=k}^{\infty} \sum_{j=-\infty}^{\infty} [f(j)]^{-1} f(i) \quad (2.13)$$

ou,

$$S_Y(k) = \sum_{i=k}^{\infty} \sum_{j=0}^{\infty} [f(j)]^{-1} f(i) \quad (2.14)$$

Assim, partindo de (2.11) e (2.13) ou (2.12) e (2.14), a função de risco da variável aleatória Y é escrita na forma:

$$h_Y(k) = \frac{f(k)}{\sum_{j=-\infty}^{\infty} f(j) S_Y(k)} \quad (2.15)$$

para $k \in \mathbb{Z}$. No caso em que o suporte é definido em $k \in \mathbb{Z}_+$, o cálculo de $h_Y(k)$ é análogo.

Alguns exemplos de distribuições discretizadas por meio deste método são: a distribuição Pearson III (HAIGHT, 1957), a distribuição de Séries de Dirichlet (SIROMONEY, 1964), a distribuição Normal (KEMP, 1997), as distribuições Exponencial e Gama (SATO et al., 1999) a distribuição Log-Normal (BI; FALOUTSOS; KORN, 2001), a distribuição de Laplace (INUSAH; KOZUBOWSKI, 2006), a distribuição Skew-Laplace (KOZUBOWSKI; INUSAH, 2006), a distribuição Half-Normal (KEMP, 2008), a distribuição Beta-Exponencial discreta (NEKOUKHOU et al., 2015), entre outras.

2.2.3 Discretização via Função de Risco

O primeiro traço desse método surgiu na literatura em 1984, no contexto de análise de sobrevivência, com uma nova discretização da distribuição Weibull proposta por Stein e Dattero (1984) com a finalidade de comparar com a distribuição Weibull discreta proposta por Nakagawa e Osaki (1975).

Nesse estudo, Stein e Dattero (1984), partindo da função de risco da distribuição Weibull contínua, apresentaram uma nova versão discreta da distribuição Weibull com função de probabilidade escrita na forma:

$$P(Y = k | c, \beta) = ck^{\beta-1} \prod_{i=1}^{k-1} (1 - ci^{\beta-1}) \quad (2.16)$$

em que $k = 1, 2, \dots, m$, $\beta > 0$ e $0 < c \leq m$. O valor m é a solução de $0 \leq h_X(x) < 1$, em que $h_X(x)$ é a função de risco da distribuição Weibull contínua, isto é, m é dado por:

$$m = \begin{cases} c^{-\lfloor 1/(\beta-1) \rfloor}, & \text{se } \beta > 1 \\ +\infty, & \text{se } \beta \leq 1. \end{cases} \quad (2.17)$$

Conforme Stein e Dattero (1984), essa nova discretização da distribuição Weibull fornece o tempo de vida exato de um sistema específico de uma série de componentes e, quando adequadamente dimensionada, converge para a Weibull contínua conforme o número

de componentes aumenta. E, neste caso, a forma da função de risco é bastante similar ao modelo contínuo.

Seguindo a notação de Stein e Dattero (1984), podemos definir uma variável aleatória discreta a partir de uma variável aleatória contínua conforme descrito na Definição 3 abaixo.

Definição 3: *Seja a variável aleatória contínua X com função de risco definida como $h_X(x) = \frac{f_X(x)}{S_X(x)}$, então a função de sobrevivência da variável aleatória discreta Y correspondente é dada por:*

$$S_Y(k) = \left(\prod_{i=1}^{k-1} [1 - h_X(i)] \right), k = 1, \dots, m \quad (2.18)$$

Logo, a função de probabilidade da variável Y é dada por:

$$P(Y = k) = \begin{cases} h_X(0), & \text{se } k = 0 \\ S_Y(k)h_X(k), & \text{se } k = 1, \dots, m \\ 0, & \text{caso contrário} \end{cases} \quad (2.19)$$

em que m é solução de $0 \leq h_X(x) < 1$.

Segundo Roy e Ghosh (2009), o alcance do suporte da variável aleatória discreta Y é o valor de m determinado de modo a satisfazer a condição $0 \leq h_X(x) < 1$ e multiplicar cada $P(Y = k)$ por uma constante normalizadora que garanta que a probabilidade total seja 1.

Partindo das condições supracitadas, é imediato verificar que $P(Y = k)$ satisfaz as seguintes propriedades:

- I. $0 \leq P(Y = k) \leq 1, k \in \mathbb{Z}$;
- II. $\sum_{k=0}^{\infty} P(Y = k) = 1$

e, portanto, a expressão definida em (2.19) é uma função de probabilidade.

Diferente dos dois métodos anteriores, este método, independente do valor de m , mantém apenas a forma funcional da função de risco, isto é, $h_Y(k) = h_X(k), k = 1, \dots, m$. A função de distribuição acumulada, neste caso, é escrita na forma:

$$F_Y(k) = 1 - \left(\prod_{i=1}^{k-1} [1 - h_X(i)] \right), k = 1, \dots, m \quad (2.20)$$

Alguns exemplos de distribuições discretizadas por meio deste método são: a distribuição Weibull proposta por (STEIN; DATTERO, 1984) e (PADGETT; SPURRIER, 1985), a distribuição Lomax (ROY; GHOSH, 2009), a distribuição Weibull Inversa (LAI, 2013), a distribuição Weibull Tipo III (BARBIERO, 2013), a distribuição Weibull Modificada, entre outras.

DISCRETIZAÇÃO DA DISTRIBUIÇÃO LINDLEY E ALGUMAS DE SUAS VARIACIONES

3.1 Introdução

A distribuição Lindley foi introduzida por Lindley (1958) no contexto da inferência Bayesiana. Por muitos anos, essa distribuição foi utilizada apenas em processos de composição junto à distribuição de Poisson (SANKARAN, 1970). Tal distribuição possui função densidade de probabilidade definida por:

$$f(x | \beta) = \frac{\beta^2}{1 + \beta} (1 + x) e^{-\beta x}, \quad (3.1)$$

em que $x > 0$ e $\beta > 0$ é o parâmetro de forma. A função de sobrevivência e de risco são definidas, respectivamente, por:

$$S(x | \beta) = \left(1 + \frac{\beta x}{1 + \beta}\right) e^{-\beta x}, \quad (3.2)$$

e,

$$h(x | \beta) = \frac{\beta^2(1 + x)}{\beta(1 + x) + 1}. \quad (3.3)$$

Essa distribuição e suas propriedades matemáticas foram estudadas, no contexto de análise de sobrevivência, por Ghitany, Al-Mutairi e Nadarajah (2008). Nesse estudo, eles mostraram através da aplicação em conjunto de dados reais, que a distribuição Lindley apresenta um melhor ajuste em relação a distribuição Exponencial. Um outro estudo de aplicação da distribuição Lindley, no campo de análise riscos competitivos, foi feito por Mazucheli e Achcar (2011).

Nas últimas décadas, diversas generalizações da distribuição Lindley vêm sendo propostas na literatura e ilustrando uma maior flexibilidade em termos da função densidade de probabilidade em comparação com os modelos tradicionais.

Dentre essas generalizações, destacamos a distribuição Lindley potência com dois parâmetros, proposta por Ghitany et al. (2013), cuja função densidade apresenta os comportamentos: decrescente, unimodal ou decrescente-crescente-decrescente. Em uma aplicação considerando dados de resistência a tração de fibras de carbono sob tensão, essa distribuição mostrou uma aplicabilidade superior as distribuições Gompertz, gama e Weibull.

Ghitany et al. (2011) propôs a distribuição Lindley ponderada com dois parâmetros para a modelagem de dados de sobrevivência. Assim como a Lindley potência, a distribuição Lindley ponderada também possui três comportamentos para a função densidade de probabilidade: decrescente, unimodal ou decrescente-crescente-decrescente. A aplicabilidade dessa distribuição, considerando os tempos de vidas de porcos da guinea infectados com o vírus da tuberculose, se mostrou superior as distribuições Weibull, exponencial potência e Lindley.

Além dessas distribuições supracitadas, diversas outras generalizações vêm sendo propostas tais como: a distribuição Quasi-Lindley proposta por Shanker e Mishra (2013), a distribuição Lindley inversa proposta por Sharma et al. (2015), a distribuição Lindley transmutada proposta por Merovci (2013) e a distribuição Lindley inversa potência proposta por Barco, Mazucheli e Janeiro (2016).

3.2 Discretizações da Distribuição Lindley

3.2.1 Distribuição Lindley Discreta (via Função de Sobrevivência)

Segundo Bakouch, Jazi e Nadarajah (2014), os modelos de probabilidade tradicionais como o de Poisson e o Geométrico, tem aplicações limitadas quando se trata de tempo de falha, confiabilidade, ou até mesmo contagem. Por exemplo, no caso do modelo geométrico, a função de risco possui comportamento constante e, no caso do modelo de Poisson, a função de risco tem comportamento crescente não-dependente do parâmetro do modelo. Essas limitações fez com que modelos discretos baseados em modelos contínuos para tais aplicações fossem desenvolvidos.

Dentre esses modelos, Gómez-Déniz e Calderín-Ojeda (2011), Bakouch, Jazi e Nadarajah (2014), usando o método de discretização proposto por Nakagawa e Osaki (1975), apresentaram uma versão discreta da distribuição Lindley (LD) com função de probabilidade dada por:

$$P(X = x | \beta) = \frac{p^x}{1 + \beta} [\beta(1 - 2p) + (1 - p)(1 + \beta x)], \quad (3.4)$$

em que $x = 0, 1, \dots$, $p = e^{-\beta}$ e $\beta > 0$.

Eles que esta nova distribuição discreta tem uma aplicabilidade melhor do que os modelos clássicos quando se trata de uma análise de dados de sobrevivência e confiabilidade e até mesmo dados de contagem. Isso faz com esse modelo seja mais versátil e menos limitado que os modelos de probabilidade tradicionais usados na análise de dados discretos.

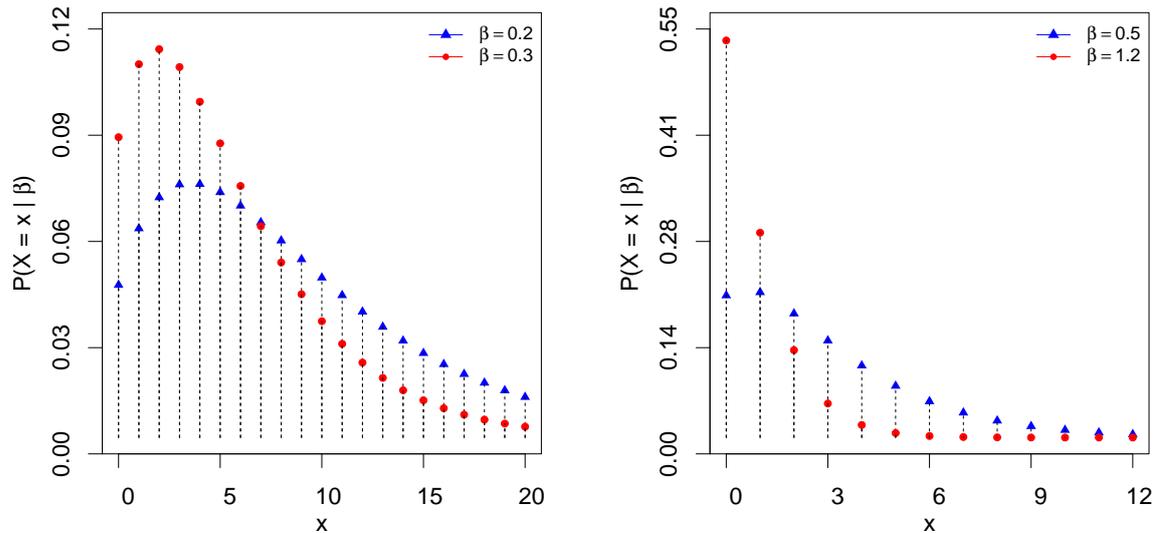


Figura 1 – Comportamento da função de probabilidade da distribuição Lindley discreta (via função de sobrevivência) para alguns valores de β .

O comportamento de (3.4) é ilustrado na Figura 1. Pode-se notar que a moda é centrada no 0 quando $\beta > 1$. Além disso, conforme Bakouch, Jazi e Nadarajah (2014), a moda pode ser expressa como:

$$x_0 : \begin{cases} x_0 = \left\lfloor \frac{2p}{1-p} - \frac{1}{\beta} \right\rfloor, & \text{se } \frac{2p}{1-p} - \frac{1}{\beta} \notin \mathbb{Z}, \\ \frac{2p}{1-p} - \frac{1}{\beta} < x_0 < \frac{2p}{1-p} - \frac{1}{\beta} - 1, & \text{se } \frac{2p}{1-p} - \frac{1}{\beta} \in \mathbb{Z} \end{cases} \quad (3.5)$$

em que $\lfloor \cdot \rfloor$ é a função menor inteiro.

A função de distribuição acumulada, a função de sobrevivência e a função de risco são, respectivamente, dadas por:

$$F(x | \beta) = 1 - \frac{1 + \beta + \beta x}{1 + \beta} p^x, \quad (3.6)$$

$$S(x | \beta) = \frac{1 + \beta + \beta x}{1 + \beta} p^x \quad (3.7)$$

e,

$$h(x | \beta) = 1 - p - \frac{p\beta}{1 + \beta + \beta x}. \quad (3.8)$$

Note, na equação (3.8), que $h(0 | \beta) = \frac{1}{1 + \beta} [(1 + \beta)(1 - p) - p\beta]$ e $h(\infty | \beta) = 1 - p$. Se β for pequeno, $h(\infty | \beta)$ é aproximadamente igual a β o que implica que $h(x | \beta)$ é limitada por β para todo x .

O comportamento de (3.8) é ilustrado na Figura 2. Pode-se notar que a função de risco dessa distribuição possui comportamento crescente para todo $\beta > 0$.

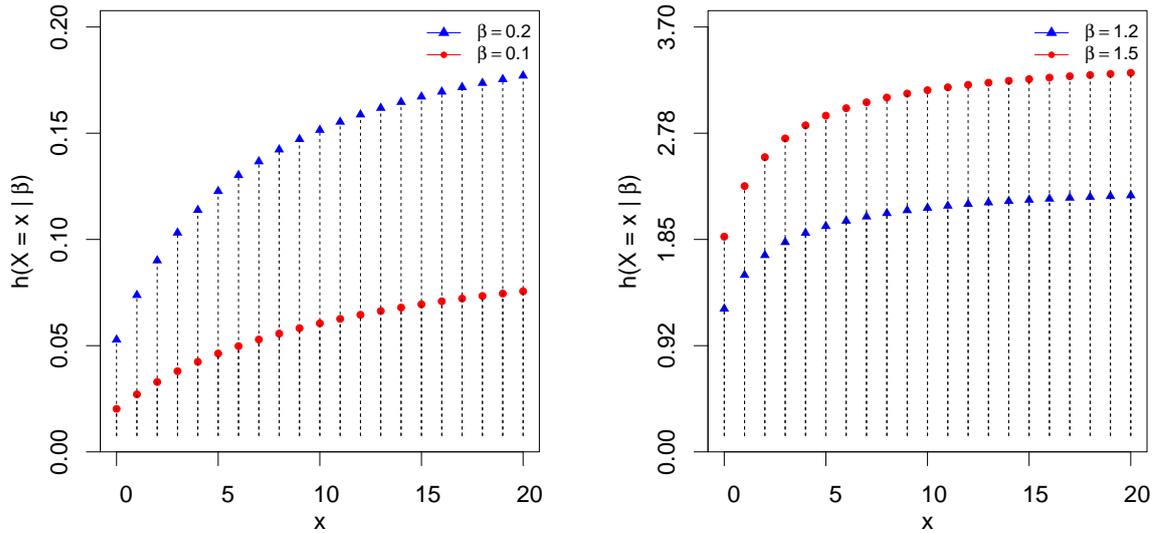


Figura 2 – Comportamento da função de risco da distribuição Lindley discreta (via função de sobrevivência) para alguns valores de β .

Partindo de (3.4), a função geradora de momentos de uma variável aleatória discreta X com distribuição Lindley discreta pode ser escrita na forma:

$$M(t) = \frac{\beta(1 - 2p) + 1 - p}{(1 + \beta)(1 - e^t p)} + \frac{\beta p(1 - 2p)e^t}{(1 + \beta)(1 - e^t p)^2} \quad (3.9)$$

Consequentemente, de (3.9), obtém-se que a esperança e a variância da variável X são dadas, respectivamente, por:

$$\mathbb{E}(X) = \frac{p(2\beta - \beta p + 1 - p)}{(1 + \beta)(1 - p)^2} \quad (3.10)$$

e,

$$\text{Var}(X) = \frac{p[(-3\beta^2 - 4\beta - 2)p + (2\beta + p^2 + 1)(\beta + 1)]}{(1 + \beta)^2(1 - p)^4}. \quad (3.11)$$

Neste caso, de (3.10) e (3.11), é fácil verificar que $\mathbb{E}(X) < \text{Var}(X)$ para todo $\beta > 0$. Portanto, a distribuição Lindley discreta pode apenas ser usada para modelar dados com superdispersão (Figura 3).

Além das propriedades supracitadas, Bakouch, Jazi e Nadarajah (2014) também apresentaram a função quartil dessa distribuição como sendo a parte inteira da solução da equação $F(Q(u)) = u$. A função quantil, então, pode ser escrita na forma:

$$Q(u) = \left\lfloor -1 - \frac{1}{\beta} - \frac{1}{\beta} W(-(1 + \beta) \exp(-1 - \beta)(1 - u)) \right\rfloor, \quad (3.12)$$

em que W representa a função W de Lambert com ramo positivo (JODRA, 2010) e $0 < u < 1$.

Além das propriedades matemáticas, uma outra característica importante quando se trabalha com modelos de probabilidade é a estimação de seus parâmetros afim de realizar

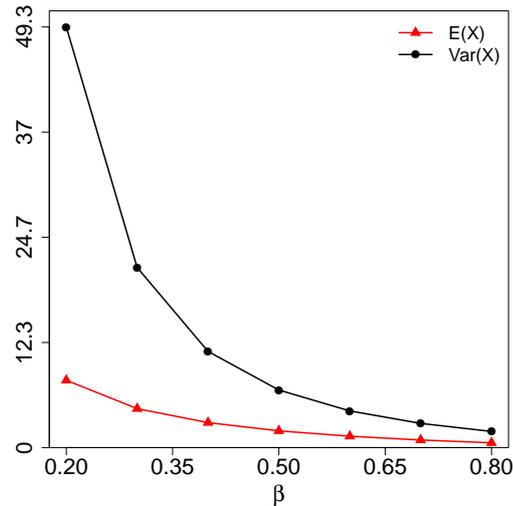


Figura 3 – Esperança e variância distribuição Lindley discreta (via função de sobrevivência) para alguns valores de β .

inferência sobre população. Neste caso, suponha uma amostra aleatória x_1, \dots, x_n iid da distribuição Lindley discreta com função de probabilidade dada por (3.4). O logaritmo da função de verossimilhança pode ser escrito como:

$$\begin{aligned} \ell(\beta | \mathbf{x}) &= -\beta \sum_{i=1}^n x_i - n \log(1 + \beta) \\ &+ \sum_{i=1}^n \log [1 + (1 + x_i)\beta - (1 + 2\beta + \beta x_i) \exp(-\beta)]. \end{aligned}$$

É importante destacar que o parâmetro β é o parâmetro de dispersão e seu estimador de máxima verossimilhança é obtido resolvendo-se $\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = 0$ em β em que:

$$\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = -n\bar{x} - \frac{n}{1 + \beta} + \sum_{i=1}^n \frac{1 + x_i - (1 - 2\beta + x_i + \beta x_i) \exp(-\beta)}{1 + (1 + x_i)\beta - (1 + 2\beta + \beta x_i) \exp(-\beta)}. \quad (3.13)$$

Note que esta expressão é não linear em β e deve ser resolvida numericamente. Entretanto $\hat{\beta} \approx -0.5 (1 - \sqrt{1 + 4\bar{x}})$ se $\exp(-\beta) \approx 1$. Exceto por essa aproximação, o estimador de máxima verossimilhança de β não possui forma explícita.

Com objetivo de avaliar o vício e o erro-quadrático-médio de $\hat{\beta}$, foi conduzido um estudo de simulação Monte Carlo considerando $\beta = 0.2, 0.5, 0.8, 1.0$ e 1.2 com tamanhos de amostra $n = 10, 20, \dots, 100$. Para cada cenário, foi calculado:

$$\text{Vício}(\hat{\theta}) = \frac{1}{B} \sum_{i=1}^B (\hat{\theta}_i - \theta_i) \quad \text{e} \quad \text{EQM}(\hat{\theta}) = \frac{1}{B} \sum_{i=1}^B (\hat{\theta}_i - \theta_i)^2$$

em que $\theta = \beta$ e $B = 10.000$ replicações.

Os resultados desse estudo estão dispostos na Tabela 1. De acordo com a Tabela 1 e a Figura 4, pode-se notar que tanto o vício quanto o EQM de $\hat{\beta}$ apresentam tendência de

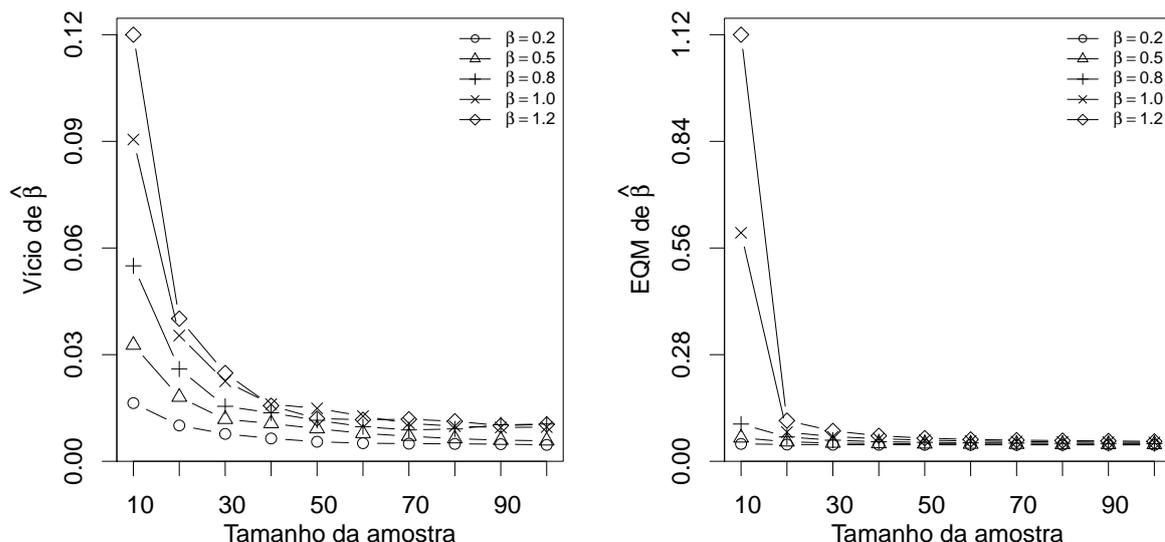


Figura 4 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta (via função de sobrevivência).

decaimento exponencial e se aproximam de 0 quando o tamanho da amostra aumenta. O vício é positivo tendo o maior valor quando $\beta = 1.2$ e menor valor quando $\beta = 0.2$. O mesmo ocorre com o EQM nos cenários considerados.

Os intervalos de confiança para $\hat{\beta}$ assim como testes de hipóteses podem ser contruídos a partir da normalidade assintóticas dos estimadores de máxima verossimilhança. Os métodos Bayesianos podem ser usados com o mesmo propósito. Para um estudo mais detalhado a respeito da inferência dessa distribuição, consultar Bakouch, Jazi e Nadarajah (2014).

Tabela 1 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta (via função de sobrevivência).

	n	Valores de β						n	Valores de β				
		0.2	0.5	0.8	1.0	1.2			0.2	0.5	0.8	1.0	1.2
Vício	10	0.0137	0.0318	0.0540	0.0910	0.1218	EQM	10	0.0028	0.0190	0.0569	0.5768	1.1153
	20	0.0072	0.0151	0.0237	0.0335	0.0385		20	0.0012	0.0083	0.0218	0.0350	0.0662
	30	0.0047	0.0081	0.0128	0.0201	0.0225		30	0.0007	0.0050	0.0130	0.0226	0.0382
	40	0.0035	0.0072	0.0108	0.0134	0.0129		40	0.0006	0.0037	0.0090	0.0172	0.0251
	50	0.0025	0.0059	0.0086	0.0122	0.0092		50	0.0004	0.0030	0.0071	0.0129	0.0186
	60	0.0020	0.0046	0.0068	0.0099	0.0088		60	0.0004	0.0025	0.0058	0.0108	0.0150
	70	0.0016	0.0038	0.0058	0.0077	0.0090		70	0.0003	0.0021	0.0051	0.0092	0.0131
	80	0.0016	0.0030	0.0061	0.0069	0.0084		80	0.0003	0.0018	0.0045	0.0083	0.0119
	90	0.0015	0.0027	0.0073	0.0065	0.0071		90	0.0002	0.0016	0.0040	0.0075	0.0110
	100	0.0013	0.0025	0.0076	0.0067	0.0075		100	0.0002	0.0014	0.0037	0.0068	0.0096

3.2.2 Distribuição Lindley Discreta (via Série Infinita)

Por meio do método de discretização baseado em uma série infinita apresentado no capítulo anterior, propomos a distribuição Lindley discreta (LDS) com função de probabilidade escrita na forma:

$$P(X = x | \beta) = (1 - p)^2(1 + x)p^x \tag{3.14}$$

em que $x = 0, 1, \dots$ e $p = e^{-\beta}$. Comparada a versão apresentada por Bakouch, Jazi e Nadarajah (2014), essa versão é mais flexível em termos da função de probabilidade.

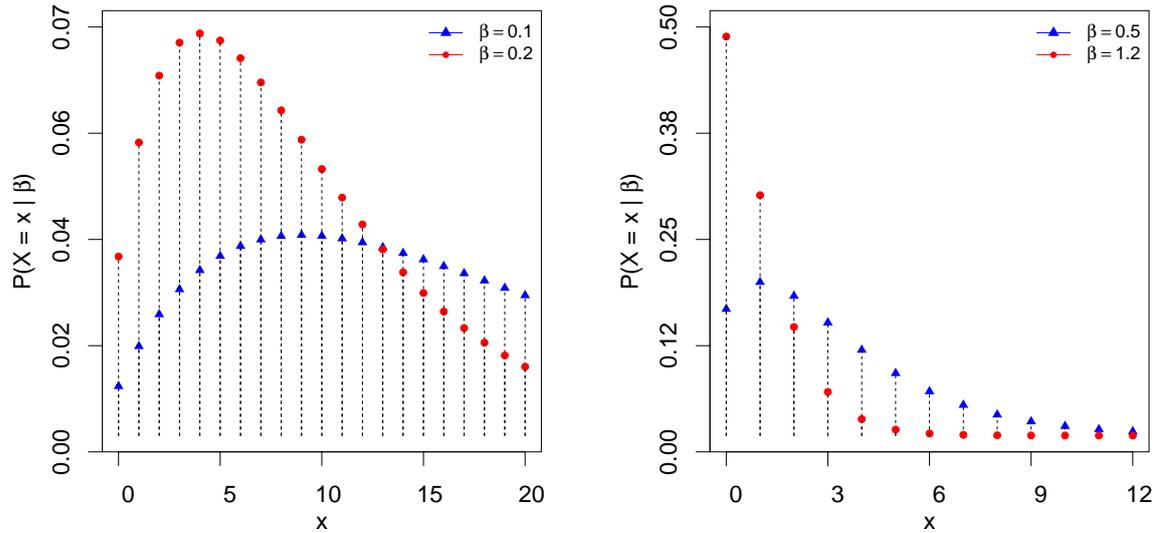


Figura 5 – Comportamento da função de probabilidade da distribuição Lindley discreta (via série infinita) para alguns valores de β .

O comportamento de (3.14) é ilustrado na Figura 5. Pode-se observar que a moda é centrada no 0 se $\beta > 1$. Além disso, a moda é definida como:

$$x_0 : \begin{cases} x_0 = \left\lfloor \frac{p}{1-p} \right\rfloor, & \text{se } \frac{p}{1-p} \notin \mathbb{Z}, \\ \frac{p}{1-p} < x_0 < \frac{p}{1-p} - 1, & \text{se } \frac{p}{1-p} \in \mathbb{Z} \end{cases} \quad (3.15)$$

É importante destacar que a distribuição Lindley discreta aqui proposta tem comportamento unimodal. De fato, note que:

$$\begin{aligned} [P(X = x)]^2 &= (1-p)^4(1+x)^2p^{2x} \\ &\geq x(x+2)p^{2x}(1-p)^4 \end{aligned}$$

Note que, o lado direito da desigualdade acima, é o mesmo que $P(X = x-1)P(X = x+1)$. Assim, a equação (3.14) satisfaz a inequação de log-concavidade $P^2(X = x) \geq P(X = x-1)P(X = x+1)$ para $x = 0, 1, 2, \dots$ e, portanto, pelo Teorema 3 de Keilson e Gerber (1971), é unimodal.

A função de distribuição acumulada, a função de sobrevivência e a função de risco são, respectivamente, dadas por:

$$F(x | \beta) = 1 - [1 + (1-p)x]p^x \quad (3.16)$$

$$S(x | \beta) = [1 + (1-p)x]p^x \quad (3.17)$$

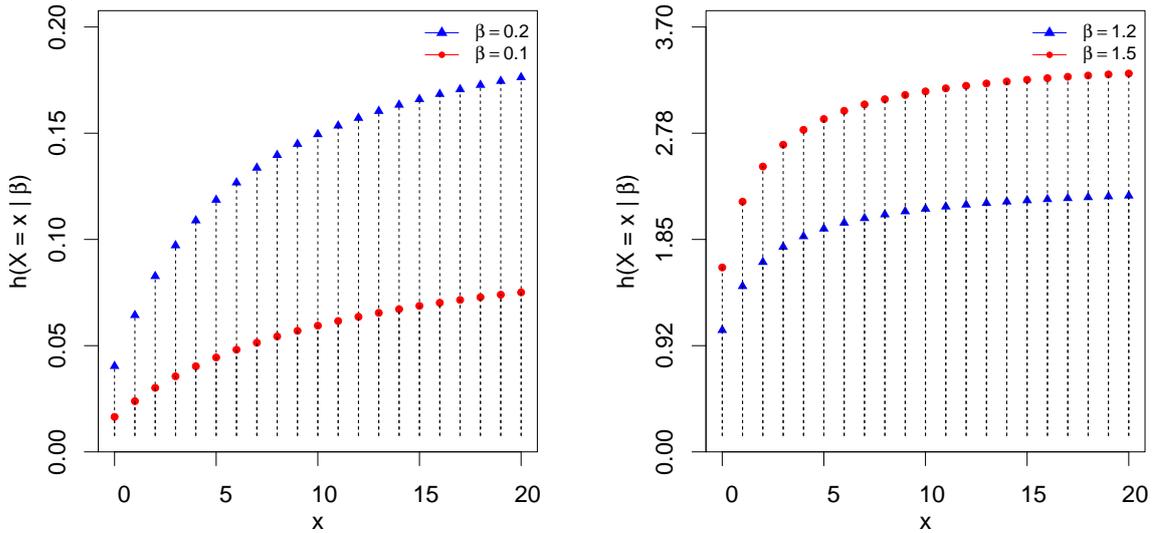


Figura 6 – Comportamento da função de risco da distribuição Lindley discreta (via série infinita) para alguns valores de β .

e,

$$h(x | \beta) = \frac{(1 - p)^2(1 + x)}{1 + x - px}. \quad (3.18)$$

Note, na equação (3.18), que $h(0 | \beta) = (1 - p)^2$ e $h(\infty | \beta) = 1 - p$. Se β for pequeno, $h(\infty | \beta)$ é aproximadamente igual a β o que implica que $h(x | \beta)$ é limitada por β para todo x .

O comportamento de (3.18) é ilustrado na Figura 6. Pode-se notar que a função de risco dessa distribuição possui comportamento crescente para todo $\beta > 0$.

De (3.14), a função geradora de momentos de uma variável aleatória discreta X que segue uma distribuição Lindley discreta (via série infinita) é escrita na forma:

$$M_X(t) = \frac{(1 - p)^2}{(pe^t - 1)^2}. \quad (3.19)$$

Logo, a partir de (3.19), a esperança e variância da variável X são definidas, respectivamente, por:

$$\mathbb{E}(X) = \frac{2p}{1 - p} \quad (3.20)$$

e,

$$\text{Var}(X) = \frac{2p}{(1 - p)^2}. \quad (3.21)$$

De (3.20) e (3.21) é fácil verificar que $\mathbb{E}(X) < \text{Var}(X)$ para todo $\beta > 0$. Da mesma maneira que a discretização feita por Bakouch, Jazi e Nadarajah (2014), essa discretização da distribuição Lindley pode ser apenas utilizada para modelar dados com superdispersão já que $\mathbb{E}(X) < \text{Var}(X), \forall \beta > 0$ (Figura 7).

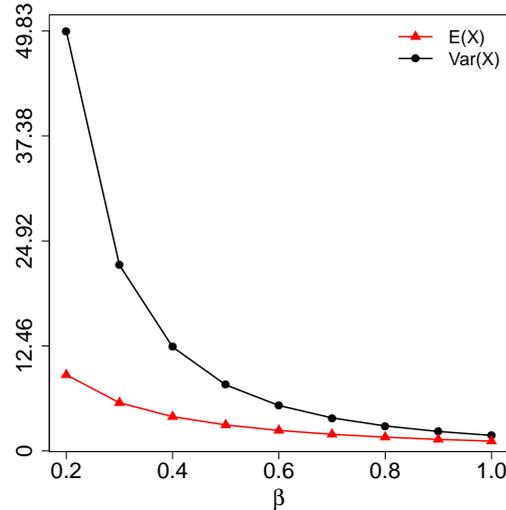


Figura 7 – Esperança e variância distribuição Lindley discreta (via série infinita) para alguns valores de β .

De maneira análoga a distribuição Lindley discreta (via função de sobrevivência), a função quartil da distribuição Lindley discreta (via série infinita) é escrita em termos da função W de Lambert e definida por:

$$Q(u) = \left\lfloor \frac{1}{p-1} + \frac{1}{\log(p)} W \left(\frac{\log(p)(u-1)e^{\frac{\log(p)}{1-p}}}{p-1} \right) \right\rfloor \quad (3.22)$$

em que W representa a função W de Lambert com ramo positivo (JODRA, 2010) e $0 < u < 1$.

Seja uma amostra aleatória x_1, \dots, x_n de tamanho n da distribuição Lindley discreta (via série infinita) com parâmetro de dispersão β e função de probabilidade dada por (3.14), o logaritmo da função de verossimilhança é proporcional a:

$$\ell(\beta | \mathbf{x}) \propto 2n \log(1-p) - \beta(2n + n\bar{x}). \quad (3.23)$$

O estimador de máxima verossimilhança do parâmetro β é obtido resolvendo-se $\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = 0$ em β em que:

$$\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = \frac{2np}{p-1} - 2n - n\bar{x}. \quad (3.24)$$

Note que a equação (3.24) é não-linear, porém sua solução é facilmente obtida sem a necessidade de métodos numéricos. Dessa forma, estimador $\hat{\beta}$ é descrito por:

$$\hat{\beta} = \log \left(1 + \frac{2}{\bar{x}} \right). \quad (3.25)$$

Para avaliar o vício e o EQM do estimador dado por (3.24), utilizamos estudos de simulação Monte Carlo adotando como cenários os tamanhos de amostra $n = 10, 20, \dots, 100$ e $\beta = 0.2, 0.5, 0.8, 1.0$ e 1.2 . Para cada cenário, calculamos o Vício($\hat{\beta}$) e o EQM($\hat{\beta}$) descritos anteriormente com 10.000 replicações.

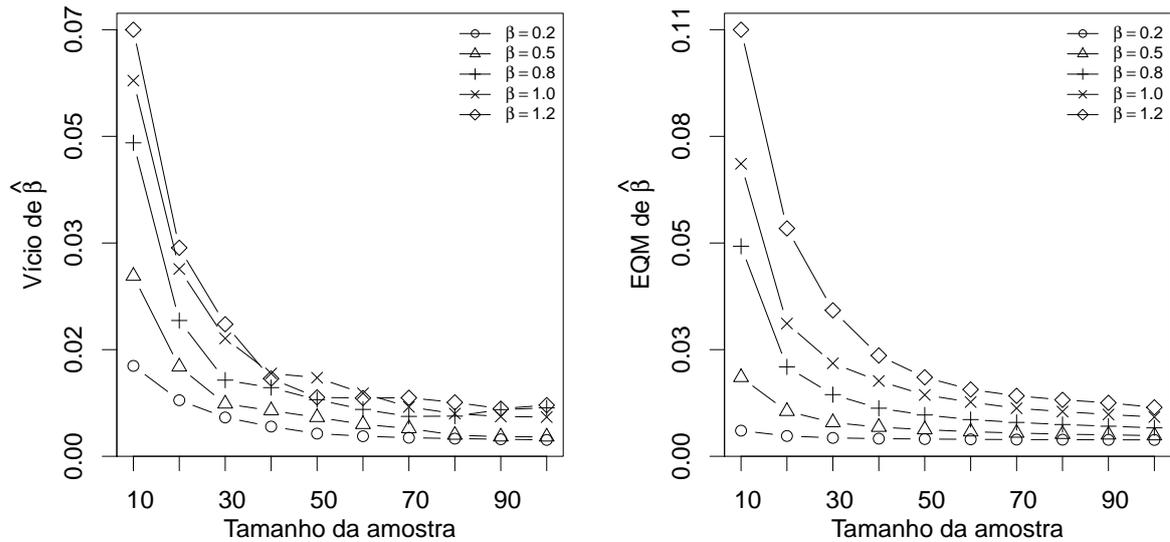


Figura 8 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta (via série infinita).

Os resultados desse estudo estão dispostos na Tabela 2. De acordo com a Tabela 2 e a Figura 8, pode-se notar que tanto o vício quanto o EQM de $\hat{\beta}$ apresentam tendência de decaimento exponencial e se aproximam de 0 quando o tamanho da amostra aumenta. O vício é positivo tendo o maior valor quando $\beta = 1.2$ e menor valor quando $\beta = 0.2$. O mesmo ocorre com o EQM nos cenários considerados.

É importante destacar que esta versão apresenta valores menores e uma tendência mais exponencial de decaimento para o EQM em comparação com a discretização apresentada por Bakouch, Jazi e Nadarajah (2014).

Quanto aos intervalos de confiança para $\hat{\beta}$, de maneira análoga as distribuições anteriores, os intervalos de confiança para $\hat{\beta}$ assim como testes de hipóteses podem ser contruídos a partir da normalidade assintóticas dos estimadores de máxima verossimilhança.

Tabela 2 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta (via série infinita).

	n	Valores de β					n	Valores de β					
		0.2	0.5	0.8	1.0			1.2	0.2	0.5	0.8	1.0	1.2
Vício	10	0.0134	0.0278	0.0499	0.0600	0.0683	EQM	10	0.0027	0.0167	0.0507	0.0722	0.1072
	20	0.0071	0.0134	0.0210	0.0294	0.0328		20	0.0012	0.0077	0.0192	0.0306	0.0554
	30	0.0046	0.0072	0.0113	0.0181	0.0204		30	0.0007	0.0047	0.0120	0.0202	0.0340
	40	0.0035	0.0064	0.0101	0.0124	0.0116		40	0.0006	0.0035	0.0085	0.0156	0.0222
	50	0.0025	0.0052	0.0082	0.0117	0.0085		50	0.0004	0.0028	0.0067	0.0119	0.0165
	60	0.0020	0.0041	0.0066	0.0092	0.0084		60	0.0004	0.0023	0.0055	0.0101	0.0134
	70	0.0016	0.0033	0.0055	0.0070	0.0085		70	0.0003	0.0020	0.0048	0.0085	0.0117
	80	0.0016	0.0026	0.0055	0.0059	0.0077		80	0.0003	0.0017	0.0042	0.0076	0.0107
	90	0.0015	0.0022	0.0066	0.0054	0.0067		90	0.0002	0.0015	0.0038	0.0068	0.0099
	100	0.0013	0.0022	0.0068	0.0054	0.0073		100	0.0002	0.0013	0.0033	0.0062	0.0087

3.2.3 Distribuição Lindley Discreta (via Função de Risco)

Por meio do método de discretização baseado na função de risco apresentado no capítulo anterior, propomos uma outra discretização da distribuição Lindley, aqui denominada

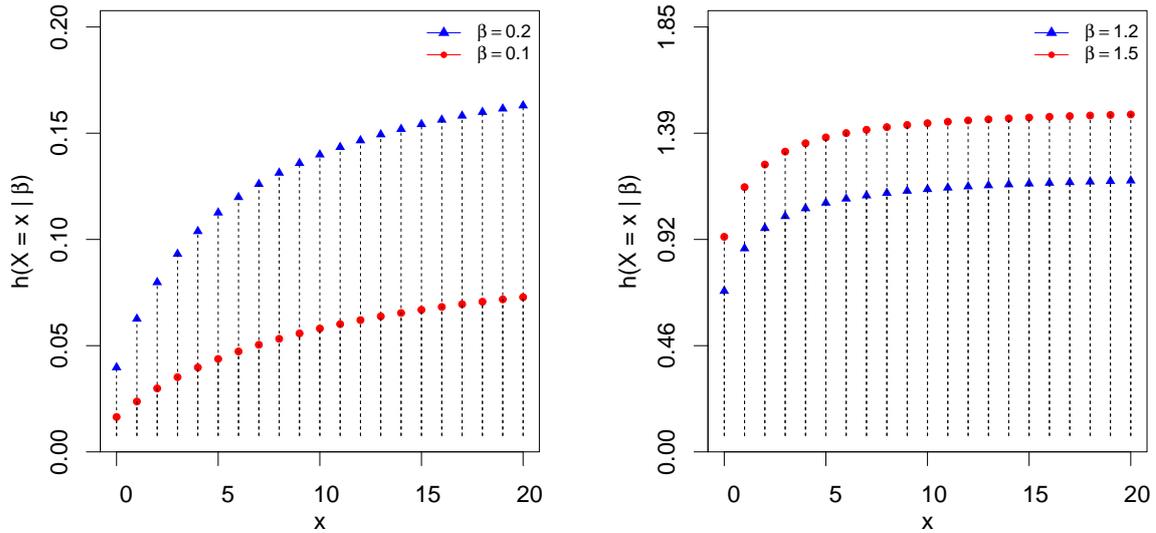


Figura 9 – Comportamento da função de risco da distribuição Lindley discreta (via função de risco) para alguns valores de β .

Lindley discreta (via função de risco) com função de probabilidade escrita na forma:

$$P(x | \beta) = \frac{\beta^2(1+x)}{\beta(x+1)+1} \prod_{i=1}^{x-1} \left(1 - \frac{\beta^2(1+i)}{\beta(i+1)+1}\right) \quad \text{se } x = 1, \dots, k \quad (3.26)$$

em que $\beta > 0$ e

$$k = \begin{cases} \left\lfloor -\frac{\beta^2 - \beta - 1}{\beta(\beta - 1)} \right\rfloor & \text{se } \beta > 1.619 \\ +\infty & \text{se } \beta \leq 1.619 \end{cases} \quad (3.27)$$

A função de distribuição acumulada, de sobrevivência e de risco para essa discretização da distribuição Lindley são descritas, respectivamente, por:

$$F(x | \beta) = 1 - \prod_{i=1}^{x-1} \frac{\beta^2(1+i)}{\beta(i+1)+1}, \quad (3.28)$$

$$S(x | \beta) = \prod_{i=1}^{x-1} \frac{\beta^2(1+i)}{\beta(i+1)+1}, \quad (3.29)$$

e,

$$h(x | \beta) = \frac{\beta^2(1+x)}{\beta(x+1)+1}. \quad (3.30)$$

em que $x = 1, \dots, k$, $\beta > 0$ e k descrito na equação (3.27). O comportamento de (3.30) é ilustrado na Figura 9 e pode-se notar, como nas distribuições anteriores, que o mesmo é crescente para todo $\beta > 0$.

Apesar desta ser uma outra discretização da distribuição Lindley, a mesma foi proposta apenas para ilustrar uma aplicação do método de discretização considerado nessa construção. Os estudos de suas propriedades matemáticas e estimação ficarão para um estudo futuro.

3.2.4 Distribuição Lindley Potência Discreta

Ghitany et al. (2013), com o objetivo de obter uma distribuição mais flexível em termos de comportamento da função densidade e de risco, introduziram a distribuição Lindley potência, que considera a potência de uma variável aleatória com distribuição Lindley. Ou seja, se uma variável aleatória Y possui distribuição Lindley, então a variável aleatória $X = Y^{\frac{1}{\alpha}}$ segue uma distribuição Lindley potência. Dessa forma, sua função densidade de probabilidade é definida por:

$$f(x | \alpha, \beta) = \frac{\alpha\beta^2}{1 + \beta}(1 + x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}, \quad (3.31)$$

em que $x > 0$ e $\alpha, \beta > 0$.

Note que, a equação acima também pode ser obtida pela mistura de duas distribuições: a distribuição Weibull, com parâmetros de escala β e forma α , e a distribuição Gama generalizada, com parâmetro de escala β e de forma $\alpha = 2$, e com probabilidade de mistura igual a $p = \frac{\beta}{\beta+1}$. A função de sobrevivência dessa distribuição é definida por:

$$S(x | \alpha, \beta) = \left(1 + \frac{\beta x^\alpha}{1 + \beta}\right) e^{-\beta x^\alpha}, \quad (3.32)$$

É importante destacar que a distribuição Lindley é um caso particular da distribuição Lindley potência, quando tomamos $\alpha = 1$. Além disso, sua função densidade têm comportamentos decrescente, unimodal ou decrescente-crescente-decrescente, dependendo dos valores tomados para α e β .

Neste trabalho, a discretização da distribuição Lindley potência foi obtida apenas pelo método de Nakagawa e Osaki (1975), pois pelos outros métodos não foi possível discretizar a mesma devidos a problemas de convergência.

Considerando a função de sobrevivência, dada por (3.32), da distribuição Lindley potência contínua, propomos a discretização da distribuição Lindley potência com função de probabilidade definida por:

$$P(X = x | \alpha, \beta) = \left(1 + \frac{\beta x^\alpha}{\beta + 1}\right) \gamma^{x^\alpha} - \left(1 + \frac{\beta(x+1)^\alpha}{\beta + 1}\right) \gamma^{(x+1)^\alpha}, \quad (3.33)$$

em que $x = 0, 1, \dots$, $\alpha, \beta > 0$ e $\gamma = \exp(-\beta)$.

Note que, assim como no caso contínuo, tem-se que a distribuição Lindley discreta, apresentada por Bakouch, Jazi e Nadarajah (2014), é um caso particular da distribuição Lindley potência discreta para $\alpha = 1$. Além disso, a equação (3.33) satisfaz a inequação de log-concavidade (KEILSON; GERBER, 1971) e, portanto, é unimodal para todo $\alpha, \beta > 0$. Na [Figura 10](#) pode-se observar tal comportamento.

A função de distribuição acumulada, de sobrevivência e de risco da distribuição Lindley potência discreta são dadas, respectivamente, por:

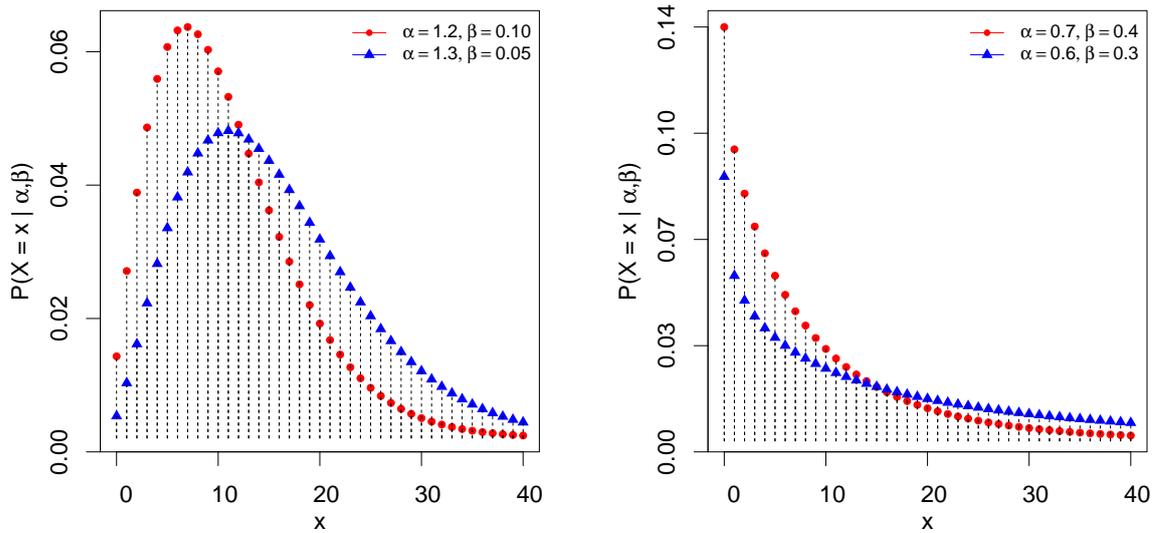


Figura 10 – Comportamento da função de probabilidade da distribuição Lindley potência discreta para diferentes valores de α e β .

$$F(x | \alpha, \beta) = 1 - \left(1 + \frac{\beta x^\alpha}{\beta + 1}\right) \gamma^{x^\alpha}, \quad (3.34)$$

$$S(x | \alpha, \beta) = \left(1 + \frac{\beta x^\alpha}{\beta + 1}\right) \gamma^{x^\alpha}, \quad (3.35)$$

e,

$$h(x | \alpha, \beta) = 1 - \left[\frac{1 + \beta + \beta(x+1)^\alpha}{1 + \beta + \beta x^\alpha} \right] \gamma^{(x+1)^\alpha - x^\alpha} \quad (3.36)$$

em que $x = 0, 1, 2, \dots$, $\alpha, \beta > 0$ e $\gamma = \exp(-\beta)$. Em relação a função de risco, para $x \rightarrow 0$ e $x \rightarrow \infty$, obtemos:

$$h(0 | \alpha, \beta) = 1 - \frac{1 + 2\beta}{1 + \beta} \gamma \quad h(\infty | \alpha, \beta) = \begin{cases} 0, & \text{if } \alpha < 1 \\ 1 - \gamma, & \text{if } \alpha = 1 \\ 1, & \text{if } \alpha > 1. \end{cases} \quad (3.37)$$

O comportamento de (3.36) é ilustrado na Figura 11. Observa-se que o mesmo é crescente para $\alpha > 1$ e decrescente para $\alpha < 1$ para um $\beta \leq 0.5$ fixo.

O momento de ordem k não tem forma fechada mas pode ser avaliado computacionalmente por:

$$\begin{aligned} \mathbb{E}(X^k) &= \sum_{x=0}^{\infty} x^k \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^k \frac{[-\beta(x+1)^\alpha]^j}{j!} \\ &+ \frac{\beta}{\beta + 1} \left[\sum_{x=0}^{\infty} x^{k+\alpha} \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^k (x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right]. \end{aligned} \quad (3.38)$$

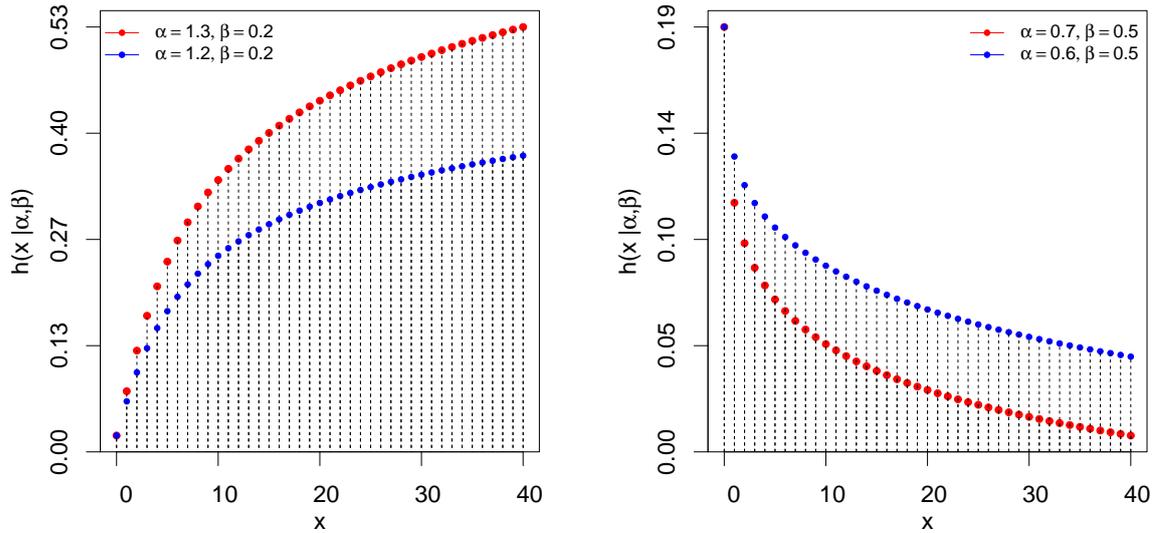


Figura 11 – Comportamento da função de risco da distribuição Lindley potência discreta para alguns valores de α e β .

Todavia, se $\frac{k}{\alpha} \in \mathbb{N}^*$ então $\mathbb{E}[X^k]$ pode ser expresso como:

$$\begin{aligned} \mathbb{E}(X^k) &= K\left(\frac{k}{\alpha}, \frac{1}{\gamma}\right) - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^k \frac{[-\beta(x+1)^\alpha]^j}{j!} \\ &+ \frac{\beta}{\beta+1} \left[K\left(\frac{k}{\alpha} + 1, \frac{1}{\gamma}\right) - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^k (x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right], \end{aligned} \quad (3.39)$$

em que:

$$K(a, b) = \frac{1}{(b-1)^{a+1}} \sum_{i=1}^a \left[\frac{1}{b^{a-i}} \sum_{j=0}^i \frac{(-1)^j (a+1)! (i-j)^a}{j! (a+1-j)!} \right] \quad a = 1, 2, \dots, b \neq 1.$$

Por meio do momento de ordem k dado pela equação (3.38), tem-se que a esperança e a variância da distribuição Lindley potência discreta são dadas, respectivamente, por:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x=0}^{\infty} x \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x \frac{[-\beta(x+1)^\alpha]^j}{j!} \\ &+ \frac{\beta}{\beta+1} \left[\sum_{x=0}^{\infty} x^{1+\alpha} \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x(x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right] \end{aligned} \quad (3.40)$$

e,

$$\begin{aligned} \text{Var}(X) &= \sum_{x=0}^{\infty} x^2 \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^2 \frac{[-\beta(x+1)^\alpha]^j}{j!} \\ &+ \frac{\beta}{\beta+1} \left[\sum_{x=0}^{\infty} x^{2+\alpha} \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^2 (x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right] \\ &- \left\{ \sum_{x=0}^{\infty} x \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x \frac{[-\beta(x+1)^\alpha]^j}{j!} \right. \\ &\left. + \frac{\beta}{\beta+1} \left[\sum_{x=0}^{\infty} x^{1+\alpha} \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x(x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right] \right\}^2 \end{aligned} \quad (3.41)$$

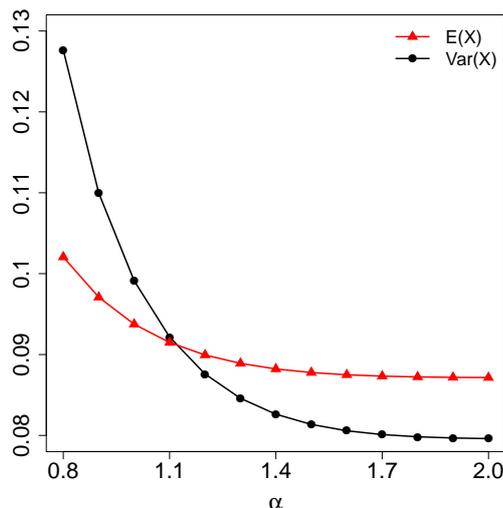


Figura 12 – Esperança e variância distribuição Lindley potência discreta para alguns valores de α e $\beta = 3$.

Neste caso, tanto a esperança quanto a variância não possuem forma fechada o que impossibilita a verificação da subdispersão e superdispersão algebricamente. Na Tabela 3 é ilustrado o índice de dispersão da distribuição Lindley potência discreta e pode-se verificar que a mesma pode ser usada tanto para dados com subdispersão quanto para dados com superdispersão (Figura 12).

Tabela 3 – Índice de dispersão da distribuição Lindley potência discreta.

$\beta \backslash \alpha$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
0.5	58.131	2.649	0.919	0.531	0.376	0.277	0.194	0.192	0.191	0.191	0.191	0.191
1.0	15.348	1.635	0.808	0.577	0.482	0.452	0.449	0.448	0.448	0.448	0.448	0.448
1.5	7.376	1.315	0.815	0.680	0.647	0.643	0.643	0.643	0.643	0.643	0.643	0.643
2.0	4.559	1.172	0.852	0.784	0.775	0.774	0.774	0.774	0.774	0.774	0.774	0.774
2.5	3.237	1.098	0.892	0.861	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859
3.0	2.510	1.057	0.926	0.913	0.912	0.912	0.912	0.912	0.912	0.912	0.912	0.912
3.5	2.066	1.033	0.952	0.946	0.946	0.946	0.946	0.946	0.946	0.946	0.946	0.946
4.0	1.777	1.020	0.969	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967
4.5	1.579	1.012	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980
5.0	1.439	1.007	0.988	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987
5.5	1.337	1.004	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992
6.0	1.261	1.002	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995

Em relação aos quantis, a função quantil para a distribuição Lindley potência discreta é definida como solução da equação $F(Q(u)) = u$. A solução dessa equação é descrita no Teorema abaixo.

Teorema 3.2.1. Para todo $\alpha, \beta > 0$, a função quantil da distribuição LPD é dada por:

$$Q(u) = \left\lceil \left[-1 - \frac{1}{\beta} - \frac{1}{\beta} W_{-1}(- (1 + \beta) e^{-(1-\beta)} (1 - u)) \right]^{\frac{1}{\alpha}} \right\rceil, \quad 0 < u < 1,$$

em que W_{-1} é o ramo inferior da função W de Lambert.

Demonstração: Seja X uma variável aleatória discreta tal que $X \sim LPD(\alpha, \beta)$. A função quantil de X , definida por $F(Q(u)) = u$ é a solução da equação:

$$1 - \left(1 + \frac{\beta[Q(u)]^\alpha}{\beta + 1}\right) e^{-\beta[Q(u)]^\alpha} = u, \quad 0 < u < 1.$$

Multiplicando ambos os lados da equação acima por $e^{-1-\beta}$, tem-se que:

$$\left(\frac{1 + \beta + \beta[Q(u)]^\alpha}{\beta + 1}\right) e^{-1-\beta-\beta[Q(u)]^\alpha} = (1 - u)e^{-1-\beta}.$$

Por outro lado, toma-se $Z(u) = -1 - \beta - \beta[Q(u)]^\alpha$, a equação acima se reduz a:

$$Z(u)e^{Z(u)} = -(1 + \beta)(1 - u)e^{(-1-\beta)}, \quad 0 < u < 1.$$

Para determinar-se a solução dessa equação necessita-se utilizar a função W de Lambert, apresentada por Jodra (2010), que trata-se de uma função complexa multivalorada definida como a solução da equação $W(z)e^{W(z)} = z$, em que z é um número complexo.

Como $-(1 + \beta)(1 - u)e^{(-1-\beta)} \leq -1$, $Z(u)e^{Z(u)}$ tem como solução uma função W de Lambert com ramo inferior, isto é,

$$Z(u) = W_{-1}(-(1 + \beta)(1 - u)e^{(-1-\beta)}).$$

Voltando a variável original, obtém-se:

$$Q(u) = \left[-1 - \frac{1}{\beta} - \frac{1}{\beta}W_{-1}(-(1 + \beta)e^{(-1-\beta)}(1 - u))\right]^{\frac{1}{\alpha}},$$

e, portanto, o resultado segue usando o fato de que X ser uma variável aleatória discreta. ■

Seja x_1, \dots, x_n uma amostra aleatória de tamanho n da distribuição LPD com parâmetros α, β e função de probabilidade dada por (3.33). A função de verossimilhança pode ser escrita na forma:

$$L(\alpha, \beta | \mathbf{x}) = \prod_{i=1}^n \left[\left(1 + \frac{\beta x_i^\alpha}{\beta + 1}\right) \gamma^{x_i^\alpha} - \left(1 + \frac{\beta(x_i + 1)^\alpha}{\beta + 1}\right) \gamma^{(x_i+1)^\alpha} \right]. \quad (3.42)$$

Aplicando a função logarítmica em (3.42) tem-se que o logaritmo da função de verossimilhança é dado por:

$$\ell(\alpha, \beta | \mathbf{x}) = -n \log(\beta + 1) + \sum_{i=1}^n \log \left[(1 + \beta + \beta x_i^\alpha) \gamma^{x_i^\alpha} - (1 + \beta + \beta(x_i + 1)^\alpha) \gamma^{(x_i+1)^\alpha} \right]. \quad (3.43)$$

É importante destacar que os estimadores de máxima verossimilhança de α e β são obtidos resolvendo-se $U(\alpha, \beta | \mathbf{x}) = \left(\frac{d}{d\alpha} \ell(\alpha, \beta | \mathbf{x}), \frac{d}{d\beta} \ell(\alpha, \beta | \mathbf{x})\right) = 0$ em α e β em que:

$$U(\alpha, \beta | \mathbf{x}) = \begin{cases} \sum_{i=1}^n \frac{\beta x_i^\alpha \log x_i}{1 + \beta + \beta x_i^\alpha} + \sum_{i=1}^n \frac{\beta(x_i + 1)^\alpha \log(1 + x_i) \gamma^{(x_i+1)^\alpha} - x_i^\alpha \log x_i \gamma^{x_i^\alpha}}{\gamma^{x_i^\alpha} - \gamma^{(x_i+1)^\alpha}} \\ \frac{-n}{1 + \beta} + \sum_{i=1}^n \frac{1 + x_i^\alpha}{1 + \beta + \beta x_i^\alpha} + \sum_{i=1}^n \frac{(x_i + 1)^\alpha \gamma^{(x_i+1)^\alpha} - x_i^\alpha \gamma^{x_i^\alpha}}{\gamma^{x_i^\alpha} - \gamma^{(x_i+1)^\alpha}} \end{cases} \quad (3.44)$$

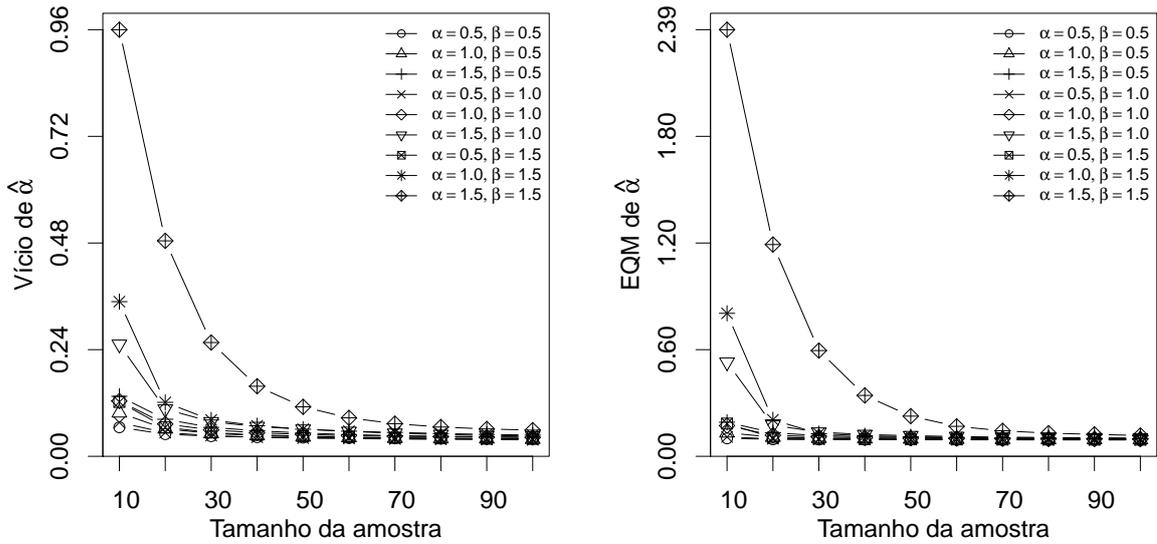


Figura 13 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta.

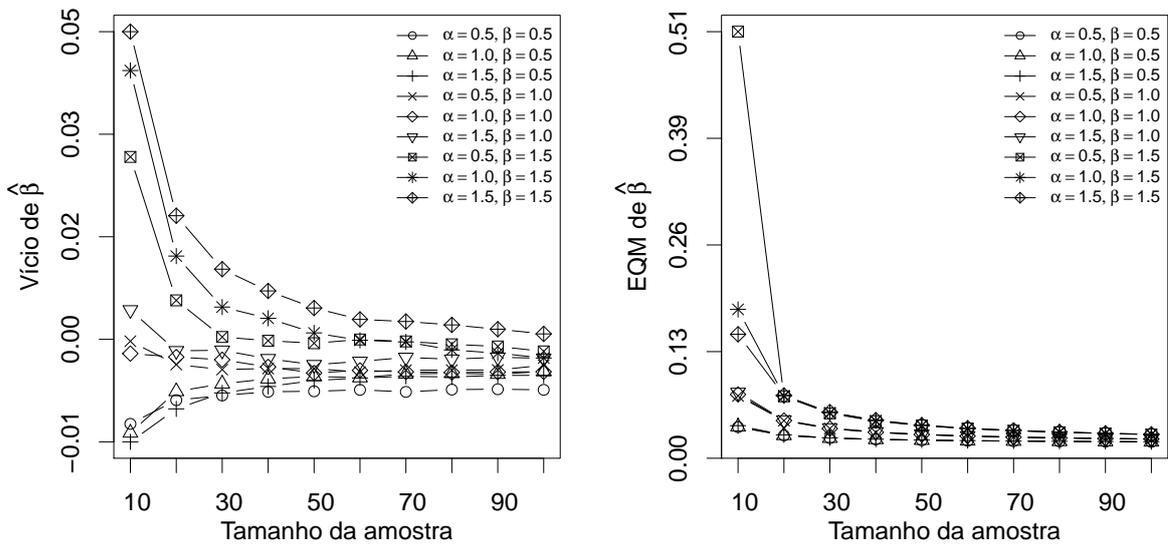


Figura 14 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta.

Note que a equação acima é não-linear e não possui solução analítica, isto é, os estimadores de máxima verossimilhança não possuem forma fechada. Analogamente as distribuições anteriores, foi realizado um estudo de simulação Monte Carlo com objetivo de avaliar algumas propriedades de $\hat{\alpha}$ e $\hat{\beta}$. Consideramos os cenários $(\alpha, \beta) \times (0.5, 1.0, 1.5)$ com tamanhos de amostra 20, 40, ..., 200 com passo $n = 20$. Os resultados estão dispostos nas Tabelas 4 e 5 e ilustrados nas Figuras 13 e 14.

Diante dos resultados apresentados, concluímos que a magnitude do vício e do EQM de $\hat{\alpha}$ e $\hat{\beta}$ convergem para zero quando o tamanho de amostra aumenta. Além disso, assim como nas distribuições anteriores, a mesma tem tendência exponencial. Os maiores valores para o vício e EQM de $\hat{\alpha}$ foram atingidos quando $\alpha = \beta = 1.5$ e para o vício de $\hat{\beta}$ quando $\alpha = 0.5$ e $\beta = 1.5$.

Tabela 4 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta.

n	$\beta = 0.5$			$\beta = 1.0$			$\beta = 1.5$			
	α			α			α			
	0.5	1.0	1.5	0.5	1.0	1.5	0.5	1.0	1.5	
Vício	20	0.0331	0.0668	0.1052	0.0442	0.0942	0.2262	0.0921	0.3254	0.9583
	40	0.0171	0.0294	0.0518	0.0208	0.0407	0.0755	0.0330	0.0911	0.4669
	60	0.0124	0.0188	0.0329	0.0137	0.0271	0.0459	0.0208	0.0501	0.2303
	80	0.0101	0.0139	0.0249	0.0098	0.0202	0.0348	0.0153	0.0367	0.1286
	100	0.0089	0.0108	0.0198	0.0077	0.0166	0.0286	0.0124	0.0287	0.0806
	120	0.0080	0.0094	0.0161	0.0069	0.0136	0.0235	0.0101	0.0241	0.0550
	140	0.0077	0.0077	0.0137	0.0058	0.0121	0.0192	0.0085	0.0212	0.0415
	160	0.0071	0.0069	0.0121	0.0050	0.0111	0.0164	0.0073	0.0186	0.0337
	180	0.0067	0.0063	0.0109	0.0045	0.0099	0.0141	0.0063	0.0166	0.0292
	200	0.0065	0.0054	0.0101	0.0037	0.0093	0.0127	0.0059	0.0149	0.0264
EQM	20	0.0099	0.0420	0.1036	0.0179	0.0838	0.4539	0.0960	0.7393	2.3946
	40	0.0041	0.0168	0.0423	0.0069	0.0283	0.0869	0.0137	0.1179	1.1408
	60	0.0026	0.0105	0.0259	0.0043	0.0178	0.0460	0.0079	0.0371	0.5220
	80	0.0019	0.0077	0.0189	0.0030	0.0129	0.0321	0.0055	0.0250	0.2602
	100	0.0015	0.0061	0.0148	0.0024	0.0102	0.0253	0.0042	0.0189	0.1394
	120	0.0012	0.0050	0.0123	0.0019	0.0083	0.0202	0.0034	0.0154	0.0798
	140	0.0010	0.0043	0.0104	0.0017	0.0071	0.0168	0.0029	0.0130	0.0538
	160	0.0009	0.0037	0.0090	0.0014	0.0061	0.0145	0.0025	0.0110	0.0401
	180	0.0008	0.0033	0.0080	0.0013	0.0055	0.0128	0.0022	0.0096	0.0330
	200	0.0007	0.0029	0.0072	0.0011	0.0049	0.0114	0.0020	0.0085	0.0263

Tabela 5 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta.

n	$\beta = 0.5$			$\beta = 1.0$			$\beta = 1.5$			
	α			α			α			
	0.5	1.0	1.5	0.5	1.0	1.5	0.5	1.0	1.5	
Vício	20	-0.0079	-0.0092	-0.0104	0.0036	0.0019	0.0080	0.0293	0.0414	0.0468
	40	-0.0046	-0.0034	-0.0058	0.0003	0.0014	0.0023	0.0093	0.0155	0.0211
	60	-0.0040	-0.0023	-0.0036	-0.0003	0.0010	0.0023	0.0042	0.0084	0.0136
	80	-0.0034	-0.0016	-0.0027	-0.0002	0.0000	0.0012	0.0037	0.0068	0.0106
	100	-0.0034	-0.0013	-0.0019	-0.0002	-0.0008	0.0003	0.0033	0.0048	0.0082
	120	-0.0032	-0.0015	-0.0015	-0.0007	-0.0005	0.0008	0.0038	0.0037	0.0066
	140	-0.0035	-0.0010	-0.0013	-0.0004	-0.0007	0.0013	0.0036	0.0035	0.0064
	160	-0.0031	-0.0009	-0.0014	-0.0004	-0.0008	0.0011	0.0032	0.0024	0.0059
	180	-0.0031	-0.0010	-0.0012	-0.0004	-0.0007	0.0014	0.0029	0.0019	0.0053
	200	-0.0032	-0.0006	-0.0012	0.0003	-0.0007	0.0012	0.0022	0.0013	0.0046
EQM	20	0.0204	0.0209	0.0222	0.0594	0.0613	0.0638	0.5149	0.1676	0.1365
	40	0.0100	0.0101	0.0107	0.0285	0.0286	0.0292	0.0589	0.0597	0.0600
	60	0.0067	0.0066	0.0070	0.0186	0.0186	0.0192	0.0376	0.0385	0.0391
	80	0.0050	0.0050	0.0053	0.0137	0.0140	0.0145	0.0278	0.0284	0.0292
	100	0.0040	0.0041	0.0042	0.0109	0.0112	0.0114	0.0225	0.0222	0.0230
	120	0.0033	0.0034	0.0036	0.0090	0.0093	0.0094	0.0186	0.0185	0.0191
	140	0.0029	0.0029	0.0031	0.0077	0.0079	0.0080	0.0159	0.0161	0.0164
	160	0.0025	0.0026	0.0027	0.0068	0.0069	0.0070	0.0140	0.0141	0.0144
	180	0.0022	0.0023	0.0024	0.0060	0.0061	0.0062	0.0123	0.0126	0.0127
	200	0.0020	0.0020	0.0021	0.0054	0.0056	0.0056	0.0110	0.0114	0.0114

Embora os estimadores de máxima verossimilhança de α e β não tem forma explícita, os intervalos de confiança para $\hat{\alpha}$ e $\hat{\beta}$ podem ser facilmente contruídos por métodos numéricos a partir da normalidade assintóticas dos dos mesmos. Os métodos Bayesianos podem ser usados com o mesmo propósito.

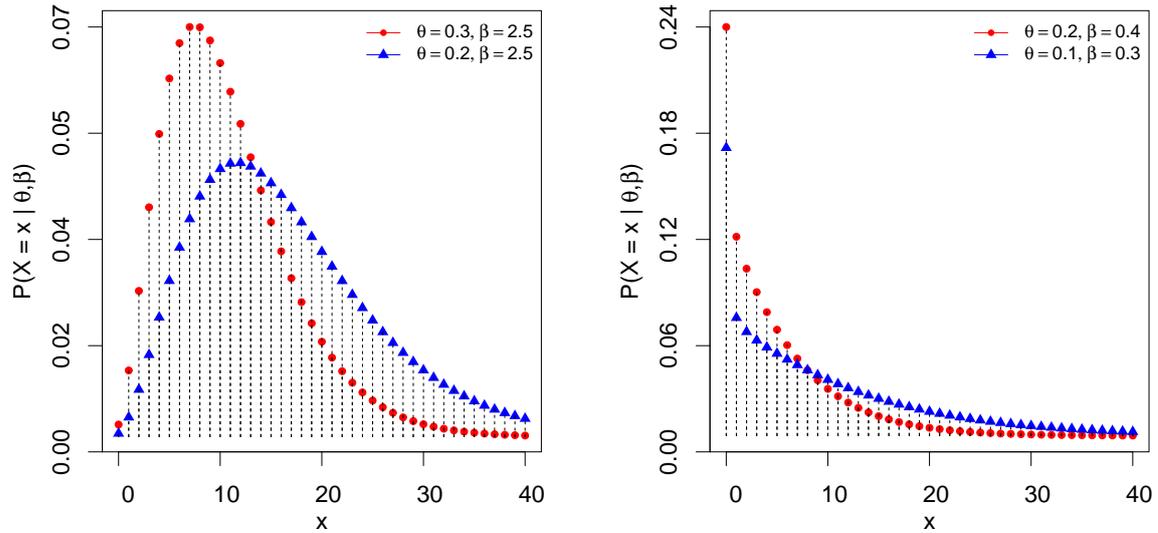


Figura 15 – Comportamento da função de probabilidade da distribuição Lindley ponderada discreta para diferentes valores de θ e β .

3.2.5 Distribuição Lindley Ponderada Discreta

Ghitany et al. (2011) introduziu a distribuição Lindley ponderada com função densidade de probabilidade dada por:

$$f(x | \theta, \beta) = \frac{\theta^{\beta+1}}{(\theta + \beta)\Gamma(\beta)} x^{\beta-1} (1+x)e^{-\theta x}, \quad x > 0, \quad \theta, \beta > 0, \quad (3.45)$$

em que:

$$\Gamma(\beta) = \int_0^{\infty} y^{\beta-1} e^{-y} dy, \quad \beta > 0 \quad (3.46)$$

é a função gama completa. A função de sobrevivência correspondente é dada por:

$$S(x | \theta, \beta) = \frac{(\theta + \beta)\Gamma(\beta, \theta x) + (\theta x)^{\beta} e^{-\theta x}}{(\theta + \beta)\Gamma(\beta)}, \quad x > 0, \quad \theta, \beta > 0, \quad (3.47)$$

em que:

$$\Gamma(a, z) = \int_z^{\infty} y^{a-1} e^{-y} dy, \quad a > 0, \quad z \geq 0, \quad (3.48)$$

é a função gama incompleta superior.

De maneira análoga as distribuições anteriores, propomos a versão discreta da distribuição Lindley ponderada por meio do método de Nakagawa e Osaki (1975) com função de probabilidade definida como:

$$P(X = x | \theta, \beta) = \frac{\Gamma(\beta, \theta x) - \Gamma(\beta, \theta(x+1))}{\Gamma(\beta)} + \frac{\theta^{\beta} e^{-\theta x} [x^{\beta} - (x+1)^{\beta} e^{-\theta}]}{(\theta + \beta)\Gamma(\beta)}, \quad (3.49)$$

em que $\theta, \beta > 0$ e o suporte em $x \in \mathbb{Z}_+$. Note que se $\beta = 1$, então a função de probabilidade $P(X = x | \theta, \beta)$ acima corresponde a função de probabilidade da versão discreta da distribuição de Lindley apresentada por Bakouch, Jazi e Nadarajah (2014).

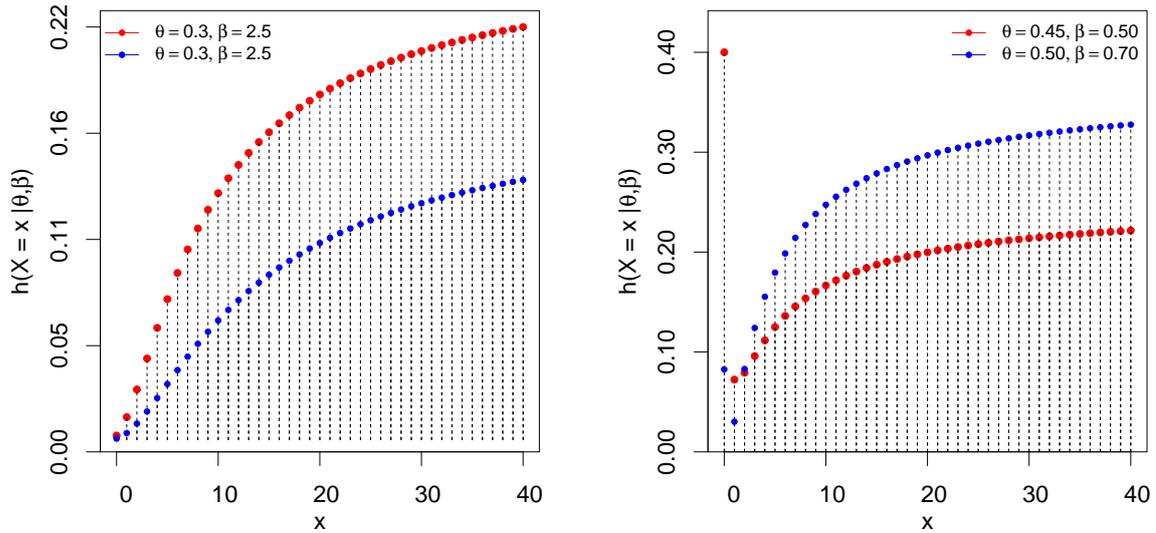


Figura 16 – Comportamento da função de risco da distribuição Lindley ponderada discreta para diferentes valores de θ e β .

Neste trabalho, a discretização da distribuição Lindley ponderada foi também obtida apenas pelo método de Nakagawa e Osaki (1975), pois os outros métodos geraram problemas de convergência.

O comportamento de (3.49) é ilustrado na Figura 15. Pode-se notar que o mesmo é unimodal visto que satisfaz a inequação de log-concavidade $[P(X = x)]^2 \geq P(X = x - 1)P(X = x + 1)$ e o resultado segue do Teorema 3 de Keilson e Gerber (1971).

A função de distribuição acumulada, de sobrevivência e de risco correspondentes são, respectivamente, dadas por:

$$F(x | \theta, \beta) = 1 - \frac{(\theta + \beta)\Gamma(\beta, \theta x) + (\theta x)^\beta e^{-\theta x}}{(\theta + \beta)\Gamma(\beta)}, \quad (3.50)$$

$$S(x | \theta, \beta) = \frac{(\theta + \beta)\Gamma(\beta, \theta x) + (\theta x)^\beta e^{-\theta x}}{(\theta + \beta)\Gamma(\beta)}, \quad (3.51)$$

e,

$$h(x | \theta, \beta) = 1 - \frac{(\theta + \beta)\Gamma(\beta, \theta x + \theta) + \theta^\beta (x + 1)^\beta e^{-\theta(x+1)}}{(\theta + \beta)\Gamma(\beta, \theta x) + (\theta x)^\beta e^{-\theta x}}. \quad (3.52)$$

O comportamento de (3.52) para alguns valores de θ e β é ilustrado na Figura 16 e pode-se notar que o mesmo é crescente e decrescente-crescente nos valores de θ e β considerados. Além disso, para $x \rightarrow 0$ e $x \rightarrow \infty$ temos:

$$h(0 | \theta, \beta) = 1 - \left[\frac{\Gamma(\beta, \theta)}{\Gamma(\beta)} + \frac{\theta^\beta e^{-\theta}}{(\theta + \beta)\Gamma(\beta)} \right] \quad (3.53)$$

e,

$$h(\infty | \theta, \beta) = 1 - e^{-\theta} \quad (3.54)$$

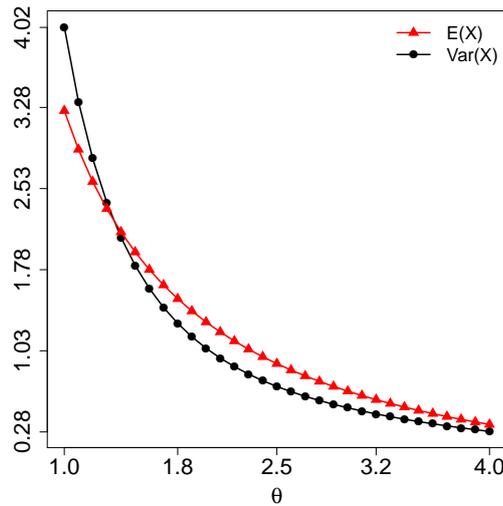


Figura 17 – Esperança e variância distribuição Lindley ponderada discreta para alguns valores de θ e $\beta = 3$.

Em relação aos momentos, obtemos de (3.49) que o momento de ordem k da distribuição Lindley ponderada discreta é dado por:

$$\mathbb{E}(X^k) = \sum_{x=0}^{\infty} \frac{x^k [\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)]}{\Gamma(\beta)} - \frac{\theta^\beta}{(\theta + \beta)\Gamma(\beta)} \left[K(k + \beta, e^\theta) - \sum_{x=0}^{\infty} x^k e^{-\theta(x+1)} (x + 1)^\beta \right] \quad (3.55)$$

em que $K(a, b)$ é o mesmo descrito na equação 3.40. De (3.55), obtemos que a esperança e variância da distribuição Lindley ponderada discreta são dadas, respectivamente, por:

$$\mathbb{E}(X) = \sum_{x=0}^{\infty} \frac{x [\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)]}{\Gamma(\beta)} + \frac{\theta^\beta}{(\theta + \beta)\Gamma(\beta)} \left[\sum_{x=0}^{\infty} x^{\beta+1} e^{-\theta x} - \sum_{x=0}^{\infty} x e^{-\theta(x+1)} (x + 1)^\beta \right] \quad (3.56)$$

e,

$$\begin{aligned} \text{Var}(X) &= \sum_{x=0}^{\infty} \frac{x^2 [\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)]}{\Gamma(\beta)} + \frac{\theta^\beta}{(\theta + \beta)\Gamma(\beta)} \left[\sum_{x=0}^{\infty} x^{\beta+2} e^{-\theta x} - \sum_{x=0}^{\infty} x^2 e^{-\theta(x+1)} (x + 1)^\beta \right] \\ &- \left\{ \sum_{x=0}^{\infty} \frac{x [\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)]}{\Gamma(\beta)} + \frac{\theta^\beta}{(\theta + \beta)\Gamma(\beta)} \left[\sum_{x=0}^{\infty} x^{\beta+1} e^{-\theta x} - \sum_{x=0}^{\infty} x e^{-\theta(x+1)} (x + 1)^\beta \right] \right\}^2 \end{aligned} \quad (3.57)$$

Assim como a distribuição Lindley potência discreta, a distribuição Lindley ponderada discreta não possui forma fechada para a esperança e a variância. Na Figura 17, ilustramos a esperança e a variância para alguns valores de θ e $\beta = 3$.

Na Tabela 6, exibimos o índice de dispersão da distribuição Lindley ponderada discreta e pode-se notar que essa distribuição pode ser usada para dados com superdispersão e subdispersão.

Tabela 6 – Índice de dispersão da distribuição Lindley ponderada discreta.

$\beta \backslash \theta$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
0.5	3.001	1.820	1.424	1.238	1.140	1.084	1.050	1.031	1.019	1.011	1.007	1.004
1.0	2.650	1.636	1.316	1.172	1.098	1.057	1.034	1.020	1.012	1.007	1.004	1.003
1.5	2.453	1.483	1.195	1.078	1.027	1.005	0.996	0.994	0.994	0.995	0.996	0.997
2.0	2.338	1.370	1.086	0.978	0.940	0.932	0.938	0.949	0.960	0.970	0.978	0.985
2.5	2.266	1.292	1.000	0.887	0.850	0.849	0.865	0.888	0.911	0.932	0.950	0.963
3.0	2.218	1.238	0.936	0.813	0.769	0.765	0.785	0.815	0.849	0.881	0.908	0.931
3.5	2.184	1.200	0.890	0.757	0.701	0.690	0.706	0.738	0.777	0.818	0.855	0.888
4.0	2.159	1.173	0.858	0.715	0.648	0.626	0.634	0.662	0.702	0.747	0.792	0.834
4.5	2.140	1.152	0.833	0.685	0.609	0.576	0.573	0.592	0.628	0.673	0.722	0.771
5.0	2.125	1.135	0.815	0.663	0.581	0.539	0.524	0.533	0.560	0.601	0.651	0.703
5.5	2.113	1.122	0.800	0.646	0.560	0.511	0.488	0.485	0.502	0.536	0.580	0.632
6.0	2.102	1.111	0.788	0.632	0.544	0.491	0.461	0.449	0.455	0.478	0.516	0.563

Seja x_1, \dots, x_n uma amostra aleatória de tamanho n da distribuição Lindley ponderada discreta. O logaritmo da função de verossimilhança é dado por:

$$\ell(\theta, \beta | \mathbf{x}) = \sum_{i=1}^n \log \left[\frac{\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)}{\Gamma(\beta)} + \frac{\theta^\beta e^{-\theta x} [x^\beta - (x + 1)^\beta e^{-\theta}]}{(\theta + \beta)\Gamma(\beta)} \right] \quad (3.58)$$

Assim como a distribuição Lindley potência discreta proposta neste trabalho, a distribuição Lindley ponderada discreta não possui forma explícita para os estimadores de verossimilhança de θ e β pois $U(\theta, \beta | \mathbf{x}) = \left(\frac{d}{d\theta} \ell(\theta, \beta | \mathbf{x}), \frac{d}{d\beta} \ell(\theta, \beta | \mathbf{x}) \right) = 0$ não possui solução explícita para θ e β . Dessa forma, é necessário fazer o uso de métodos numéricos para avaliar os estimadores de máxima verossimilhança.

Então, a fim de avaliar as propriedades $\hat{\theta}$ e $\hat{\beta}$, realizamos um estudo de simulação Monte Carlo no qual foi avaliado o vício e o EQM dos mesmos. Como cenários, foi considerado todas combinações de (θ, β) , em que $\theta = 0.5, 0.8, 1.0$ e $\beta = 0.5, 0.6, 0.7$, com tamanhos de amostra 30, 50, ..., 150. Os resultados deste estudo de simulação são apresentados nas Tabelas 7 e 8 e ilustrados nas Figuras 18 e 19.

Conforme as Tabelas 7 e 8, concluímos que a magnitude do vício e do EQM de $\hat{\theta}$ e $\hat{\beta}$ é positiva em todos os cenários considerados. Nas Figuras 18 e 19, pode-se notar que essa magnitude apresenta comportamento exponencial e decai para 0 conforme o tamanho amostral aumenta. Por fim, para alguns tamanhos de amostras, nota-se um valor alto para o vício e o EQM tanto de $\hat{\theta}$ quanto para $\hat{\beta}$.

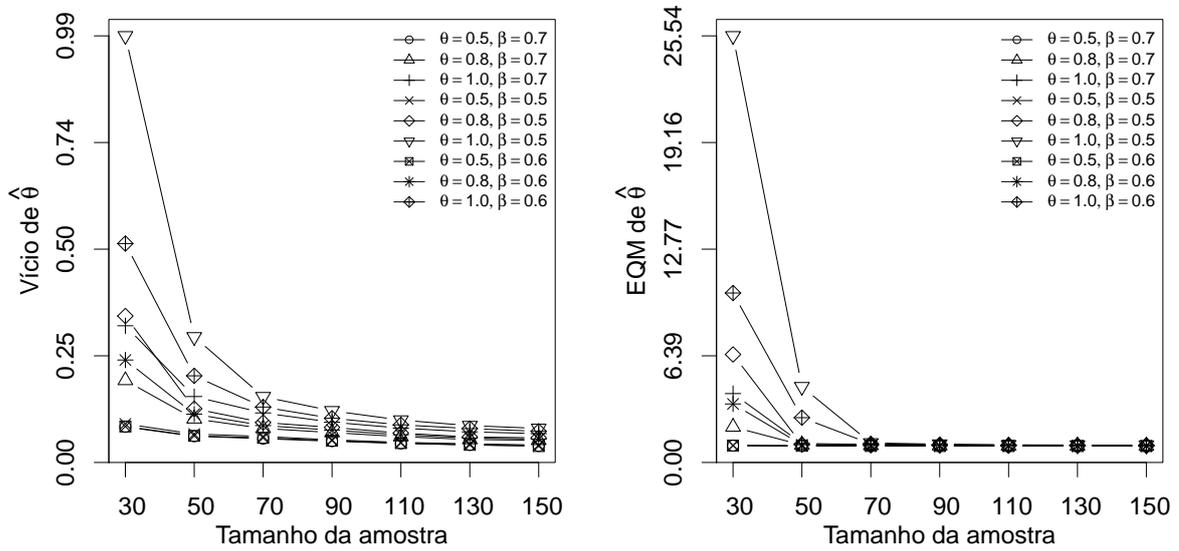


Figura 18 – Vício e EQM de $\hat{\theta}$ da distribuição Lindley ponderada discreta.

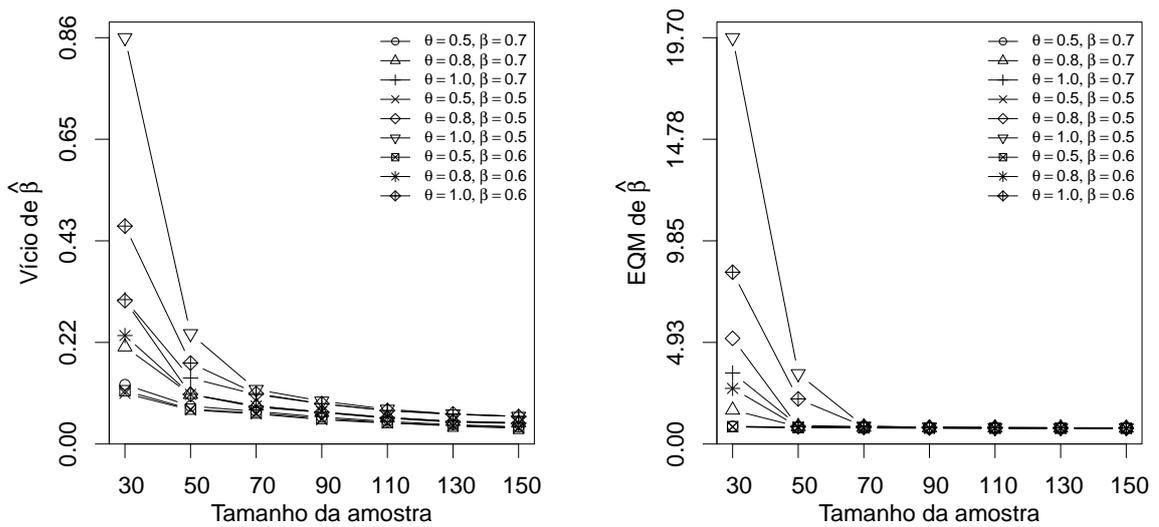


Figura 19 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley ponderada discreta.

Tabela 7 – Vício e EQM de $\hat{\theta}$ da distribuição Lindley ponderada discreta.

n	$\beta = 0.7$			$\beta = 0.5$			$\beta = 0.6$			
	θ			θ			θ			
	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0	
Vício	30	0.0569	0.1679	0.2985	0.0638	0.3218	0.9913	0.0578	0.2166	0.4949
	50	0.0344	0.0761	0.1293	0.0397	0.1002	0.2719	0.0353	0.0873	0.1788
	70	0.0286	0.0531	0.0902	0.0340	0.0672	0.1289	0.0305	0.0599	0.1042
	90	0.0225	0.0433	0.0680	0.0259	0.0555	0.0952	0.0238	0.0491	0.0774
	110	0.0171	0.0335	0.0538	0.0198	0.0420	0.0730	0.0181	0.0380	0.0617
	130	0.0137	0.0267	0.0449	0.0164	0.0327	0.0598	0.0138	0.0303	0.0527
	150	0.0112	0.0247	0.0407	0.0135	0.0306	0.0532	0.0110	0.0277	0.0462
EQM	30	0.0311	1.1950	3.2711	0.0399	5.7061	25.5402	0.0339	2.6259	9.5315
	50	0.0170	0.0677	0.1649	0.0206	0.1027	3.6971	0.0176	0.0836	1.7709
	70	0.0118	0.0425	0.1063	0.0146	0.0616	0.1987	0.0126	0.0501	0.1409
	90	0.0083	0.0315	0.0728	0.0104	0.0440	0.1227	0.0090	0.0358	0.0905
	110	0.0064	0.0228	0.0545	0.0078	0.0301	0.0835	0.0069	0.0258	0.0660
	130	0.0053	0.0195	0.0429	0.0064	0.0238	0.0667	0.0056	0.0217	0.0525
	150	0.0044	0.0165	0.0382	0.0053	0.0207	0.0563	0.0047	0.0184	0.0455

Tabela 8 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley ponderada discreta.

n	$\beta = 0.7$			$\beta = 0.5$			$\beta = 0.6$			
	θ			θ			θ			
	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0	
Vício	30	0.1115	0.1928	0.2949	0.0926	0.2926	0.8608	0.0982	0.2173	0.4538
	50	0.0649	0.0902	0.1255	0.0565	0.0904	0.2213	0.0584	0.0903	0.1579
	70	0.0535	0.0654	0.0898	0.0480	0.0627	0.1008	0.0500	0.0650	0.0920
	90	0.0413	0.0513	0.0700	0.0355	0.0512	0.0758	0.0380	0.0525	0.0702
	110	0.0317	0.0400	0.0550	0.0281	0.0386	0.0581	0.0293	0.0406	0.0554
	130	0.0249	0.0319	0.0468	0.0233	0.0301	0.0479	0.0217	0.0322	0.0481
	150	0.0200	0.0290	0.0427	0.0192	0.0275	0.0428	0.0168	0.0288	0.0423
EQM	30	0.1198	0.9351	2.8021	0.0818	4.5481	19.7014	0.0981	2.0254	7.8823
	50	0.0649	0.0995	0.1558	0.0428	0.0861	2.7663	0.0514	0.0937	1.4950
	70	0.0456	0.0634	0.1023	0.0307	0.0524	0.1194	0.0367	0.0575	0.1095
	90	0.0332	0.0461	0.0723	0.0227	0.0367	0.0725	0.0269	0.0408	0.0708
	110	0.0257	0.0345	0.0540	0.0175	0.0258	0.0494	0.0211	0.0302	0.0514
	130	0.0211	0.0299	0.0439	0.0145	0.0206	0.0398	0.0170	0.0254	0.0421
	150	0.0173	0.0262	0.0392	0.0118	0.0184	0.0337	0.0143	0.0224	0.0365

3.3 Modelos Discretos Truncados em Zero

Nesta seção, propomos as versões truncada em zero das discretizações da distribuição Lindley e algumas de suas variantes apresentadas na seção anterior. Dizemos que uma variável aleatória discreta é dita truncada em zero se seu suporte (domínio) é estritamente positivo (GHITANY; AL-MUTAIRI; NADARAJAH, 2008). Dessa forma, sua função de probabilidade é escrita na forma:

$$P(X = x) = \frac{P(X = x)}{1 - P(X = 0)}. \tag{3.59}$$

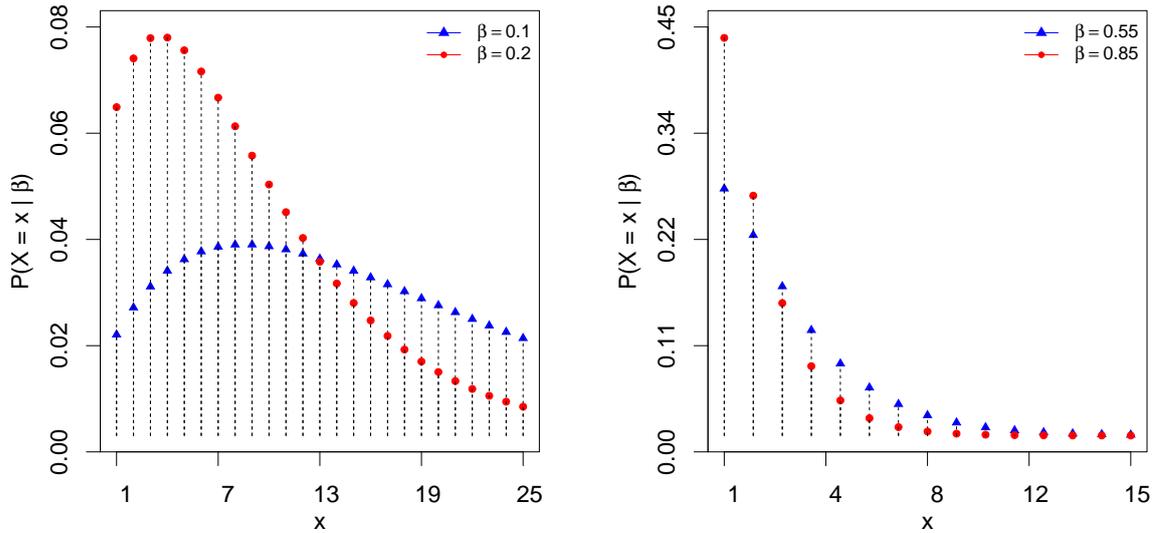


Figura 20 – Comportamento da função de probabilidade da distribuição Lindley discreta truncada em zero para alguns valores de β .

3.3.1 Distribuição Lindley Discreta Truncada em Zero (via Função de Sobre- vivência)

A distribuição Lindley discreta apresentada por Bakouch, Jazi e Nadarajah (2014) tem suporte em $x = 0, 1, \dots$ o que impossibilita a mesma de ser aplicada diretamente, por exemplo, para modelar a distribuição do comprimento da sequência de dias chuvosos ou secos (DEKA; BORAH; KAKATY, 2010), o número de dias de estadia em um hospital, entre outros (BÖHNING; KUHNERT, 2006) já que para este tipo de situação, o suporte é estritamente positivo. Assim, para esta de situação, propomos a versão truncada em zero da distribuição Lindley discreta com função de probabilidade escrita na forma:

$$P(X = x | \beta) = \frac{e^{-\beta x} [\beta(1 - 2e^{-\beta}) + (1 - e^{-\beta})(1 + \beta x)]}{e^{-\beta}(1 + 2\beta)}, \quad (3.60)$$

em que $x = 1, 2, \dots$ e $\beta > 0$.

O comportamento de (3.60) é ilustrado na Figura 20. Além disso, facilmente verificar-se que (3.60) satisfaz a equação de log-concavidade, $[P(X = x)]^2 \leq P(X = x-1)P(X = x+1)$, e, pelo Teorema 3 de Keilson e Gerber (1971), o comportamento da mesma é unimodal.

A função de distribuição acumulada, de sobrevivência e risco são dadas, respectivamente, por:

$$F(x | \beta) = 1 - \frac{e^{-\beta x}(1 + \beta + \beta x)}{e^{-\beta}(2\beta + 1)} \quad (3.61)$$

$$S(x | \beta) = \frac{e^{-\beta x}(1 + \beta + \beta x)}{e^{-\beta}(2\beta + 1)} \quad (3.62)$$

e,

$$h(x | \beta) = \frac{e^{-\beta x} [\beta(1 - 2e^{-\beta}) + (1 - e^{-\beta})(1 + \beta x)]}{e^{-\beta x}(1 + \beta + \beta x)}. \quad (3.63)$$

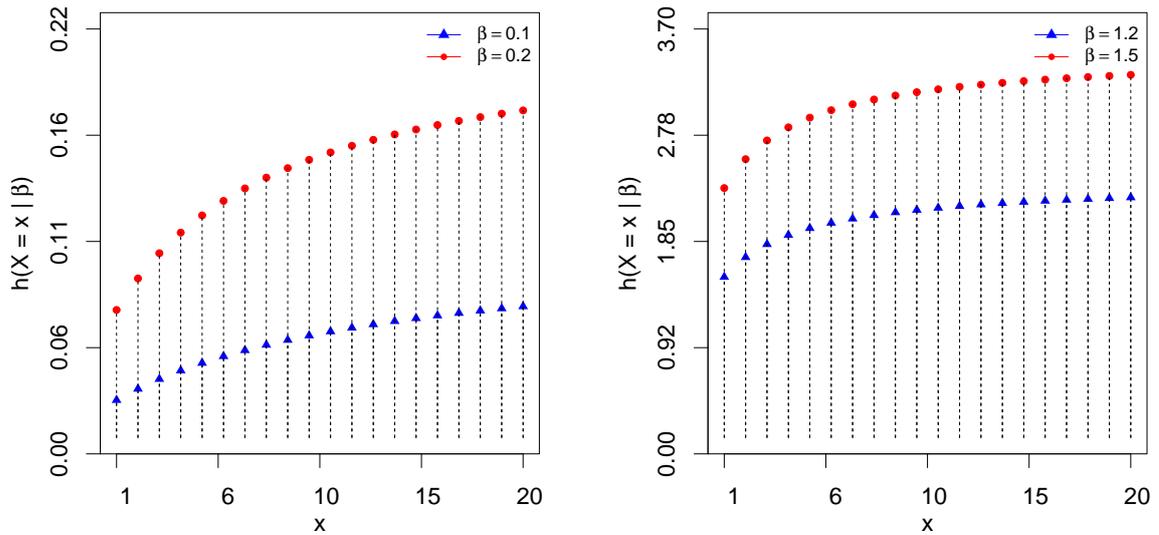


Figura 21 – Comportamento da função de risco da distribuição Lindley discreta truncada em zero (via função de sobrevivência) para diferentes valores de β .

em que $x = 1, 2, \dots$ e $\beta > 0$. O comportamento da função de risco é ilustrado na Figura 21 e o mesmo é crescente para $\beta > 0$.

A partir de (3.60), tem-se que a esperança e a variância são descritas, respectivamente, como:

$$\mathbb{E}(X) = \frac{2\beta - \beta e^{-\beta} + 1 - e^{-\beta}}{(1 + 2\beta)(1 - e^{-\beta})^2} \tag{3.64}$$

e,

$$\text{Var}(X) = \frac{(\beta e^{-\beta} - 4\beta - 1)e^{-\beta}}{(1 + 2\beta)(1 - e^{-\beta})^3} - \frac{\mathbb{E}(X)}{e^{-\beta} - 1}. \tag{3.65}$$

Uma ilustração das expressões para a esperança e variância descritas acima para diferentes valores de β pode ser vista na Figura 22.

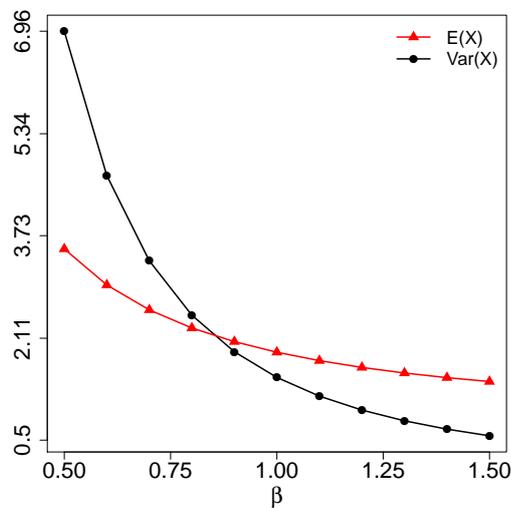


Figura 22 – Esperança e variância distribuição Lindley discreta truncada em zero (via função de sobrevivência) para alguns valores de β .

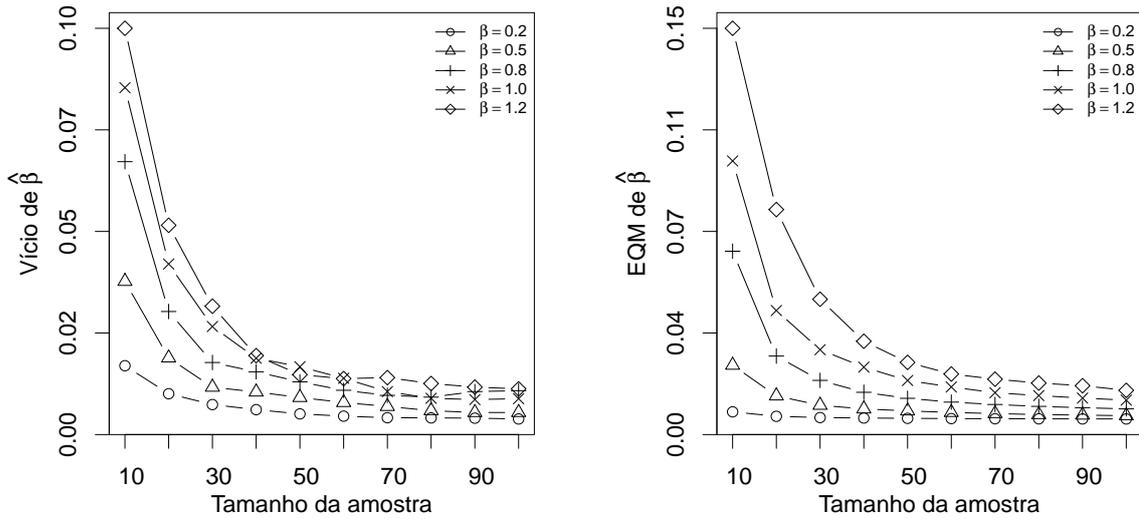


Figura 23 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta truncada em zero (via função de sobrevivência).

Por meio de métodos numéricos, foi avaliado a razão entre $\mathbb{E}(X)$ e $\text{Var}(X)$ e foi verificado que $\mathbb{E}(X) < \text{Var}(X)$ para $\beta < 0.8505$ e $\mathbb{E}(X) > \text{Var}(X)$ caso contrário.

Neste caso, a versão truncada em zero pode ser utilizada tanto para dados com superdispersão quanto para dados com subdispersão oferecendo uma maior vantagem em comparação com sua versão normal.

A função quantil de uma variável aleatória X que segue uma distribuição Lindley discreta truncada em zero é escrita como:

$$Q(u) = \left\lceil -1 - \frac{1}{\beta} - \frac{1}{\beta} W(-(1 + 2\beta) \exp(-2\beta - 1)(1 - u)) \right\rceil \quad (3.66)$$

em que W representa a função W de Lambert com ramo positivo (JODRA, 2010) e $0 < u < 1$.

Quanto a estimação, supondo uma amostra aleatória x_1, \dots, x_n de (3.60), tem-se que o logaritmo da função de verossimilhança para a distribuição Lindley discreta truncada em zero é dado por:

$$\begin{aligned} \ell(\beta | \mathbf{x}) &= n\beta - n \log(1 + 2\beta) - n\beta\bar{x} \\ &+ \sum_{i=1}^n \log [1 + (1 + x_i)\beta - (1 + 2\beta + \beta x_i) \exp(-\beta)] \end{aligned}$$

Note que, para essa distribuição, o estimador $\hat{\beta}$ do parâmetro de dispersão β não possui forma explícita. Sendo assim, por meio de um estudo de simulações Monte Carlo, avaliamos as propriedades de $\hat{\beta}$ usando tamanhos de amostra $n = 10, 20, \dots, 100$ e $\beta = 0.2, 0.5, 0.8, 1.0$ e 1.2. Os resultados desse estudo são dispostos na Tabela 9.

De acordo com a Tabela 9, tem-se que o vício de $\hat{\beta}$ é positivo e sua magnitude decai para 0 quando o tamanho da amostra aumenta. A Figura 23, ilustra tais conclusões e também uma tendência exponencial tanto para o vício quanto para o EQM.

Tabela 9 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta truncada em zero (via função de sobrevivência).

	n	Valores de β						n	Valores de β				
		0.2	0.5	0.8	1.0	1.2			0.2	0.5	0.8	1.0	1.2
Vício	10	0.0143	0.0346	0.0635	0.0814	0.0957	EQM	10	0.0030	0.0208	0.0645	0.0991	0.1500
	20	0.0075	0.0162	0.0274	0.0388	0.0482		20	0.0012	0.0091	0.0243	0.0418	0.0805
	30	0.0049	0.0091	0.0150	0.0237	0.0286		30	0.0008	0.0055	0.0150	0.0268	0.0461
	40	0.0037	0.0079	0.0128	0.0161	0.0167		40	0.0006	0.0041	0.0104	0.0201	0.0300
	50	0.0026	0.0066	0.0104	0.0140	0.0121		50	0.0005	0.0033	0.0081	0.0150	0.0219
	60	0.0021	0.0054	0.0083	0.0112	0.0112		60	0.0004	0.0027	0.0067	0.0125	0.0175
	70	0.0017	0.0044	0.0071	0.0080	0.0114		70	0.0003	0.0023	0.0058	0.0103	0.0155
	80	0.0017	0.0034	0.0067	0.0064	0.0100		80	0.0003	0.0019	0.0050	0.0092	0.0140
	90	0.0016	0.0030	0.0080	0.0061	0.0091		90	0.0002	0.0017	0.0045	0.0082	0.0130
	100	0.0014	0.0030	0.0082	0.0063	0.0087		100	0.0002	0.0015	0.0041	0.0074	0.0112

3.3.2 Distribuição Lindley Discreta Truncada em Zero (via Série Infinita)

Assim como a discretização apresentada por Bakouch, Jazi e Nadarajah (2014), a distribuição Lindley discreta (via série infinita) também não pode ser aplicada diretamente a análise de sequências de dias chuvosos ou secos, por exemplo. Sendo assim, propomos a versão truncada em zero da distribuição Lindley discreta (via série infinita) com função de probabilidade escrita na forma:

$$P(X = x | \beta) = \frac{(1 + x)e^{-\beta(x+2)}(e^\beta - 1)^2}{1 - e^{-2\beta}(e^\beta - 1)^2}. \tag{3.67}$$

em que $x > 0$ e $\beta > 0$. O comportamento de (3.67) é unimodal e ilustrado na Figura 24.

A função de distribuição acumulada, de sobrevivência e de risco são dadas, respectivamente, por:

$$F(x | \beta) = \frac{2e^\beta - 1 - e^{-\beta(x+1)}(x + 2) + e^{-\beta x}(x + 1)}{2e^\beta - 1} \tag{3.68}$$

$$S(x | \beta) = 1 - \frac{2e^\beta - 1 - e^{-\beta(x+1)}(x + 2) + e^{-\beta x}(x + 1)}{2e^\beta - 1} \tag{3.69}$$

e,

$$h(x | \beta) = \frac{e^{-\beta x} (1 + x) (e^\beta - 1)^2}{e^{-\beta(x-1)}x - e^{-\beta}xx + 2e^{-\beta(x-1)} - e^{-\beta x}}. \tag{3.70}$$

em que $x = 1, 2, \dots$ e $\beta > 0$. O comportamento da função de risco é ilustrado na Figura 21 e o mesmo é crescente para $\beta > 0$.

De (3.67), obtemos que a esperança e a variância são descritas respectivamente, como:

$$\mathbb{E}(X) = \frac{2e^{2\beta}}{(e^\beta - 1)(2e^\beta - 1)} \tag{3.71}$$

e,

$$\text{Var}(X) = \frac{2e^{2\beta} (3e^\beta - 2)}{(e^\beta - 1)^2 (2e^\beta - 1)^2}. \tag{3.72}$$

Uma ilustração das expressões para a esperança e variância descritas acima para diferentes valores de β pode ser vista na Figura 26. Analogamente as distribuições anteriores, de

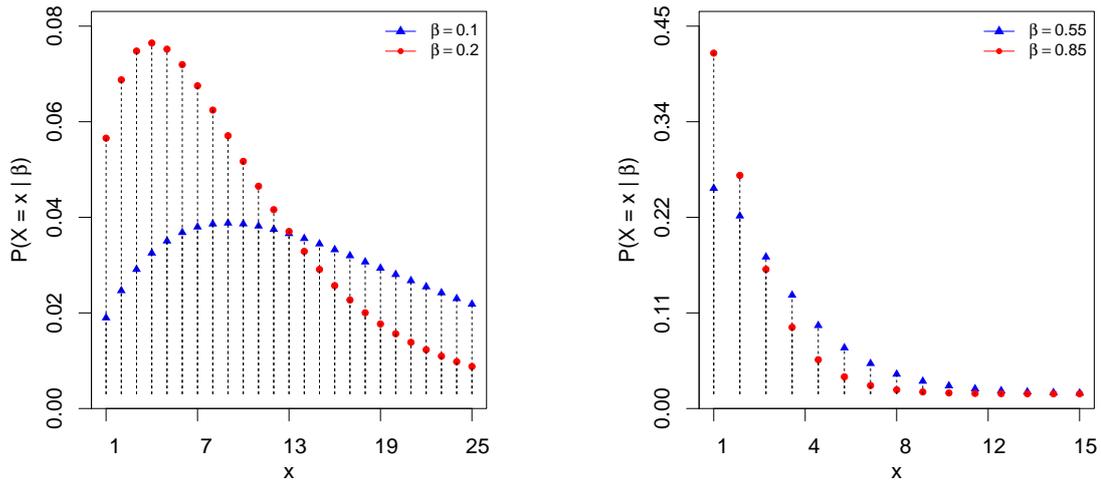


Figura 24 – Comportamento da função de probabilidade da distribuição Lindley discreta truncada em zero (via série infinita) para alguns valores de β .

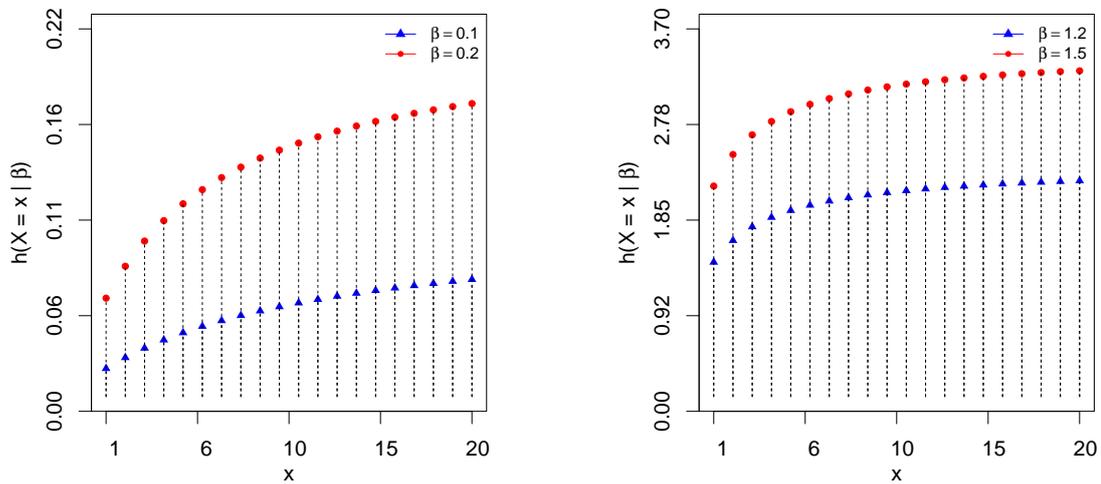


Figura 25 – Comportamento da função de risco da distribuição Lindley discreta truncada em zero (via série infinita) para alguns valores de β .

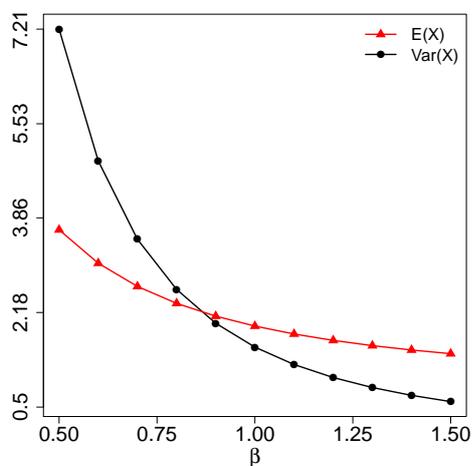


Figura 26 – Esperança e variância distribuição Lindley discreta truncada em zero (via série infinita) para alguns valores de β .

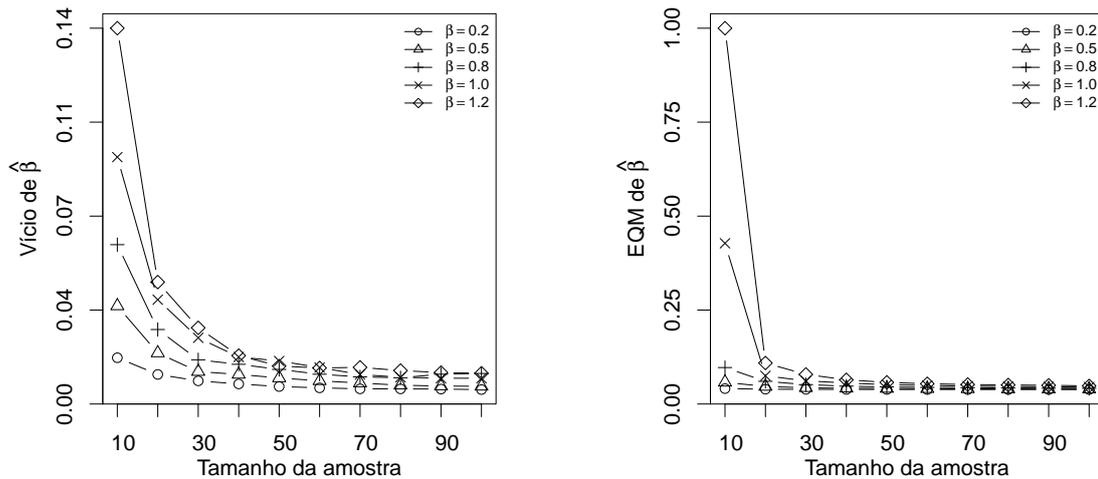


Figura 27 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta truncada em zero (via série infinita).

(3.71) e (3.72) verifica-se, numericamente, que $\mathbb{E}(X) < \text{Var}(X)$ para $0 < \beta < 0.812$, isto é, esta distribuição pode ser usada na análise de dados com superdispersão e/ou subdispersão.

A função quantil, de maneira semelhante as distribuições descritas anteriormente, também é definida em termos da função W de Lambert e é escrita na forma:

$$Q(u) = \left[\frac{1 - 2e^\beta}{(e^\beta - 1)} - W\left(\frac{\beta(2e^\beta - 1)(u - 1)e^{\frac{-2e^\beta \beta + \beta}{e^\beta - 1}}}{e^\beta - 1}\right) \beta^{-1} \right] \quad 0 < u < 1 \quad (3.73)$$

Em termos de inferência, supondo uma amostra aleatória x_1, \dots, x_n de (3.67), o estimador do parâmetro de dispersão β é obtido resolvendo-se $\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = \frac{2ne^\beta}{e^\beta - 1} - 2e^{-2\beta}(e^\beta - 1) - 2n - n\bar{x} = 0$ em β , porém, a mesma é não-linear e não possui solução explícita. Assim, para avaliar as propriedades de $\hat{\beta}$, foi realizado um estudo de simulação Monte Carlo. Os resultados desse estudo estão dispostos na Tabela 10.

Tabela 10 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta truncada em zero (via série infinita).

n	Valores de β					n	Valores de β						
	0.2	0.5	0.8	1.0	1.2		0.2	0.5	0.8	1.0	1.2		
Vício	10	0.0139	0.0344	0.0586	0.0932	0.1442	EQM	10	0.0029	0.0204	0.0612	0.4063	1.0033
	20	0.0073	0.0158	0.0251	0.0369	0.0438		20	0.0012	0.0087	0.0230	0.0383	0.0739
	30	0.0048	0.0084	0.0131	0.0218	0.0257		30	0.0008	0.0052	0.0139	0.0244	0.0422
	40	0.0035	0.0073	0.0113	0.0142	0.0147		40	0.0006	0.0039	0.0097	0.0184	0.0273
	50	0.0025	0.0059	0.0093	0.0126	0.0106		50	0.0005	0.0031	0.0075	0.0139	0.0201
	60	0.0020	0.0048	0.0074	0.0100	0.0099		60	0.0004	0.0026	0.0062	0.0116	0.0163
	70	0.0016	0.0039	0.0063	0.0073	0.0101		70	0.0003	0.0022	0.0054	0.0097	0.0142
	80	0.0016	0.0031	0.0060	0.0062	0.0088		80	0.0003	0.0018	0.0047	0.0088	0.0129
	90	0.0015	0.0027	0.0074	0.0058	0.0080		90	0.0002	0.0016	0.0042	0.0078	0.0120
	10	0.0013	0.0026	0.0077	0.0058	0.0078		100	0.0002	0.0014	0.0038	0.0072	0.0104

De acordo com a Tabela 10, nota-se que tanto o vício e quanto o EQM de $\hat{\beta}$ é positivo. Além disso, pode-se notar também que a magnitude tanto do vício e quanto do EQM de β sempre decaía para 0 quando o tamanho da amostra aumenta. A Figura 27 ilustra esse resultado.

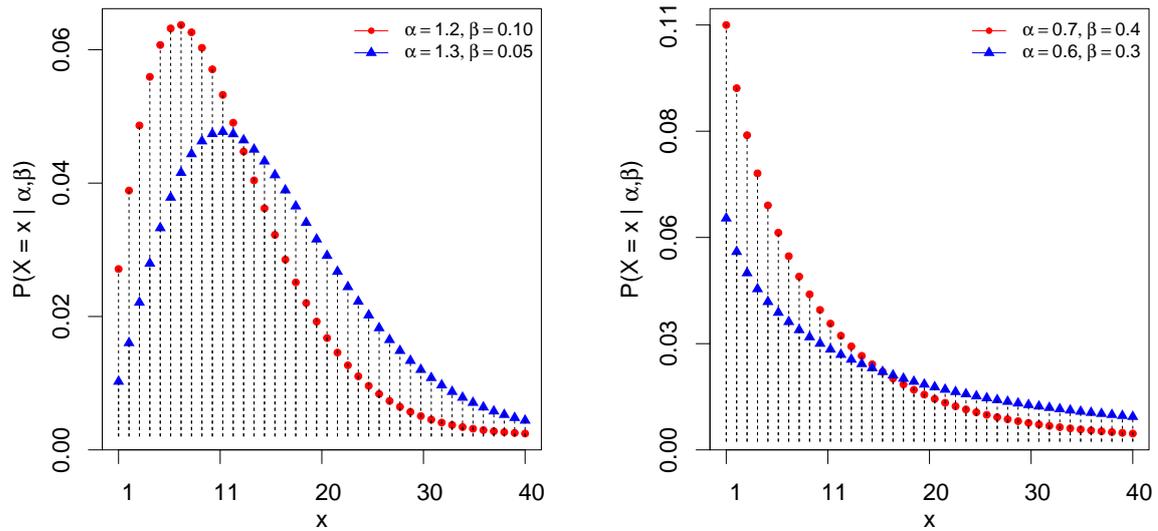


Figura 28 – Comportamento da função de probabilidade da distribuição Lindley potência discreta truncada em zero para diferentes valores de α e β .

3.3.3 Distribuição Lindley Potência Discreta Truncada em Zero

Analogamente as distribuições Lindley uni-paramétricas, propomos a versão truncada em zero da distribuição Lindley potência discreta com função de probabilidade definida por:

$$P(X = x | \alpha, \beta) = \frac{\left(1 + \frac{\beta x^\alpha}{\beta + 1}\right) \gamma^{x^\alpha} - \left(1 + \frac{\beta (x + 1)^\alpha}{\beta + 1}\right) \gamma^{(x+1)^\alpha}}{\left(1 + \frac{\beta}{\beta + 1}\right) \gamma} \quad (3.74)$$

em que $\gamma = \exp(-\beta)$. Na Figura 28 pode-se observar que o comportamento de (3.74) para diferentes valores de α e β é unimodal.

A função de distribuição acumulada, de sobrevivência e de risco são descritas, respectivamente, por:

$$F(x | \alpha, \beta) = 1 - \frac{[1 + \beta + \beta(x + 1)^\alpha] e^{\beta(1-(x+1)^\alpha)}}{(2\beta + 1)} \quad (3.75)$$

$$S(x | \alpha, \beta) = \frac{[1 + \beta + \beta(x + 1)^\alpha] e^{\beta(1-(x+1)^\alpha)}}{(2\beta + 1)} \quad (3.76)$$

e,

$$h(x | \alpha, \beta) = \frac{\gamma^{(x^\alpha - (1+x)^\alpha)} (\beta x^\alpha + \beta + 1)}{\beta(1+x)^\alpha + \beta + 1} - 1 \quad (3.77)$$

em que $\gamma = \exp(-\beta)$.

O comportamento da função de risco é ilustrado na Figura 29 e pode-se notar que o mesmo assume comportamento crescente e também decrescente para os valores de parâmetros considerados.

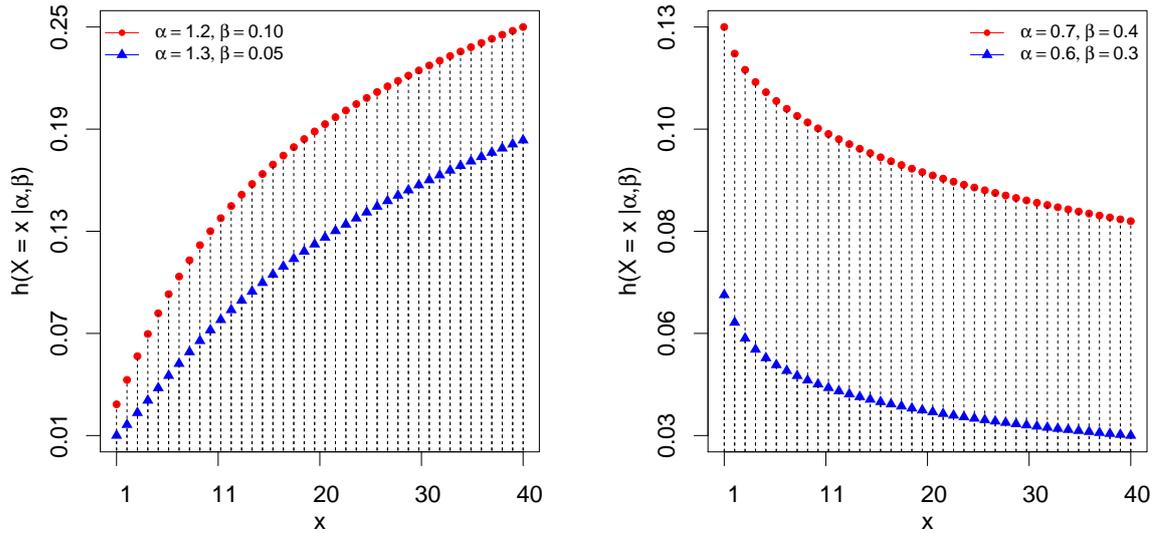


Figura 29 – Comportamento da função de probabilidade da distribuição Lindley potência discreta truncada em zero para diferentes valores de α e β .

Da equação (3.74), nota-se que o momento de ordem k não tem forma fechada mas pode ser avaliado computacionalmente por:

$$\begin{aligned} \mathbb{E}(X^k) &= \left[\left(1 + \frac{\beta}{\beta + 1} \right) \gamma \right]^{-1} \left\{ \sum_{x=0}^{\infty} x^k \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^k \frac{[-\beta(x+1)^\alpha]^j}{j!} \right. \\ &\quad \left. + \frac{\beta}{\beta + 1} \left[\sum_{x=0}^{\infty} x^{k+\alpha} \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^k (x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right] \right\}. \end{aligned} \quad (3.78)$$

E de (3.78), obtemos que a esperança e a variância podem ser avaliadas computacionalmente, respectivamente, por:

$$\begin{aligned} \mathbb{E}(X) &= \left[\left(1 + \frac{\beta}{\beta + 1} \right) \gamma \right]^{-1} \left\{ \sum_{x=0}^{\infty} x \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x \frac{[-\beta(x+1)^\alpha]^j}{j!} \right. \\ &\quad \left. + \frac{\beta}{\beta + 1} \left[\sum_{x=0}^{\infty} x^{1+\alpha} \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x(x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right] \right\} \end{aligned} \quad (3.79)$$

e,

$$\begin{aligned} \text{Var}(X) &= \left[\left(1 + \frac{\beta}{\beta + 1} \right) \gamma \right]^{-1} \left\{ \sum_{x=0}^{\infty} x^2 \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^2 \frac{[-\beta(x+1)^\alpha]^j}{j!} \right. \\ &\quad \left. + \frac{\beta}{\beta + 1} \left[\sum_{x=0}^{\infty} x^{2+\alpha} \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^2 (x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right] \right\} \\ &\quad - \left[\left(1 + \frac{\beta}{\beta + 1} \right) \gamma \right]^{-2} \left\{ \sum_{x=0}^{\infty} x \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x \frac{[-\beta(x+1)^\alpha]^j}{j!} \right. \\ &\quad \left. + \frac{\beta}{\beta + 1} \left[\sum_{x=0}^{\infty} x^{1+\alpha} \gamma^{x^\alpha} - \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x(x+1)^\alpha \frac{[-\beta(x+1)^\alpha]^j}{j!} \right] \right\}^2 \end{aligned} \quad (3.80)$$

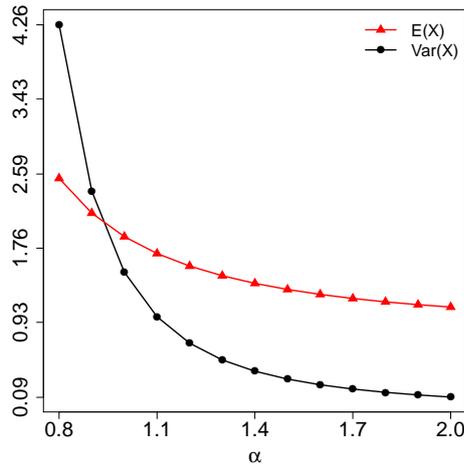


Figura 30 – Esperança e variância distribuição Lindley potência discreta truncada em zero para alguns valores de α e $\beta = 1$.

A Figura 30 ilustra a o comportamento da esperança e variância descrita nas expressões anteriores para alguns valores de α e $\beta = 3$. Pode-se notar que há evidências de superdispersão e subdispersão para esta distribuição.

No caso dos quantis, assim como todas as distribuições anteriores, a função quantil para a distribuição Lindley potência discreta truncada em zero é definida em termos da função W de Lambert e é descrita por:

$$Q(u) = \left[\left(-\frac{\beta + W_{-1} \left((u - 1) (2\beta + 1) e^{-2\beta - 1} \right) + 1}{\beta} \right)^{\alpha - 1} - 1 \right] \quad (3.81)$$

em que W_{-1} é o ramo negativo da função W de Lambert e $0 < u < 1$.

Quanto a estimação dos parâmetros, considerando uma amostra aleatória x_1, \dots, x_n de tamanho n da distribuição Lindley potência discreta truncada em zero com parâmetros α, β e função de probabilidade dada por (3.74), tem-se que o logaritmo da função de verossimilhança é dado por:

$$\begin{aligned} \ell(\alpha, \beta | \mathbf{x}) &= -n \log(2\beta + 1) - n\beta \\ &+ \sum_{i=1}^n \log \left[(1 + \beta + \beta x_i^\alpha) \gamma^{x_i^\alpha} - (1 + \beta + \beta(x_i + 1)^\alpha) \gamma^{(x_i + 1)^\alpha} \right]. \end{aligned} \quad (3.82)$$

É importante destacar que os estimadores de máxima verossimilhança de α e β são obtidos resolvendo-se $U(\alpha, \beta | \mathbf{x}) = \left(\frac{d}{d\alpha} \ell(\alpha, \beta | \mathbf{x}), \frac{d}{d\beta} \ell(\alpha, \beta | \mathbf{x}) \right) = 0$ em α e β em que:

$$U(\alpha, \beta | \mathbf{x}) = \begin{cases} \sum_{i=1}^n \frac{\beta x_i^\alpha \log x_i}{1 + \beta + \beta x_i^\alpha} + \sum_{i=1}^n \frac{\beta (x_i + 1)^\alpha \log(1 + x_i) \gamma^{(x_i + 1)^\alpha} - x_i^\alpha \log x_i \gamma^{x_i^\alpha}}{\gamma^{x_i^\alpha} - \gamma^{(x_i + 1)^\alpha}} \\ \frac{-n(3 + 2\beta)}{1 + 2\beta} + \sum_{i=1}^n \frac{1 + x_i^\alpha}{1 + \beta + \beta x_i^\alpha} + \sum_{i=1}^n \frac{(x_i + 1)^\alpha \gamma^{(x_i + 1)^\alpha} - x_i^\alpha \gamma^{x_i^\alpha}}{\gamma^{x_i^\alpha} - \gamma^{(x_i + 1)^\alpha}} \end{cases} \quad (3.83)$$

Os estimadores de máxima verossimilhança de α e β , assim como na distribuição anterior, não possuem forma explícita. Sendo assim, o sistema descrito por (3.83) deve ser solucionado por meio de métodos numéricos como, por exemplo, o método de Newton-Raphson.

Para esse estudo de simulação Monte Carlo, consideramos os cenários de todas combinações de (n, α, β) , em que $\alpha = 0.5, 0.8, 1.0$ e $\beta = 0.4, 0.5, 0.7$, com tamanhos de amostra $n = 30, 50, \dots, 150$.

Para cada cenário, foi calculado Vício($\hat{\theta}$) e EQM($\hat{\theta}$), em que $\theta = (\alpha, \beta)$, descritos anteriormente com 10.000 replicações. Os resultados desse estudo estão dispostos nas Tabelas 11 e 12 e ilustrados nas Figuras 31 e 32.

Tabela 11 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta truncada em zero.

n	$\beta = 0.7$			$\beta = 0.5$			$\beta = 0.4$			
	α			α			α			
	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0	
Vício	30	0.0434	0.0758	0.0923	0.0349	0.0597	0.0733	0.0315	0.0533	0.0627
	50	0.0269	0.0437	0.0579	0.0217	0.0362	0.0474	0.0194	0.0326	0.0405
	70	0.0226	0.0327	0.0417	0.0184	0.0269	0.0330	0.0166	0.0241	0.0272
	90	0.0174	0.0275	0.0355	0.0142	0.0232	0.0280	0.0127	0.0208	0.0240
	110	0.0138	0.0221	0.0273	0.0112	0.0190	0.0226	0.0100	0.0174	0.0194
	130	0.0111	0.0180	0.0227	0.0088	0.0154	0.0196	0.0078	0.0144	0.0165
	150	0.0089	0.0165	0.0205	0.0070	0.0145	0.0177	0.0062	0.0134	0.0150
EQM	30	0.0187	0.0535	0.0880	0.0130	0.0351	0.0569	0.0108	0.0288	0.0449
	50	0.0106	0.0289	0.0492	0.0073	0.0192	0.0329	0.0061	0.0158	0.0258
	70	0.0078	0.0198	0.0358	0.0054	0.0132	0.0231	0.0045	0.0109	0.0179
	90	0.0058	0.0154	0.0278	0.0041	0.0101	0.0177	0.0034	0.0082	0.0138
	110	0.0046	0.0113	0.0212	0.0032	0.0076	0.0138	0.0026	0.0063	0.0109
	130	0.0038	0.0097	0.0173	0.0027	0.0065	0.0114	0.0022	0.0054	0.0089
	150	0.0032	0.0084	0.0154	0.0022	0.0057	0.0101	0.0018	0.0046	0.0079

Tabela 12 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta truncada em zero.

n	$\beta = 0.7$			$\beta = 0.5$			$\beta = 0.4$			
	α			α			α			
	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0	
Vício	30	-0.0007	-0.0002	0.0279	-0.0051	-0.0085	-0.0027	-0.0061	-0.0089	-0.0032
	50	-0.0014	0.0026	0.0036	-0.0036	-0.0055	-0.0037	-0.0040	-0.0052	-0.0035
	70	-0.0060	-0.0021	0.0036	-0.0061	-0.0057	-0.0018	-0.0056	-0.0052	-0.0012
	90	-0.0037	-0.0049	-0.0028	-0.0041	-0.0074	-0.0049	-0.0037	-0.0064	-0.0043
	110	-0.0031	-0.0056	-0.0012	-0.0032	-0.0069	-0.0041	-0.0029	-0.0062	-0.0034
	130	-0.0017	-0.0038	-0.0011	-0.0017	-0.0051	-0.0041	-0.0017	-0.0050	-0.0031
	150	0.0003	-0.0031	-0.0016	-0.0004	-0.0049	-0.0040	-0.0006	-0.0047	-0.0032
EQM	30	0.1063	0.1337	0.7231	0.0447	0.0457	0.0557	0.0266	0.0258	0.0319
	50	0.0623	0.0648	0.0733	0.0267	0.0239	0.0278	0.0159	0.0148	0.0157
	70	0.0436	0.0457	0.0538	0.0189	0.0182	0.0205	0.0115	0.0111	0.0119
	90	0.0339	0.0331	0.0385	0.0146	0.0132	0.0151	0.0089	0.0081	0.0088
	110	0.0267	0.0249	0.0301	0.0117	0.0102	0.0121	0.0072	0.0063	0.0072
	130	0.0216	0.0206	0.0254	0.0096	0.0086	0.0103	0.0059	0.0053	0.0060
	150	0.0191	0.0186	0.0220	0.0085	0.0077	0.0090	0.0051	0.0047	0.0052

De acordo com as Tabelas 11 e 12, concluímos que a magnitude do vício e do EQM de $\hat{\alpha}$ e $\hat{\beta}$ convergem para zero quando o tamanho de amostra aumenta. O vício de $\hat{\alpha}$ é positivo

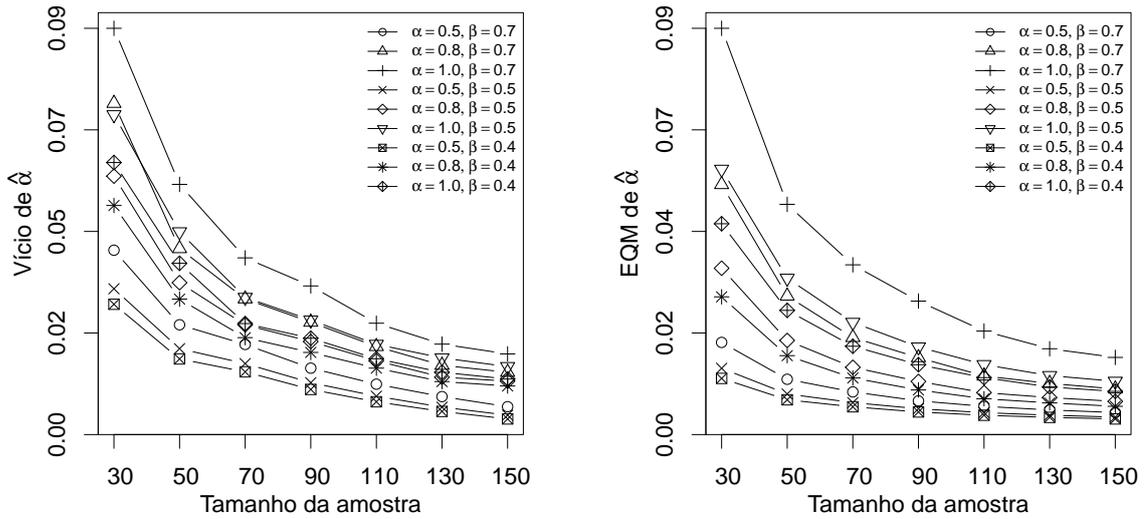


Figura 31 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta truncada em zero.

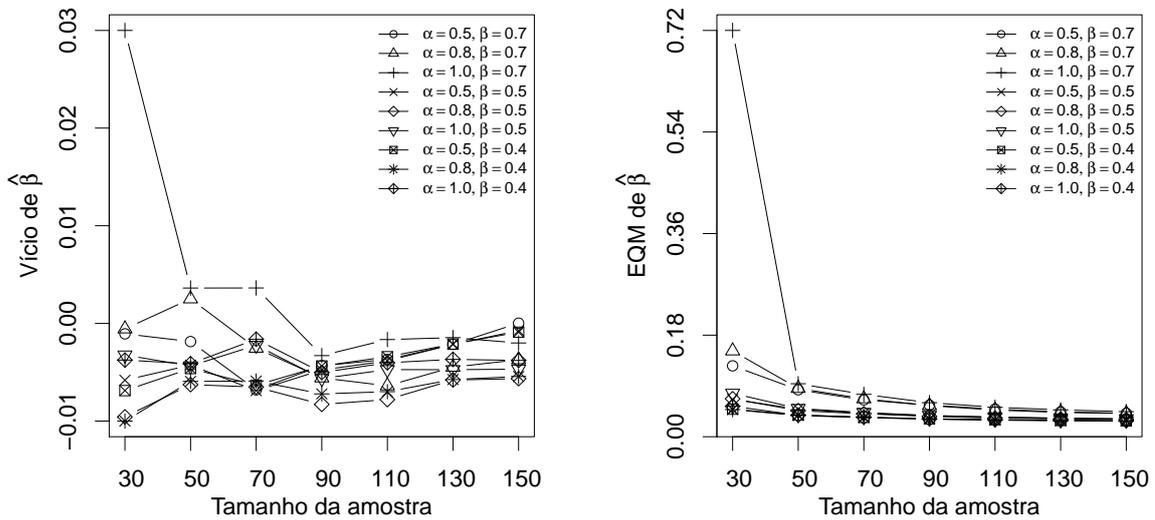


Figura 32 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta truncada em zero.

em todos os cenários. O mesmo não ocorre com o vício de $\hat{\beta}$ que se apresenta negativo em alguns cenários.

Já nas Figuras 31 e 32, assim como nas distribuições anteriores, pode-se notar que a magnitude do vício quanto do EQM tem tendência exponencial convergente para 0 quando o tamanho amostral aumenta.

3.3.4 Distribuição Lindley Ponderada Discreta Truncada em Zero

Considerando as mesmas condições das propostas dos modelos truncados anteriores, propomos a versão truncada em zero da distribuição Lindley ponderada discreta com função de probabilidade escrita na forma:

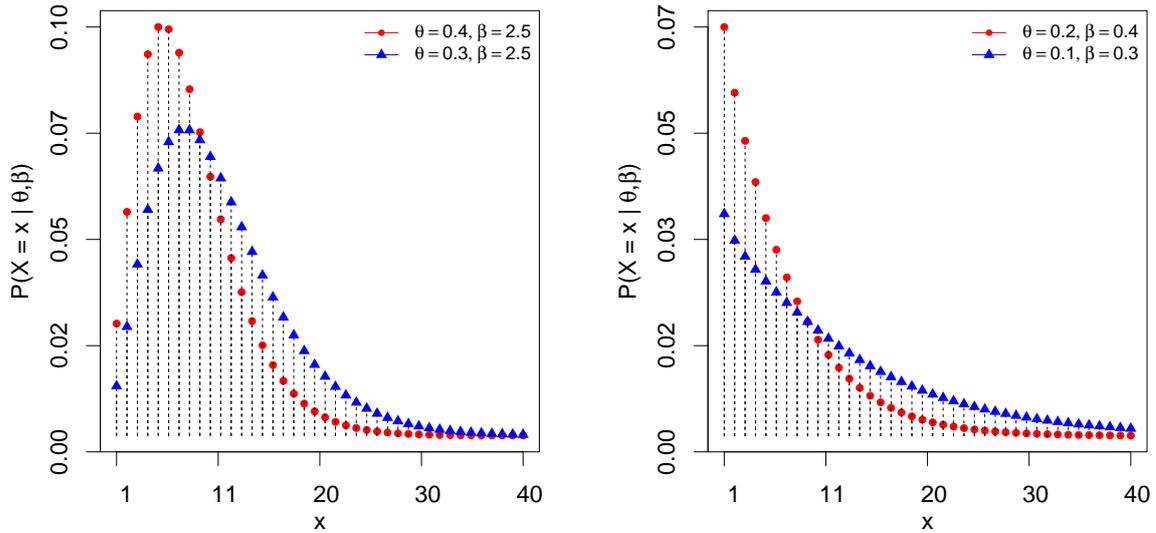


Figura 33 – Comportamento da função de probabilidade da distribuição Lindley ponderada discreta truncada em zero para diferentes valores de θ e β .

$$P(X = x | \theta, \beta) = \frac{(\theta + \beta)[\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)]}{(\theta + \beta)\Gamma(\beta, \theta) + \theta^\beta e^{-\theta}} + \frac{\theta^\beta e^{-\theta x} [x^\beta - (x + 1)^\beta e^{-\theta}]}{(\theta + \beta)\Gamma(\beta, \theta) + \theta^\beta e^{-\theta}}, \quad (3.84)$$

em que $\theta, \beta > 0$. É importante destacar que a expressão (3.84) satisfaz a inequação de log-concavidade (KEILSON; GERBER, 1971) e, portanto, seu comportamento é unimodal. Tal comportamento é ilustrado na Figura 33.

A função de distribuição acumulada, de sobrevivência e risco correspondentes são, respectivamente, dadas por:

$$F(x | \theta, \beta) = \frac{[e^{\theta x + \theta} - (x + 1)^\beta e^\theta] \theta^\beta}{[\Gamma(\beta, \theta)(\theta + \beta)e^\theta + \theta^\beta] e^{\theta x + \theta}} + \frac{[\Gamma(\beta, \theta) - \Gamma(\beta, \theta x + \theta)](\theta + \beta) e^\theta}{\Gamma(\beta, \theta)(\theta + \beta)e^\theta + \theta^\beta} \quad (3.85)$$

$$S(x | \theta, \beta) = 1 - \left[\frac{[e^{\theta x + \theta} - (x + 1)^\beta e^\theta] \theta^\beta}{[\Gamma(\beta, \theta)(\theta + \beta)e^\theta + \theta^\beta] e^{\theta x + \theta}} + \frac{[\Gamma(\beta, \theta) - \Gamma(\beta, \theta x + \theta)](\theta + \beta) e^\theta}{\Gamma(\beta, \theta)(\theta + \beta)e^\theta + \theta^\beta} \right] \quad (3.86)$$

e,

$$h(x | \theta, \beta) = \frac{\frac{(\theta + \beta)[\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)]}{(\theta + \beta)\Gamma(\beta, \theta) + \theta^\beta e^{-\theta}} + \frac{\theta^\beta e^{-\theta x} [x^\beta - (x + 1)^\beta e^{-\theta}]}{(\theta + \beta)\Gamma(\beta, \theta) + \theta^\beta e^{-\theta}}}{1 - \left[\frac{[e^{\theta x + \theta} - (x + 1)^\beta e^\theta] \theta^\beta}{[\Gamma(\beta, \theta)(\theta + \beta)e^\theta + \theta^\beta] e^{\theta x + \theta}} + \frac{[\Gamma(\beta, \theta) - \Gamma(\beta, \theta x + \theta)](\theta + \beta) e^\theta}{\Gamma(\beta, \theta)(\theta + \beta)e^\theta + \theta^\beta} \right]} \quad (3.87)$$

O comportamento da função de risco é ilustrado na Figura 34. Pode notar que o mesmo é decrescente para alguns valores e decrescente-crescente para outros valores.

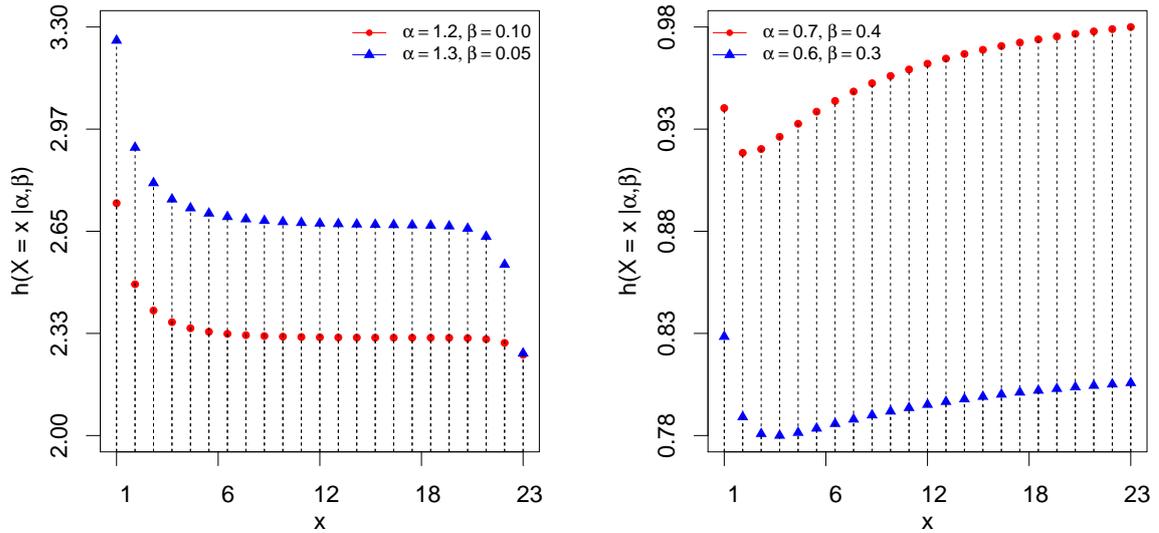


Figura 34 – Comportamento da função de risco da distribuição Lindley ponderada discreta truncada em zero para diferentes valores de θ e β .

Partindo de (3.84) e (3.85), observa-se que a distribuição Lindley ponderada discreta truncada em zero não possui forma fechada tanto para os momentos quanto para a função quantil. Logo, para avaliar ambos, faz-se necessário o uso de métodos numéricos.

Seja x_1, \dots, x_n uma amostra aleatória de tamanho n da distribuição Lindley ponderada discreta truncada em zero com parâmetros θ, β . O logaritmo da função de verossimilhança é dado por:

$$\ell(\theta, \beta | \mathbf{x}) = \sum_{i=1}^n \log \left[\frac{(\theta + \beta)[\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)]}{(\theta + \beta)\Gamma(\beta, \theta) + \theta^\beta e^{-\theta}} + \frac{\theta^\beta e^{-\theta x} [x^\beta - (x + 1)^\beta e^{-\theta}]}{(\theta + \beta)\Gamma(\beta, \theta) + \theta^\beta e^{-\theta}} \right]$$

Considerando a log-verossimilhança acima, é possível verificar que, assim como os momentos, os estimadores de máxima verossimilhança de θ e β para esta distribuição não possuem forma fechada sendo necessário também o uso de métodos numéricos para a avaliação das propriedades dos mesmos. Além disso, para esta distribuição, não foi realizado um estudo de simulação Monte Carlo devido a problemas de instabilidade computacional e convergência do método numérico para os estimadores de máxima verossimilhança em questão.

3.4 Modelos Discretos Deslocados da Origem

Além do truncamento em zero, pode-se considerar o deslocamento da origem. A versão deslocada é simplesmente uma transformação de uma variável aleatória discreta com suporte em $0, 1, \dots$ em uma variável aleatória discreta estritamente positiva. Isto é, dada uma variável aleatória discreta X , pode-se definir a versão deslocada de X como $Y = X + 1$ e então sua função de probabilidade é definida por:

$$P(Y = y) = P(X = y - 1). \tag{3.88}$$

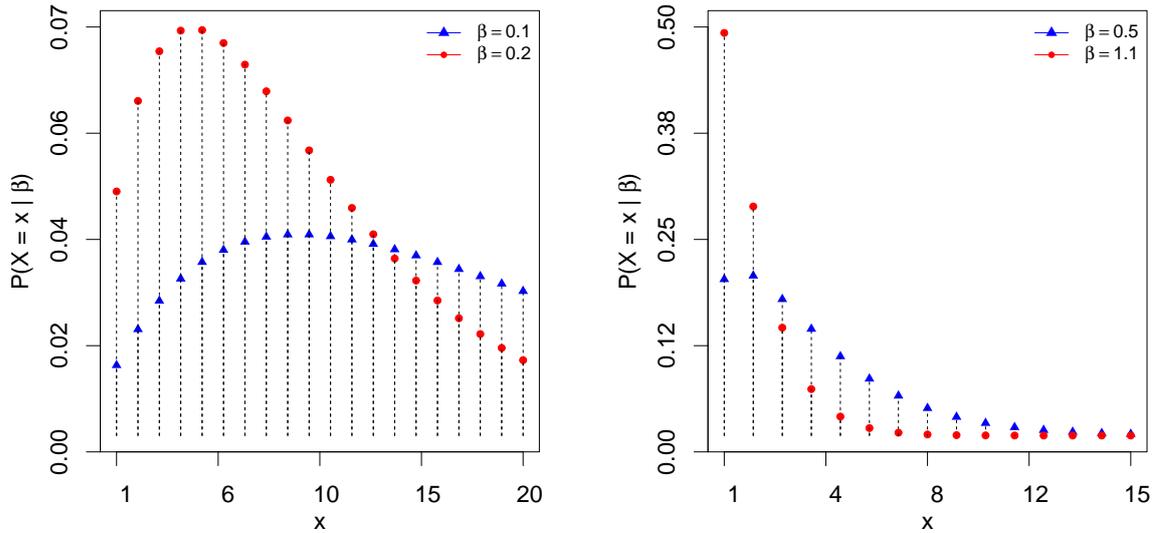


Figura 35 – Comportamento da função de probabilidade da distribuição Lindley discreta deslocada da origem (via função de sobrevivência) para alguns valores de β .

3.4.1 Distribuição Lindley Discreta Deslocada da Origem (via Função de Sobrevivência)

Como uma alternativa ao truncamento em zero, de (3.4), propomos a versão deslocada da origem da distribuição Lindley discreta com função de probabilidade dada por:

$$P(X = x | \beta) = \frac{e^{-\beta(x-1)}}{1 + \beta} \left[\beta(1 - 2e^{-\beta}) + (1 - e^{-\beta})(1 + \beta x - \beta) \right] \quad (3.89)$$

em que $x = 1, 2, \dots$ e $\beta > 0$.

Por meio da equação (3.89), é possível verificar que a mesma satisfaz a equação de log-concavidade, $[P(X = x)]^2 \leq P(X = x - 1)P(X = x + 1)$, e, pelo Teorema 3 de Keilson e Gerber (1971), é unimodal. O comportamento de (3.89) é apresentado na Figura 35.

A função de distribuição acumulada, a função de sobrevivência e a função de risco da distribuição Lindley discreta deslocada da origem são dadas, respectivamente, por:

$$F(x | \beta) = 1 - \frac{e^{-\beta(x+1)}(1 + \beta + \beta x)}{e^{-\beta}(\beta + 1)} \quad (3.90)$$

$$S(x | \beta) = \frac{e^{-\beta(x+1)}(1 + \beta + \beta x)}{e^{-\beta}(\beta + 1)}. \quad (3.91)$$

e,

$$h(x | \beta) = \frac{e^{-\beta} \left[\beta(1 - 2e^{-\beta}) + (1 - e^{-\beta})(1 + \beta x - \beta) \right]}{1 + \beta + \beta x}. \quad (3.92)$$

em que $x = 1, 2, \dots$ e $\beta > 0$. O comportamento da função de risco é ilustrado na Figura 36 para diferentes valores de β .

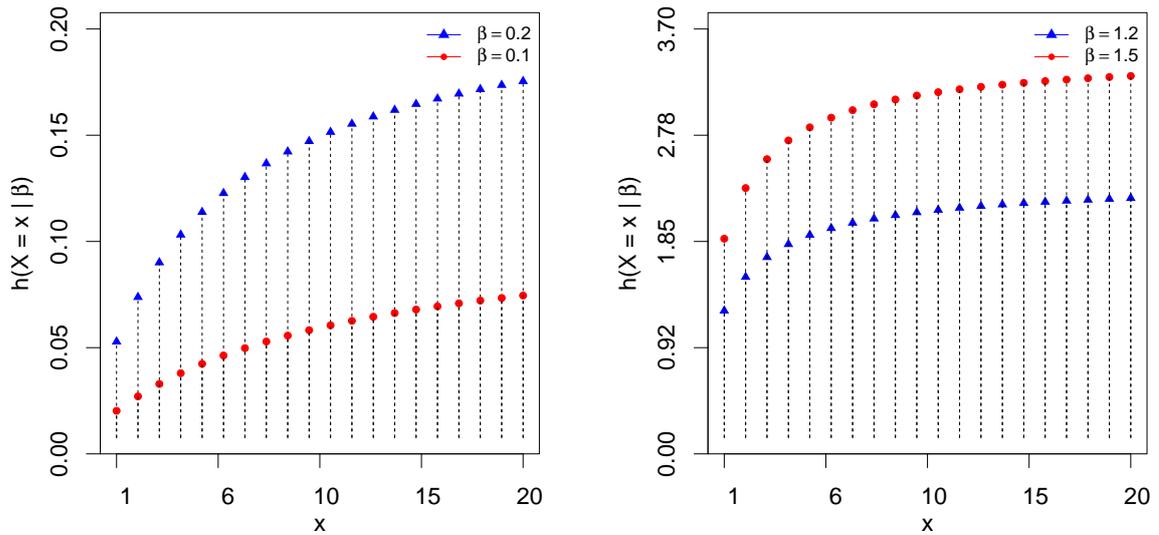


Figura 36 – Comportamento da função de risco da distribuição Lindley discreta deslocada da origem (via função de sobrevivência) para alguns valores de β .

Quanto aos momentos, tem-se que a esperança e a variância são dadas, respectivamente, por:

$$\mathbb{E}(X) = \frac{\beta - e^{-\beta} + 1}{(e^{-\beta} - 1)^2 (\beta + 1)}, \tag{3.93}$$

e,

$$\text{Var}(X) = \frac{e^{-\beta} [(\beta + 1) e^{-\beta^2} + (-3\beta^2 - 4\beta - 2) e^{-\beta} + 2\beta^2 + 3\beta + 1]}{(e^{-\beta} - 1)^4 (\beta + 1)^2}. \tag{3.94}$$

De maneira análoga as distribuições anteriores, avaliando a razão entre $\mathbb{E}(X)$ e $\text{Var}(X)$, facilmente verifica-se, numericamente, que $\mathbb{E}(X) < \text{Var}(X)$ para $\beta < 0.8755$. Para $\beta > 0.8755$ tem-se $\mathbb{E}(X) > \text{Var}(X)$ (Figura 37). Assim, como na versão truncada, a versão deslocada também pode ser utilizada na análise de dados com superdispersão ou subdispersão.

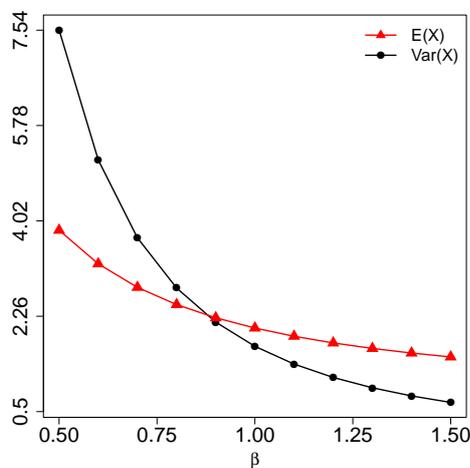


Figura 37 – Esperança e variância distribuição Lindley discreta deslocada da origem (via função de sobrevivência) para alguns valores de β .

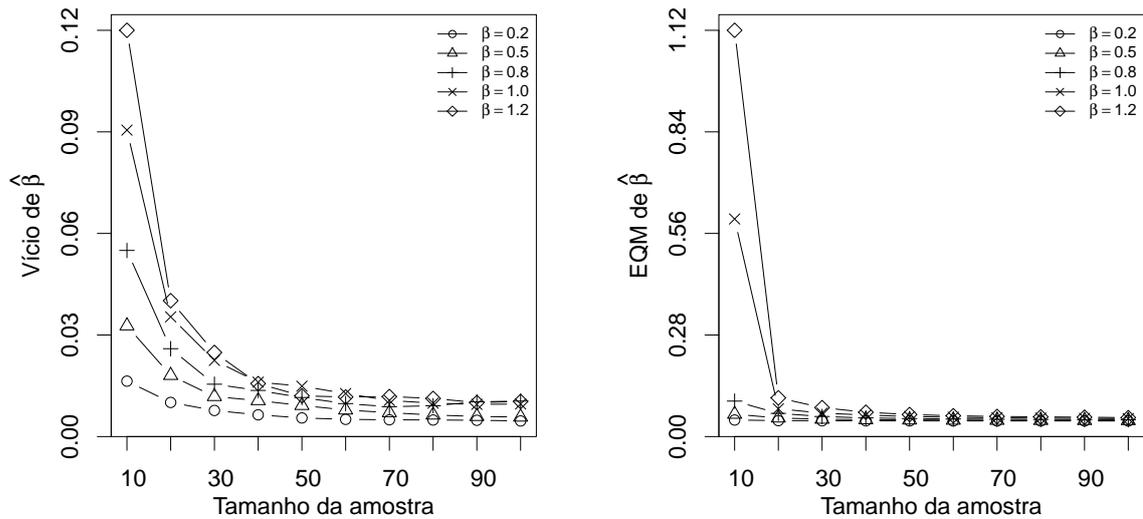


Figura 38 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta deslocada da origem (via função de sobrevivência).

Supondo uma amostra aleatória x_1, \dots, x_n de (3.89), o estimador do parâmetro de dispersão β é obtido resolvendo-se $\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = 0$ em β em que:

$$\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = n - n\bar{x} - \frac{n}{1 + \beta} + \sum_{i=1}^n \frac{(2 - 3\beta + (1 - \beta)x_i) \exp(-\beta)}{1 + \beta x_i - (1 + 3\beta + \beta x_i) \exp(-\beta)}. \quad (3.95)$$

Note que a equação acima é não-linear e deve ser resolvida numericamente. Assim, com a finalidade de estudar algumas propriedades desse estimador, realizamos um estudo de simulação Monte Carlo. Os cenários utilizados no estudo foram os mesmos considerados nas distribuições uni-paramétricas anteriores e os estão dispostos na Tabela 13.

Tabela 13 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta deslocada da origem (via função de sobrevivência).

n	Valores de β					Valores de β									
	0.2	0.5	0.8	1.0	1.2	10	20	30	40	50	60	70	80	90	100
Vício	0.0137	0.0318	0.0540	0.0910	0.1218	0.0028	0.0190	0.0569	0.5768	1.1153					
	0.0072	0.0151	0.0237	0.0335	0.0385	0.0012	0.0083	0.0218	0.0350	0.0662					
	0.0047	0.0081	0.0128	0.0201	0.0225	0.0007	0.0050	0.0130	0.0226	0.0382					
	0.0035	0.0072	0.0108	0.0134	0.0129	0.0006	0.0037	0.0090	0.0172	0.0251					
	0.0025	0.0059	0.0086	0.0122	0.0092	0.0004	0.0030	0.0071	0.0129	0.0186					
	0.0020	0.0046	0.0068	0.0099	0.0088	0.0004	0.0025	0.0058	0.0108	0.0150					
	0.0016	0.0038	0.0058	0.0077	0.0090	0.0003	0.0021	0.0051	0.0092	0.0131					
	0.0016	0.0030	0.0061	0.0069	0.0084	0.0003	0.0018	0.0045	0.0083	0.0119					
	0.0015	0.0027	0.0073	0.0065	0.0071	0.0002	0.0016	0.0040	0.0075	0.0110					
	0.0013	0.0025	0.0076	0.0067	0.0075	0.0002	0.0014	0.0037	0.0068	0.0096					
EQM															

De acordo com a Tabela 13, concluímos que tanto o vício e quanto o EQM de $\hat{\beta}$ são positivos. Além disso, pode-se notar também que a magnitude tanto do vício e quanto do EQM de $\hat{\beta}$ sempre decaí para 0 quando o tamanho da amostra aumenta e apresenta uma tendência de decaimento exponencial. A Figura 38 ilustra tais conclusões.

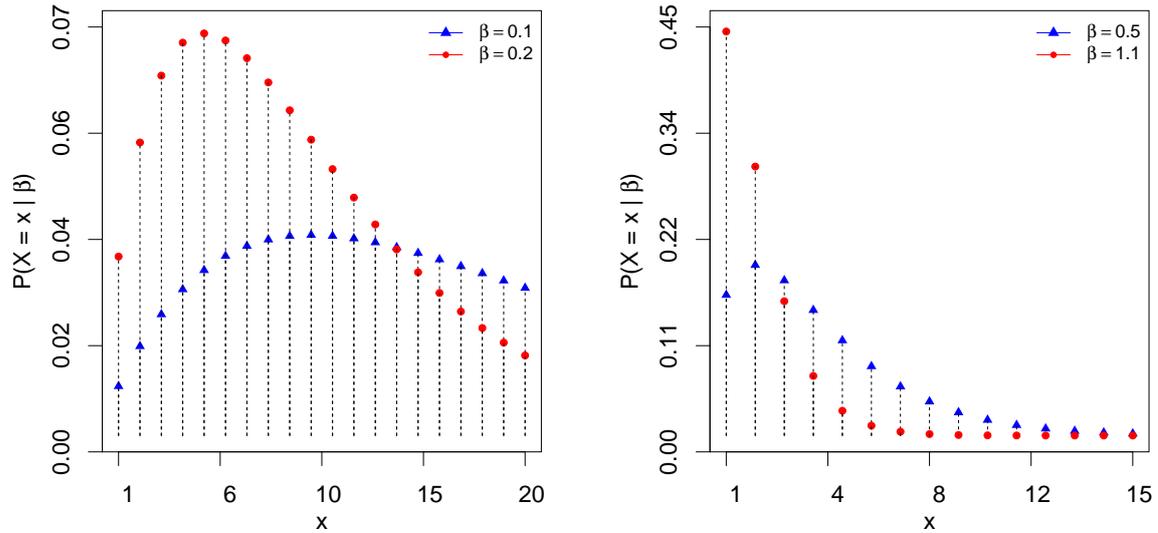


Figura 39 – Comportamento da função de probabilidade da distribuição Lindley discreta deslocada da origem (via série infinita) para diferentes valores de β .

3.4.2 Distribuição Lindley Discreta Deslocada da Origem (via Série Infinita)

Analogamente a Lindley discreta deslocada da origem (via função de sobrevivência), propomos a versão deslocada da distribuição Lindley discreta (via série infinita) com função de probabilidade escrita na forma:

$$P(X = x | \beta) = x e^{-\beta(1+x)} (e^{\beta} - 1)^2. \quad (3.96)$$

em que $x = 1, \dots$ e $\beta > 0$. O comportamento de (3.96) é unimodal e ilustrado na Figura 39 para alguns valores de β .

Neste caso, a função de distribuição acumulada e a função de sobrevivência são dadas, respectivamente, por:

$$F(x | \beta) = x e^{-\beta(x+1)} - e^{-\beta x}(x + 1) + 1, \quad (3.97)$$

$$S(x | \beta) = x e^{-\beta(x+1)} - e^{-\beta x}(x + 1), \quad (3.98)$$

e,

$$h(x | \beta) = \frac{x e^{-\beta(1+x)} (e^{\beta} - 1)^2}{x e^{-\beta(x+1)} - e^{-\beta x}(x + 1)}. \quad (3.99)$$

em que $x = 1, 2, \dots$ e $\beta > 0$. O comportamento da função de risco é crescente e ilustrado na Figura 40 para diferentes valores de β .

De maneira análoga as distribuições anteriores, por meio de (3.96), tem-se que a esperança e a variância da distribuição Lindley discreta deslocada da origem (via série infinita) são dadas, respectivamente, por:

$$\mathbb{E}(X) = \frac{1 + e^{\beta}}{e^{\beta} - 1}, \quad (3.100)$$

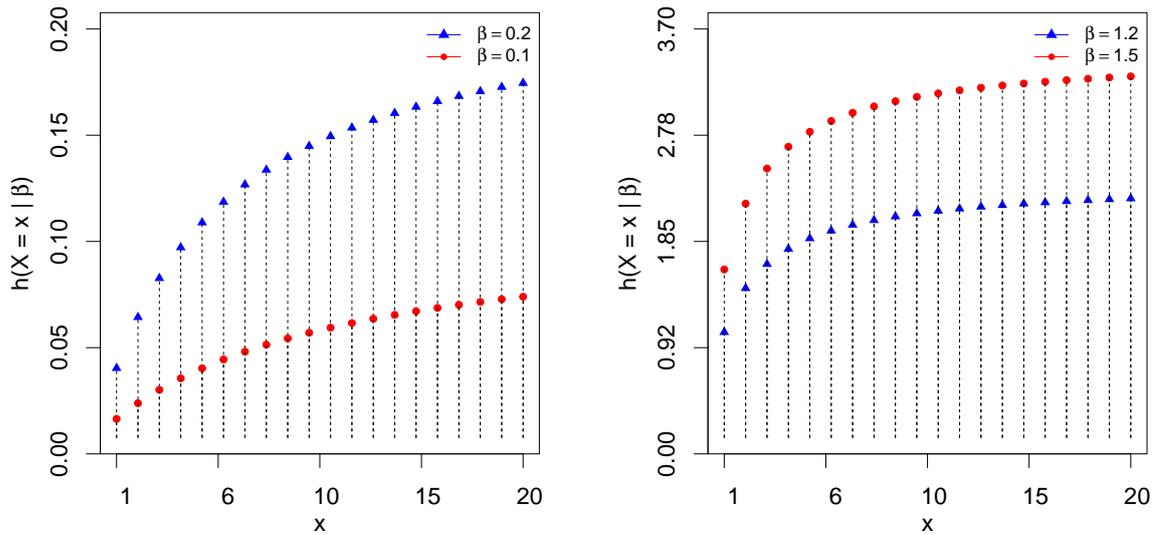


Figura 40 – Comportamento da função de risco da distribuição Lindley discreta deslocada da origem (via série infinita) para alguns valores de β .

e,

$$\text{Var}(X) = -\frac{(1 + e^\beta)^2}{(e^\beta - 1)^2} + \frac{e^{2\beta} + 4e^\beta + 1}{-2e^\beta + e^{2\beta} + 1}. \tag{3.101}$$

Além disso, avaliando numericamente a razão entre $\mathbb{E}(X)$ e $\text{Var}(X)$, tem-se que $\mathbb{E}(X) < \text{Var}(X)$ para $\beta < 0.8814$ (Figura 41). Assim, conclui-se que esta versão comporta dados com subdispersão e superdispersão.

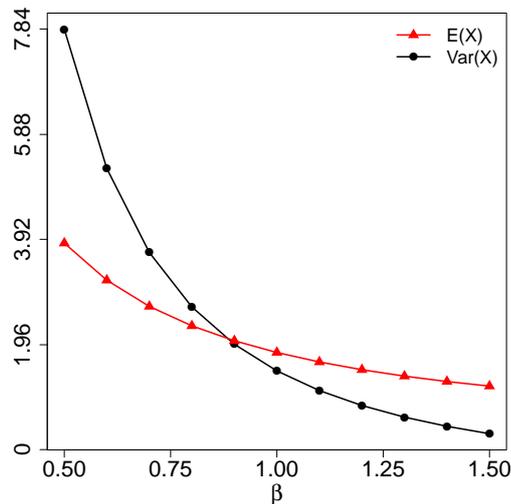


Figura 41 – Esperança e variância distribuição Lindley discreta deslocada da origem (via série infinita) para alguns valores de β .

Dada x_1, \dots, x_n uma amostra aleatória iid da distribuição Lindley discreta deslocada da origem (via série infinita), o logaritmo da função de verossimilhança é definido por:

$$\ell(\beta | \mathbf{x}) \propto 2n \log(e^\beta - 1) - \beta(n + n\bar{x}). \tag{3.102}$$

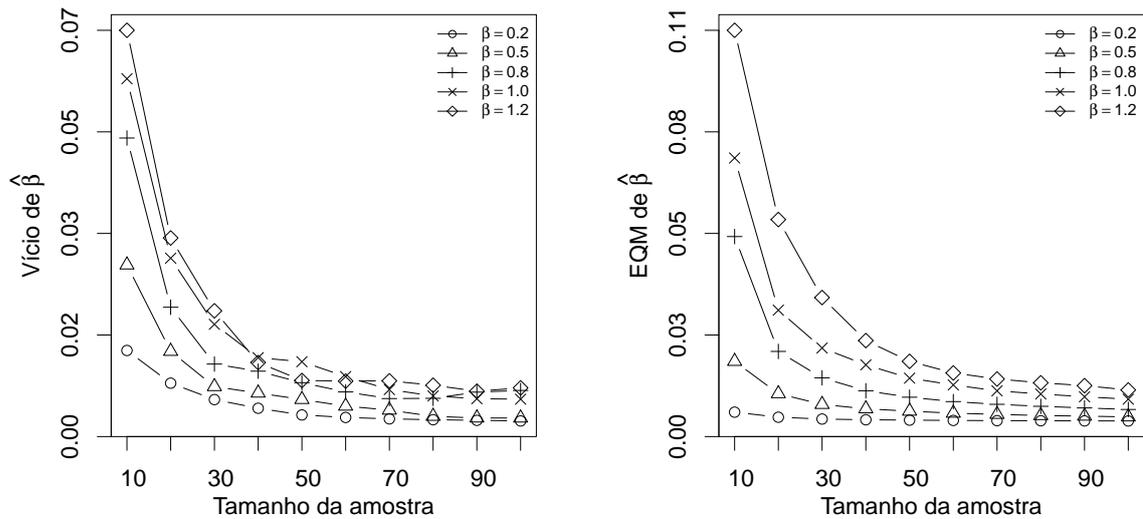


Figura 42 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta deslocada da origem (via série infinita).

Logo, o estimador de máxima verossimilhança de β é escrito na forma:

$$\hat{\beta} = \log \left(\frac{1 + \bar{x}}{\bar{x} - 1} \right). \tag{3.103}$$

Na Tabela 14 são apresentados os resultados do estudo de simulação Monte Carlo realizado para avaliar as propriedades de $\hat{\beta}$ considerando os mesmos cenários das distribuições uni-paramétricas anteriores.

Tabela 14 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta deslocada da origem (via série infinita).

n	Valores de β					n	Valores de β						
	0.2	0.5	0.8	1.0	1.2		0.2	0.5	0.8	1.0	1.2		
Vício	10	0.0134	0.0278	0.0499	0.0600	0.0683	EQM	10	0.0027	0.0167	0.0507	0.0722	0.1072
	20	0.0071	0.0134	0.0210	0.0294	0.0328		20	0.0012	0.0077	0.0192	0.0306	0.0554
	30	0.0046	0.0072	0.0113	0.0181	0.0204		30	0.0007	0.0047	0.0120	0.0202	0.0340
	40	0.0035	0.0064	0.0101	0.0124	0.0116		40	0.0006	0.0035	0.0085	0.0156	0.0222
	50	0.0025	0.0052	0.0082	0.0117	0.0085		50	0.0004	0.0028	0.0067	0.0119	0.0165
	60	0.0020	0.0041	0.0066	0.0092	0.0084		60	0.0004	0.0023	0.0055	0.0101	0.0134
	70	0.0016	0.0033	0.0055	0.0070	0.0085		70	0.0003	0.0020	0.0048	0.0085	0.0117
	80	0.0016	0.0026	0.0055	0.0059	0.0077		80	0.0003	0.0017	0.0042	0.0076	0.0107
	90	0.0015	0.0022	0.0066	0.0054	0.0067		90	0.0002	0.0015	0.0038	0.0068	0.0099
	100	0.0013	0.0022	0.0068	0.0054	0.0073		100	0.0002	0.0013	0.0033	0.0062	0.0087

De maneira análoga aos estudos anteriores, conclui-se que a magnitude tanto do vício quanto do EQM de $\hat{\beta}$ sempre decaí para 0 quando o tamanho da amostra aumenta e a mesma tende a ser exponencial. A Figura 42 ilustra tal resultado.

3.4.3 Distribuição Lindley Potência Discreta Deslocada da Origem

Nesta seção, propomos a versão deslocada da origem da distribuição Lindley potência discreta com função de probabilidade definida por:

$$P(X = x | \alpha, \beta) = \left(1 + \frac{\beta(x-1)^\alpha}{\beta+1} \right) \gamma^{(x-1)^\alpha} - \left(1 + \frac{\beta x^\alpha}{\beta+1} \right) \gamma^{x^\alpha}, \tag{3.104}$$

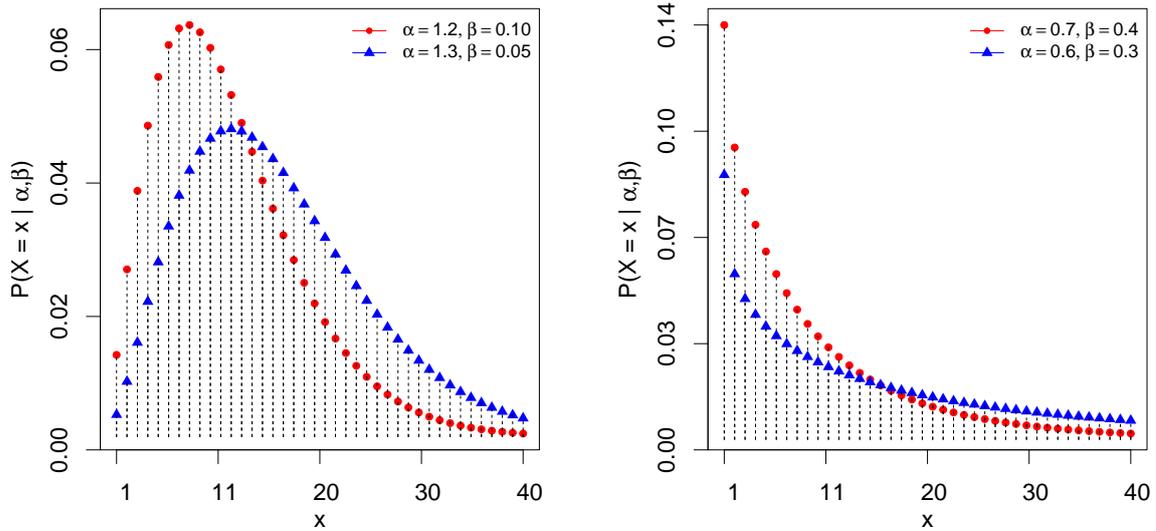


Figura 43 – Comportamento da função de probabilidade da distribuição Lindley potência deslocada da origem para diferentes valores de α e β .

em que $\gamma = \exp(-\beta)$. Na Figura 43 é ilustrado o comportamento de (3.104) e pode-se observar que o mesmo é unimodal.

A função de distribuição acumulada, a função de sobrevivência e a função de risco da distribuição Lindley potência deslocada da origem são dadas, respectivamente, por:

$$F(x | \alpha, \beta) = 1 - \left(1 + \frac{\beta x^\alpha}{\beta + 1}\right) \gamma^{x^\alpha}, \quad (3.105)$$

$$S(x | \alpha, \beta) = \left(1 + \frac{\beta x^\alpha}{\beta + 1}\right) \gamma^{x^\alpha}, \quad (3.106)$$

e,

$$h(x | \beta) = \frac{1 + \beta + \beta(x - 1)^\alpha}{1 + \beta + \beta x^\alpha} \gamma^{(x-1)^\alpha - x^\alpha} - 1 \quad (3.107)$$

em que $\gamma = \exp(-\beta)$. O comportamento da função de risco é ilustrado na Figura 44 para diferentes valores de α e β .

O momento de ordem k é descrito por:

$$\begin{aligned} \mathbb{E}(X^k) &= \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^k \frac{[-\beta(x-1)^\alpha]^j}{j!} - \sum_{x=0}^{\infty} x^k \gamma^{x^\alpha} \\ &+ \frac{\beta}{\beta + 1} \left[\sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^k (x-1)^\alpha \frac{[-\beta(x-1)^\alpha]^j}{j!} - \sum_{x=0}^{\infty} x^{k+\alpha} \gamma^{x^\alpha} \right]. \end{aligned} \quad (3.108)$$

De (3.108), obtém-se que a esperança e a variância para essa distribuição são dadas, respectivamente, por:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x \frac{[-\beta(x-1)^\alpha]^j}{j!} - \sum_{x=0}^{\infty} x \gamma^{x^\alpha} \\ &+ \frac{\beta}{\beta + 1} \left[\sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x(x-1)^\alpha \frac{[-\beta(x-1)^\alpha]^j}{j!} - \sum_{x=0}^{\infty} x^{1+\alpha} \gamma^{x^\alpha} \right] \end{aligned} \quad (3.109)$$

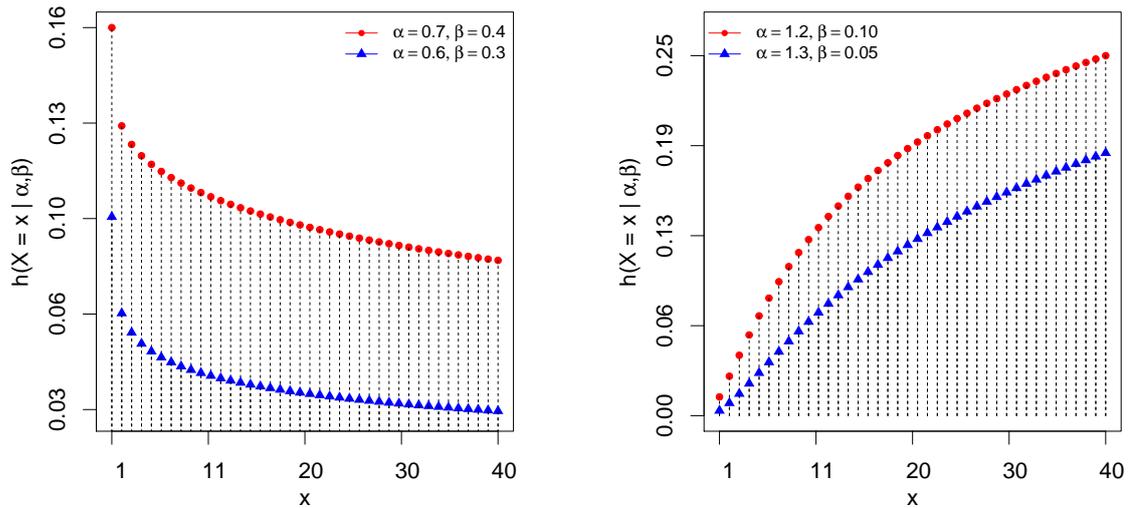


Figura 44 – Comportamento da função de risco da distribuição Lindley potência discreta deslocada da origem para diferentes valores de α e β .

$$\begin{aligned}
 \text{Var}(X) &= \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^2 \frac{[-\beta(x-1)^\alpha]^j}{j!} - \sum_{x=0}^{\infty} x^2 \gamma^{x^\alpha} \\
 &+ \frac{\beta}{\beta+1} \left[\sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x^2 (x-1)^\alpha \frac{[-\beta(x-1)^\alpha]^j}{j!} - \sum_{x=0}^{\infty} x^{2+\alpha} \gamma^{x^\alpha} \right] \\
 &- \left\{ \sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x \frac{[-\beta(x-1)^\alpha]^j}{j!} - \sum_{x=0}^{\infty} x \gamma^{x^\alpha} \right. \\
 &\left. + \frac{\beta}{\beta+1} \left[\sum_{x=0}^{\infty} \sum_{j=0}^{\infty} x(x-1)^\alpha \frac{[-\beta(x-1)^\alpha]^j}{j!} - \sum_{x=0}^{\infty} x^{1+\alpha} \gamma^{x^\alpha} \right] \right\}^2 \quad (3.110)
 \end{aligned}$$

Uma ilustração do comportamento da esperança e variância descritas acima pode ser vista na Figura 45. Pode-se notar que essa distribuição comporta também a subdispersão quanto a superdispersão.

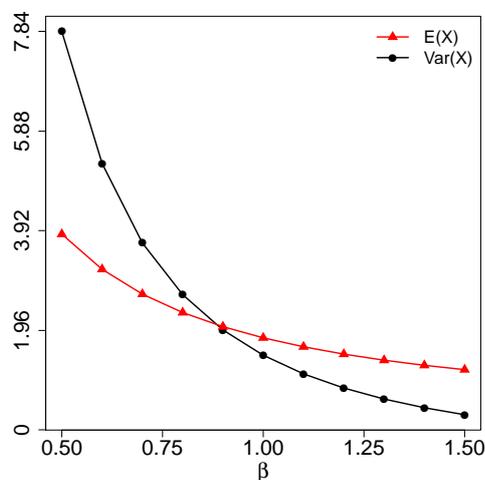


Figura 45 – Esperança e variância distribuição Lindley potência discreta deslocada da origem para alguns valores de α e $\beta = 1$.

Dada uma amostra aleatória x_1, \dots, x_n de tamanho n da distribuição Lindley potência discreta deslocada da origem com parâmetros α, β e função de probabilidade dada por (3.104), o logaritmo da função de verossimilhança é dado por:

$$\ell(\alpha, \beta | \mathbf{x}) = \sum_{i=1}^n \log \left(\left(1 + \frac{\beta(x-1)^\alpha}{\beta+1} \right) \gamma^{(x-1)^\alpha} - \left(1 + \frac{\beta x^\alpha}{\beta+1} \right) \gamma^{x^\alpha} \right).$$

Os estimadores de máxima verossimilhança de α e β são obtidos resolvendo-se a equação $U(\alpha, \beta | \mathbf{x}) = \left(\frac{d}{d\alpha} \ell(\alpha, \beta | \mathbf{x}), \frac{d}{d\beta} \ell(\alpha, \beta | \mathbf{x}) \right) = 0$ em α e β . Entretanto, o sistema gerado em $U(\alpha, \beta | \mathbf{x})$ não possui solução analítica para α e nem para β e isso implica que os estimadores de máxima verossimilhança dessa distribuição não possuem forma fechada. Logo, métodos numéricos devem ser aplicados com a finalidade de se encontrar uma solução para $U(\alpha, \beta | \mathbf{x})$.

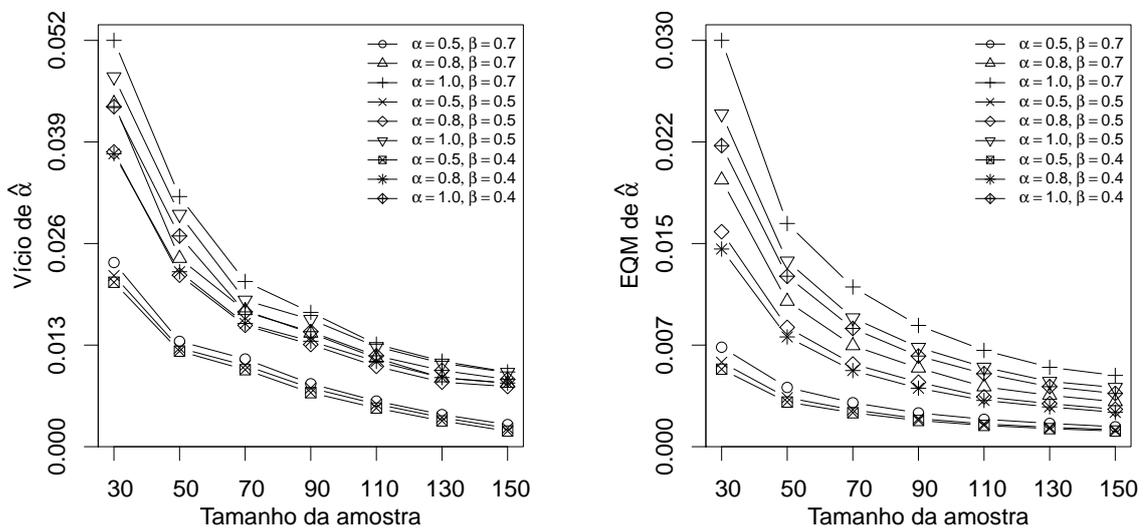


Figura 46 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta deslocada da origem.

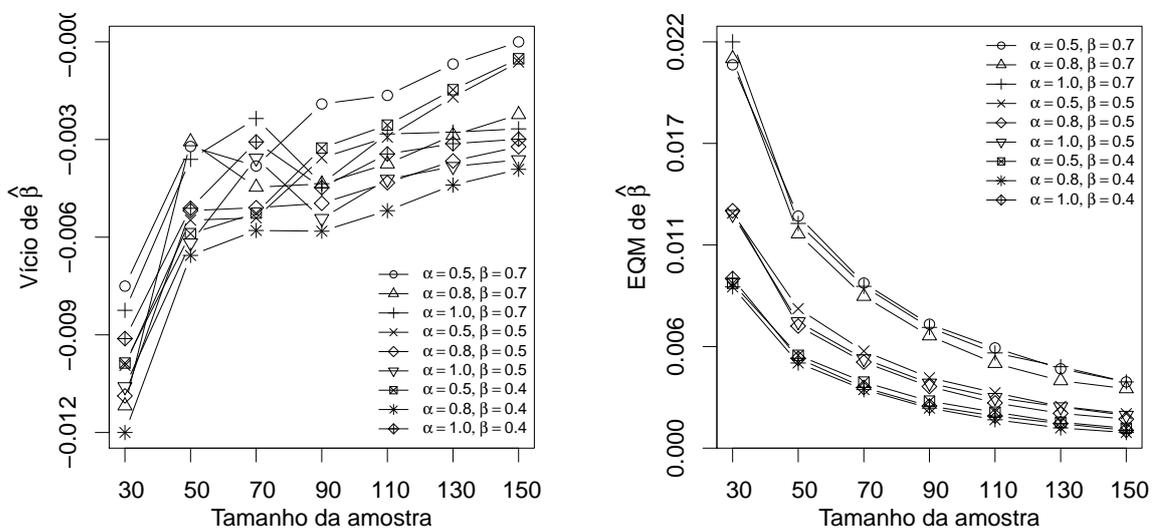


Figura 47 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta deslocada da origem.

De maneira análoga ao que foi feito para as distribuições anteriores, realizamos um estudo de simulação Monte Carlo para avaliar algumas propriedades, como o vício e o EQM, de $\hat{\alpha}$ e $\hat{\beta}$. Para este estudo de simulação, consideramos como cenários todas combinações de (α, β) , em que $\alpha = 0.5, 0.8, 1.0$ e $\beta = 0.4, 0.5, 0.7$, e tamanhos de amostra variando entre 30 e 150 com passo $n = 20$. Os resultados deste estudo são explícitos nas Tabelas 15 e 16 e ilustrados nas Figuras 46 e 47.

Tabela 15 – Vício e EQM de $\hat{\alpha}$ da distribuição Lindley potência discreta deslocada da origem.

n	$\beta = 0.7$			$\beta = 0.5$			$\beta = 0.4$			
	α			α			α			
	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0	
Vício	30	0.0246	0.0439	0.0516	0.0230	0.0380	0.0471	0.0222	0.0378	0.0435
	50	0.0150	0.0251	0.0326	0.0142	0.0230	0.0304	0.0138	0.0235	0.0278
	70	0.0129	0.0187	0.0223	0.0121	0.0169	0.0200	0.0116	0.0172	0.0186
	90	0.0098	0.0160	0.0185	0.0093	0.0146	0.0177	0.0088	0.0150	0.0162
	110	0.0077	0.0130	0.0147	0.0074	0.0120	0.0144	0.0069	0.0126	0.0132
	130	0.0061	0.0106	0.0126	0.0057	0.0100	0.0123	0.0053	0.0106	0.0115
	150	0.0049	0.0099	0.0113	0.0044	0.0095	0.0112	0.0041	0.0099	0.0104
EQM	30	0.0071	0.0193	0.0296	0.0060	0.0155	0.0242	0.0054	0.0143	0.0218
	50	0.0041	0.0104	0.0161	0.0034	0.0085	0.0134	0.0030	0.0078	0.0123
	70	0.0030	0.0072	0.0115	0.0025	0.0058	0.0092	0.0023	0.0053	0.0084
	90	0.0022	0.0055	0.0087	0.0018	0.0045	0.0070	0.0017	0.0041	0.0064
	110	0.0018	0.0042	0.0068	0.0014	0.0034	0.0056	0.0013	0.0032	0.0051
	130	0.0015	0.0035	0.0056	0.0012	0.0029	0.0045	0.0011	0.0027	0.0042
	150	0.0012	0.0031	0.0050	0.0010	0.0025	0.0041	0.0009	0.0023	0.0037

Tabela 16 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley potência discreta deslocada da origem.

n	$\beta = 0.7$			$\beta = 0.5$			$\beta = 0.4$			
	α			α			α			
	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0	
Vício	30	-0.0075	-0.0109	-0.0082	-0.0097	-0.0106	-0.0103	-0.0097	-0.0116	-0.0090
	50	-0.0035	-0.0033	-0.0038	-0.0056	-0.0053	-0.0062	-0.0060	-0.0066	-0.0052
	70	-0.0040	-0.0046	-0.0027	-0.0055	-0.0052	-0.0038	-0.0054	-0.0059	-0.0033
	90	-0.0022	-0.0045	-0.0045	-0.0038	-0.0051	-0.0055	-0.0035	-0.0059	-0.0046
	110	-0.0020	-0.0040	-0.0031	-0.0032	-0.0045	-0.0044	-0.0029	-0.0053	-0.0037
	130	-0.0011	-0.0031	-0.0031	-0.0021	-0.0039	-0.0040	-0.0018	-0.0046	-0.0034
	150	-0.0005	-0.0026	-0.0030	-0.0011	-0.0035	-0.0038	-0.0010	-0.0041	-0.0033
EQM	30	0.0210	0.0214	0.0222	0.0133	0.0134	0.0132	0.0096	0.0094	0.0099
	50	0.0131	0.0122	0.0127	0.0083	0.0074	0.0076	0.0059	0.0055	0.0057
	70	0.0096	0.0089	0.0095	0.0061	0.0055	0.0057	0.0045	0.0041	0.0042
	90	0.0075	0.0069	0.0073	0.0047	0.0042	0.0044	0.0035	0.0031	0.0032
	110	0.0062	0.0054	0.0060	0.0039	0.0034	0.0037	0.0029	0.0025	0.0027
	130	0.0052	0.0045	0.0053	0.0032	0.0028	0.0032	0.0024	0.0021	0.0023
	150	0.0045	0.0041	0.0045	0.0028	0.0025	0.0027	0.0020	0.0018	0.0019

De acordo com as Tabelas 15 e 16, concluímos que a magnitude do vício e do EQM de $\hat{\alpha}$ e $\hat{\beta}$ convergem para zero quando o tamanho de amostra aumenta. Porém, o vício $\hat{\beta}$ é negativo em todos os cenários considerados. Além disso, os valores para o vício e para o EQM, independente do cenário, são baixos tanto para $\hat{\alpha}$ quanto para $\hat{\beta}$.

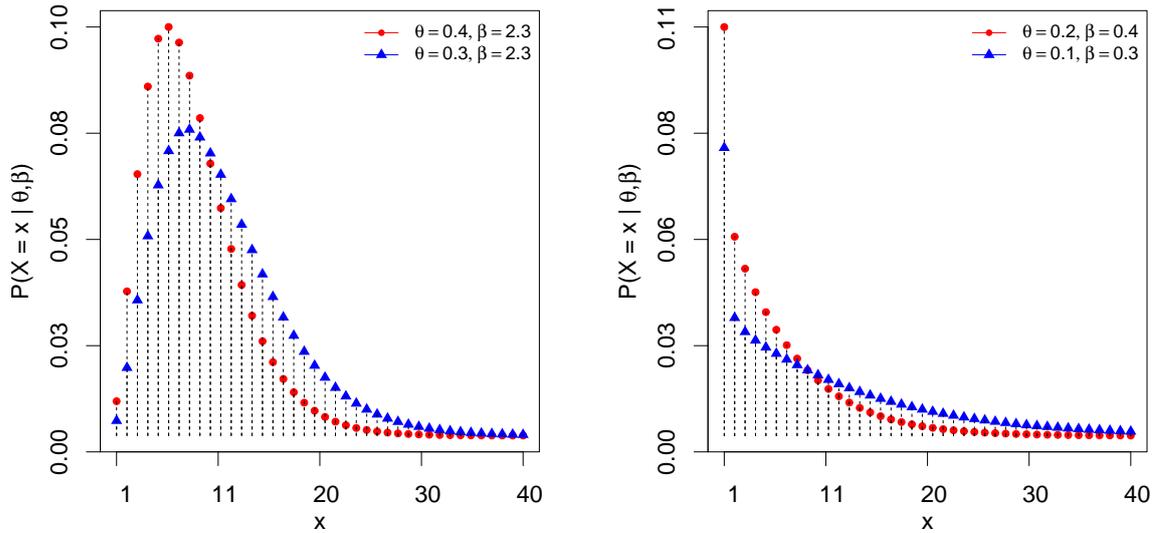


Figura 48 – Comportamento da função de probabilidade da distribuição Lindley ponderada discreta deslocada da origem para diferentes valores de θ e β .

Nas Figuras 46 e 47, pode-se notar uma tendência exponencial para a magnitude tanto do vício quanto do EQM. E também a convergência para zero conforme explícito nas Tabelas 15 e 16.

3.4.4 Distribuição Lindley Ponderada Deslocada da Origem

Por fim, nesta seção, propomos a versão deslocada da origem da distribuição Lindley ponderada discreta com função de probabilidade definida por:

$$P(X = x | \theta, \beta) = \frac{\Gamma(\beta, \theta(x - 1)) - \Gamma(\beta, \theta(x - 1) + \theta)}{\Gamma(\beta)} + \frac{\theta^\beta e^{-\theta(x-1)} [(x - 1)^\beta - x^\beta e^{-\theta}]}{(\theta + \beta)\Gamma(\beta)}. \quad (3.111)$$

em que $\theta, \beta > 0$. O comportamento de (3.111) é ilustrado na Figura 48 e pode-se observar que o mesmo é unimodal para os valores de θ e β considerados.

A função de distribuição acumulada, de sobrevivência e de risco correspondentes são, respectivamente, dadas por:

$$F(x | \theta, \beta) = 1 - \frac{(\theta + \beta)\Gamma(\beta, \theta x) + (\theta x)^\beta e^{-\theta x}}{(\theta + \beta)\Gamma(\beta)}, \quad (3.112)$$

$$S(x | \theta, \beta) = \frac{(\theta + \beta)\Gamma(\beta, \theta x) + (\theta x)^\beta e^{-\theta x}}{(\theta + \beta)\Gamma(\beta)}, \quad (3.113)$$

e,

$$h(x | \theta, \beta) = \frac{(\theta + \beta)\Gamma(\beta, \theta(x - 1)) + (\theta(x - 1))^\beta e^{-\theta(x-1)}}{(\theta + \beta)\Gamma(\beta, \theta x) + (\theta x)^\beta e^{-\theta x}} - 1. \quad (3.114)$$

Na Figura 49, ilustramos o comportamento da função de risco. Para os valores de θ e β considerados, nota-se uma evidência de comportamento decrescente-crescente.

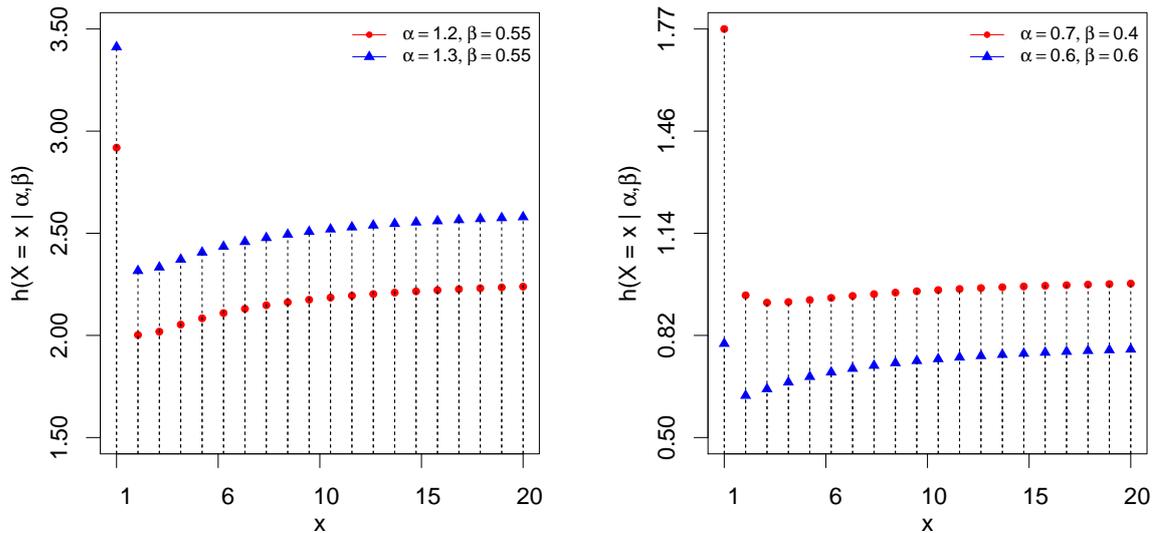


Figura 49 – Comportamento da função de risco da distribuição Lindley ponderada discreta deslocada da origem para diferentes valores de θ e β .

Dada x_1, \dots, x_n uma amostra aleatória de tamanho n da distribuição Lindley ponderada discreta deslocada da origem, temos que o logaritmo da função de verossimilhança é dado por:

$$\ell(\alpha, \beta | \mathbf{x}) = \sum_{i=1}^n \log \left[\frac{\Gamma(\beta, \theta(x-1)) - \Gamma(\beta, \theta(x-1) + \theta)}{\Gamma(\beta)} + \frac{\theta^\beta e^{-\theta(x-1)} [(x-1)^\beta - x^\beta e^{-\theta}]}{(\theta + \beta)\Gamma(\beta)} \right]$$

De (3.115), concluímos que os estimadores de máxima verossimilhança de θ e β não possuem forma analítica. Logo, por meio de um estudo de simulação Monte Carlo, avaliamos o vício e o EQM de $\hat{\theta}$ e $\hat{\beta}$ considerando os cenários de todas combinações de (θ, β) , em que $\theta = 0.5, 0.8, 1.0$ e $\beta = 0.5, 0.6, 0.7$, com tamanhos de amostra variando entre 30 e 150 com passo $n = 20$.

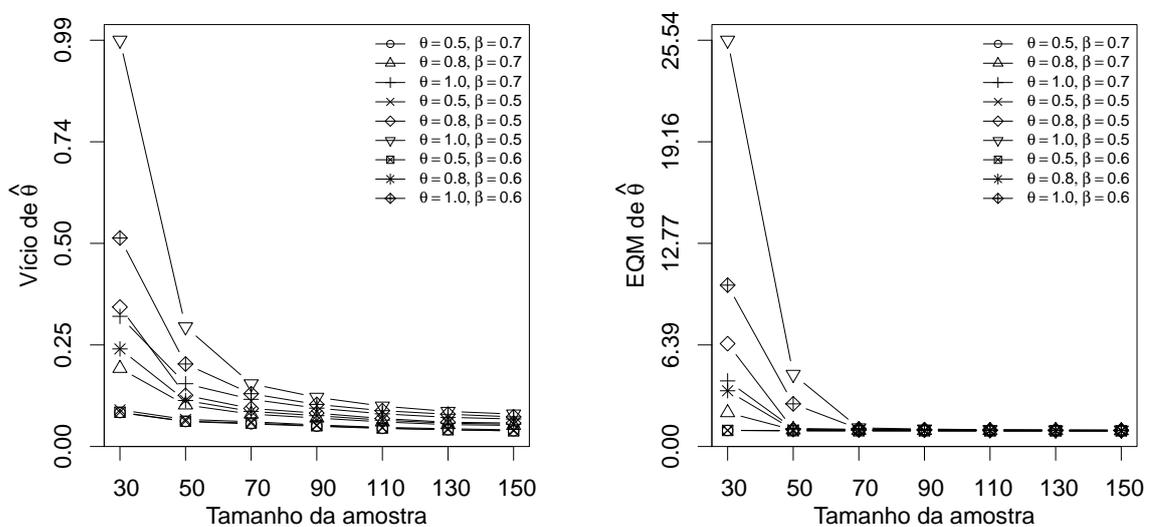


Figura 50 – Vício e EQM de $\hat{\theta}$ da distribuição Lindley ponderada discreta deslocada da origem.

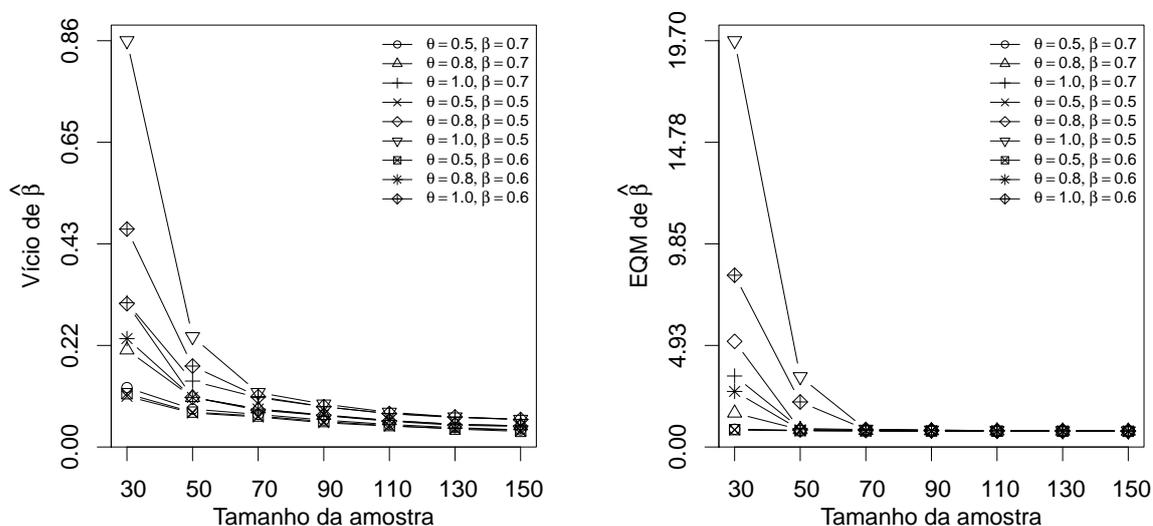


Figura 51 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley ponderada discreta deslocada da origem.

De acordo com as Tabelas 17 e 18, observa-se que o vício e o EQM são positivos em todos os cenários para $\hat{\theta}$ e $\hat{\beta}$. Porém observa-se alguns valores altos para o EQM tanto de $\hat{\theta}$ quanto de $\hat{\beta}$ em certos tamanhos de amostra. As Figuras 50 e 51 ilustram essa conclusão e nelas ainda pode-se notar um aspecto exponencial tanto para o vício quanto para o EQM de $\hat{\theta}$ e $\hat{\beta}$.

Tabela 17 – Vício e EQM de $\hat{\theta}$ da distribuição Lindley ponderada discreta deslocada da origem.

n	$\beta = 0.7$			$\beta = 0.5$			$\beta = 0.6$			
	θ			θ			θ			
	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0	
Vício	30	0.0569	0.1679	0.2985	0.0638	0.3218	0.9913	0.0578	0.2166	0.4949
	50	0.0344	0.0761	0.1293	0.0397	0.1002	0.2719	0.0353	0.0873	0.1788
	70	0.0286	0.0531	0.0902	0.0340	0.0672	0.1289	0.0305	0.0599	0.1042
	90	0.0225	0.0433	0.0680	0.0259	0.0555	0.0952	0.0238	0.0491	0.0774
	110	0.0171	0.0335	0.0538	0.0198	0.0420	0.0730	0.0181	0.0380	0.0617
	130	0.0137	0.0267	0.0449	0.0164	0.0327	0.0598	0.0138	0.0303	0.0527
	150	0.0112	0.0247	0.0407	0.0135	0.0306	0.0532	0.0110	0.0277	0.0462
EQM	30	0.0311	1.1950	3.2711	0.0399	5.7061	25.5402	0.0339	2.6259	9.5315
	50	0.0170	0.0677	0.1649	0.0206	0.1027	3.6971	0.0176	0.0836	1.7709
	70	0.0118	0.0425	0.1063	0.0146	0.0616	0.1987	0.0126	0.0501	0.1409
	90	0.0083	0.0315	0.0728	0.0104	0.0440	0.1227	0.0090	0.0358	0.0905
	110	0.0064	0.0228	0.0545	0.0078	0.0301	0.0835	0.0069	0.0258	0.0660
	130	0.0053	0.0195	0.0429	0.0064	0.0238	0.0667	0.0056	0.0217	0.0525
	150	0.0044	0.0165	0.0382	0.0053	0.0207	0.0563	0.0047	0.0184	0.0455

Tabela 18 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley ponderada discreta deslocada da origem.

n	$\beta = 0.7$			$\beta = 0.5$			$\beta = 0.6$			
	θ			θ			θ			
	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0	
Vício	30	0.1115	0.1928	0.2949	0.0926	0.2926	0.8608	0.0982	0.2173	0.4538
	50	0.0649	0.0902	0.1255	0.0565	0.0904	0.2213	0.0584	0.0903	0.1579
	70	0.0535	0.0654	0.0898	0.0480	0.0627	0.1008	0.0500	0.0650	0.0920
	90	0.0413	0.0513	0.0700	0.0355	0.0512	0.0758	0.0380	0.0525	0.0702
	110	0.0317	0.0400	0.0550	0.0281	0.0386	0.0581	0.0293	0.0406	0.0554
	130	0.0249	0.0319	0.0468	0.0233	0.0301	0.0479	0.0217	0.0322	0.0481
	150	0.0200	0.0290	0.0427	0.0192	0.0275	0.0428	0.0168	0.0288	0.0423
	EQM	30	0.1198	0.9351	2.8021	0.0818	4.5481	19.7014	0.0981	2.0254
50		0.0649	0.0995	0.1558	0.0428	0.0861	2.7663	0.0514	0.0937	1.4950
70		0.0456	0.0634	0.1023	0.0307	0.0524	0.1194	0.0367	0.0575	0.1095
90		0.0332	0.0461	0.0723	0.0227	0.0367	0.0725	0.0269	0.0408	0.0708
110		0.0257	0.0345	0.0540	0.0175	0.0258	0.0494	0.0211	0.0302	0.0514
130		0.0211	0.0299	0.0439	0.0145	0.0206	0.0398	0.0170	0.0254	0.0421
150		0.0173	0.0262	0.0392	0.0118	0.0184	0.0337	0.0143	0.0224	0.0365

3.5 Modelos Ponderados

Por outro lado, como outra alternativa ao truncamento e ao deslocamento da origem, pode-se considerar a ponderação de uma variável aleatória discreta. De acordo com Rao (1965) e Patil e Rao (1978), uma variável aleatória discreta ponderada tem função de probabilidade escrita na forma:

$$P(X = x) = \frac{x}{\mathbb{E}(X)} P_0(X = x), \tag{3.115}$$

em que $P_0(X = x)$ representa a função de probabilidade decorrente da observação x_0 e a constante normalizadora é $\mathbb{E}(X)$.

3.5.1 Distribuição Lindley Discreta Ponderada (via Função de Sobrevivência)

Como a distribuição Lindley discreta apresentada por Bakouch, Jazi e Nadarajah (2014) possui forma explícita tanto para a esperança quanto para a variância, neste caso, uma versão ponderada da mesma pode ser criada. De (3.4) e a definição de ponderação de Patil e Rao (1978), propomos a versão ponderada da distribuição Lindley discreta com função de probabilidade dada por:

$$P(X = x | \beta) = \frac{x (e^\beta - 1)^2 (e^{-\beta})^x [\beta (1 - 2e^{-\beta}) + (1 - e^{-\beta})(\beta x + 1)]}{2\beta e^\beta + e^\beta - \beta - 1}. \tag{3.116}$$

em que $\beta > 0$. Neste caso, temos que o suporte é $x = 1, 2, \dots$

O comportamento de (3.116) é ilustrado na Figura 52. Assim, como nas distribuições anteriores, a distribuição Lindley discreta ponderada também satisfaz a equação de log-concavidade, $[P(X = x)]^2 \leq P(X = x - 1)P(X = x + 1)$, e, pelo Teorema 3 de Keilson e Gerber (1971) é unimodal.

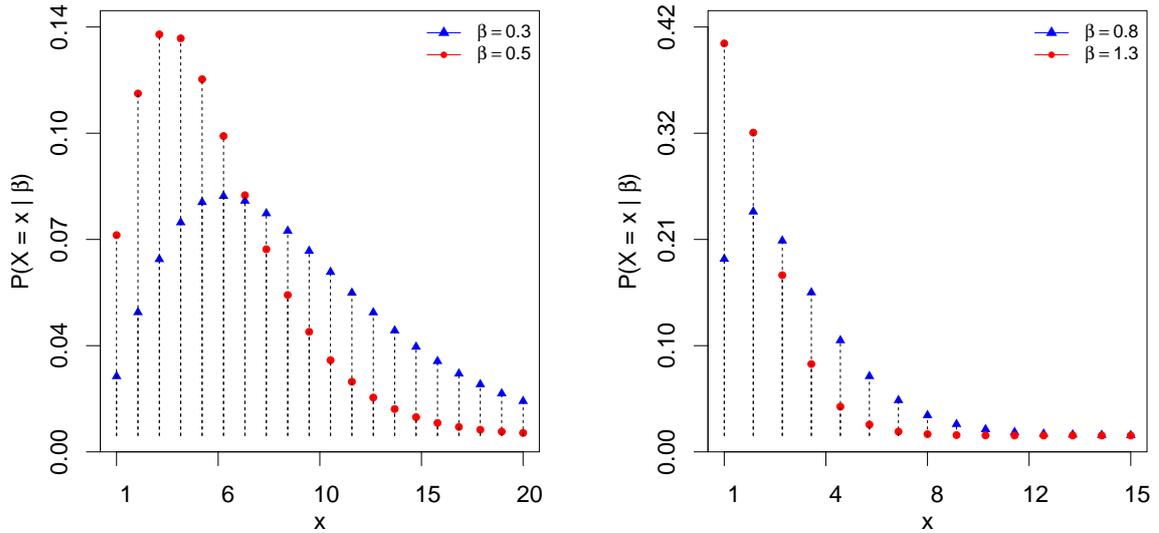


Figura 52 – Comportamento da função de probabilidade da distribuição Lindley discreta ponderada (via função de sobrevivência) para alguns valores de β .

A função de distribuição acumulada, de sobrevivência e de risco correspondentes são, respectivamente, dadas por:

$$F(x | \theta, \beta) = \sum_{j=1}^x \frac{j (e^\beta - 1)^2 (e^{-\beta})^j (\beta (1 - 2e^{-\beta}) + (1 - e^{-\beta})(\beta j + 1))}{2\beta e^\beta + e^\beta - \beta - 1}, \quad (3.117)$$

$$S(x | \theta, \beta) = 1 - \sum_{j=1}^x \frac{j (e^\beta - 1)^2 (e^{-\beta})^j (\beta (1 - 2e^{-\beta}) + (1 - e^{-\beta})(\beta j + 1))}{2\beta e^\beta + e^\beta - \beta - 1}, \quad (3.118)$$

e,

$$h(x | \theta, \beta) = \frac{x (e^\beta - 1)^2 (e^{-\beta})^x (\beta (1 - 2e^{-\beta}) + (1 - e^{-\beta})(\beta x + 1))}{1 - \sum_{j=1}^x j (e^\beta - 1)^2 (e^{-\beta})^j (\beta (1 - 2e^{-\beta}) + (1 - e^{-\beta})(\beta j + 1))} \quad (3.119)$$

Na Figura 53, ilustramos o comportamento da função de risco. Neste caso, para os valores de β considerados, nota-se uma evidência de um comportamento crescente.

Em relação aos momentos, por meio de (3.116), tem-se que a esperança e a variância são dadas, respectivamente, por:

$$\mathbb{E}(X) = \frac{2\beta e^{2\beta} + e^{2\beta} + 3\beta e^\beta - \beta - 1}{(2\beta e^\beta + e^\beta - \beta - 1)(e^\beta - 1)} \quad (3.120)$$

e,

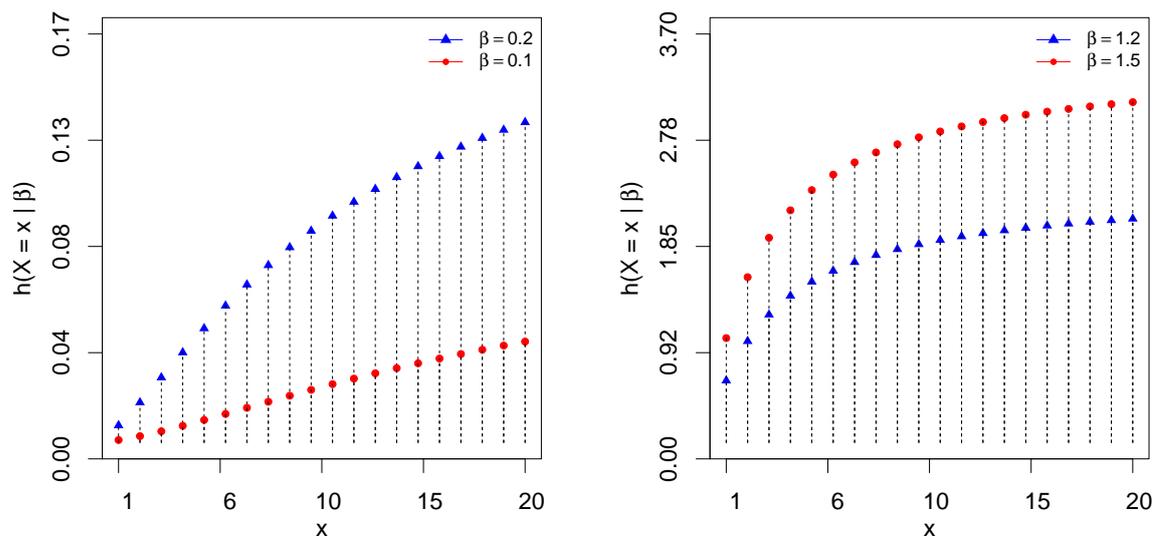


Figura 53 – Comportamento da função de risco da distribuição Lindley discreta ponderada (via função de sobrevivência) para alguns valores de β .

$$\begin{aligned} \text{Var}(X) = & -\frac{(2\beta e^{2\beta} + e^{2\beta} + 3\beta e^{\beta} - \beta - 1)^2}{(2\beta e^{\beta} + e^{\beta} - \beta - 1)^2 (e^{\beta} - 1)^2} \\ & + \frac{(e^{\beta} - 1)^2 (2\beta e^{3\beta} + e^{3\beta} + 13\beta e^{2\beta} + 3e^{2\beta} + 4\beta e^{\beta} - 3e^{\beta} - \beta - 1)}{(2\beta e^{\beta} + e^{\beta} - \beta - 1) (e^{2\beta} - 2e^{\beta} + 1)^2}. \end{aligned} \tag{3.121}$$

Analogamente as outras distribuições, avaliando a razão entre $\mathbb{E}(X)$ e $\text{Var}(X)$ tem-se, numericamente, que $\mathbb{E}(X) < \text{Var}(X)$ para $\beta < 1$ e a Figura 54 ilustra esse resultado. Então, conclui-se que a versão ponderada suporta tanto a superdispersão quanto a subdispersão.

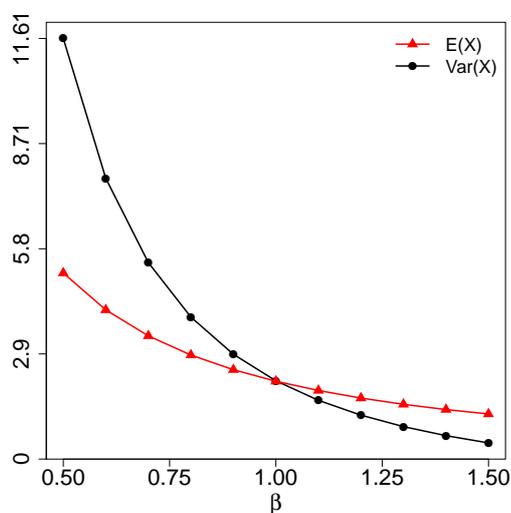


Figura 54 – Esperança e variância distribuição Lindley discreta ponderada (via função de sobrevivência) para alguns valores de β .

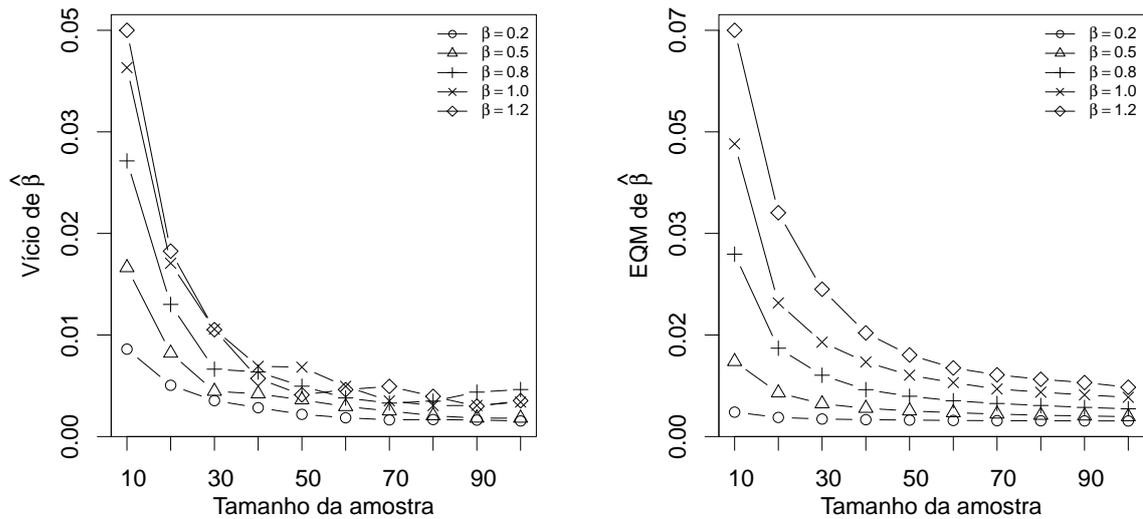


Figura 55 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta ponderada (via função de sobrevivência).

Dada uma amostra aleatória x_1, \dots, x_n de (3.116), o estimador do parâmetro de dispersão β é solução de $\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = 0$ em β em que:

$$\frac{d}{d\beta} \ell(\beta | \mathbf{x}) = \frac{2ne^\beta}{e^\beta - 1} - \frac{e^\beta(2n + n\beta) - n}{e^\beta(2\beta + 1 - \beta - 1)} + \sum_{i=1}^n \frac{1 + x_i - (1 - 2\beta + x_i + \beta x_i)e^{-\beta}}{1 + (1 + x_i)\beta - (1 + 2\beta + \beta x_i)e^{-\beta}}$$

De modo análogo as distribuições anteriores, avaliamos algumas propriedades de $\hat{\beta}$ por meio de simulações Monte Carlo. Os cenários considerados foram os mesmos das distribuições uni-paramétricas anteriores e os resultados são dispostos na Tabela 19.

Tabela 19 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta ponderada (via função de sobrevivência).

	n	Valores de β						n	Valores de β				
		0.2	0.5	0.8	1.0	1.2			0.2	0.5	0.8	1.0	1.2
Vício	10	0.0092	0.0185	0.0308	0.0415	0.0458	EQM	10	0.0017	0.0103	0.0287	0.0475	0.0669
	20	0.0050	0.0087	0.0143	0.0190	0.0204		20	0.0007	0.0049	0.0126	0.0203	0.0357
	30	0.0033	0.0043	0.0069	0.0115	0.0114		30	0.0005	0.0031	0.0080	0.0136	0.0227
	40	0.0024	0.0040	0.0065	0.0072	0.0058		40	0.0004	0.0023	0.0055	0.0102	0.0152
	50	0.0017	0.0033	0.0049	0.0071	0.0039		50	0.0003	0.0019	0.0044	0.0080	0.0114
	60	0.0013	0.0026	0.0036	0.0049	0.0045		60	0.0002	0.0016	0.0036	0.0067	0.0092
	70	0.0010	0.0021	0.0030	0.0032	0.0049		70	0.0002	0.0013	0.0031	0.0056	0.0080
	80	0.0011	0.0015	0.0032	0.0027	0.0037		80	0.0002	0.0011	0.0028	0.0051	0.0073
	90	0.0010	0.0012	0.0042	0.0027	0.0026		90	0.0002	0.0010	0.0025	0.0046	0.0067
	100	0.0009	0.0012	0.0045	0.0031	0.0032		100	0.0001	0.0009	0.0022	0.0042	0.0059

Analisando-se a Tabela 19, concluímos que tanto o vício e quanto o EQM de $\hat{\beta}$ são positivos. Além disso, a magnitude dos mesmos tem tendência de decaimento exponencial. A Figura 55 ilustra tais conclusões.

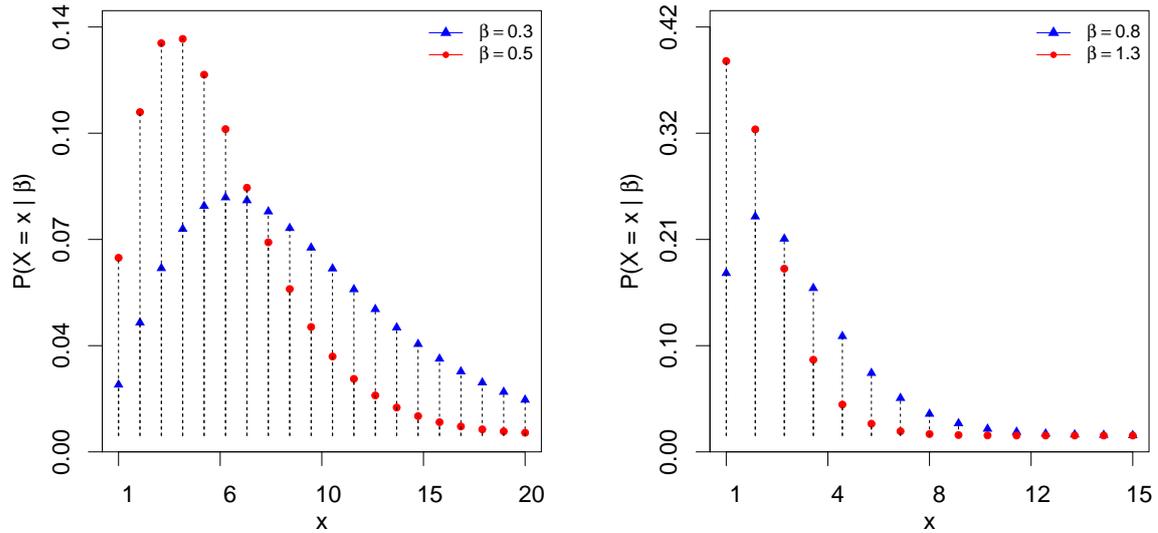


Figura 56 – Comportamento da função de probabilidade da distribuição Lindley discreta ponderada (via série infinita) para alguns valores de β .

3.5.2 Distribuição Lindley Discreta Ponderada (via Série Infinita)

Assim como foi feito para a distribuição Lindley discreta (via função de sobrevivência) de Bakouch, Jazi e Nadarajah (2014), propomos a versão ponderada da distribuição Lindley discreta (via série infinita) com função de probabilidade dada por:

$$P(X = x | \beta) = 0.5x (e^\beta - 1)^3 (1 + x) e^{-\beta(x+2)} \quad (3.122)$$

em que $x = 1, \dots$ e $\beta > 0$. Na Figura 56 pode-se observar que o comportamento de (3.122) é unimodal e, além disso, a inequação de log-concavidade (KEILSON; GERBER, 1971) é satisfeita o que garante o comportamento unimodal.

A função de distribuição acumulada, de sobrevivência e de risco correspondentes são, respectivamente, dadas por:

$$F(x | \theta, \beta) = -0.5e^\beta \left[e^{-\beta(x+3)}(x + (-2x^2 - 4x)e^\beta + e^{2\beta}(x^2 + 3x + 2)x^2) - 2e^{-\beta} \right], \quad (3.123)$$

$$S(x | \theta, \beta) = 1 + 0.5e^\beta \left[e^{-\beta(x+3)}(x + (-2x^2 - 4x)e^\beta + e^{2\beta}(x^2 + 3x + 2)x^2) - 2e^{-\beta} \right], \quad (3.124)$$

e,

$$h(x | \theta, \beta) = \frac{x (e^\beta - 1)^3 (1 + x) e^{-\beta(x+2)}}{e^\beta ((x^2 + 3x + 2) e^{2\beta} + (-2x^2 - 4x) e^\beta + x^2 + x) e^{-\beta(x+3)}} \quad (3.125)$$

Na Figura 57, ilustramos o comportamento da função de risco. Neste caso, para os valores de β considerados, nota-se uma evidência de um comportamento crescente.

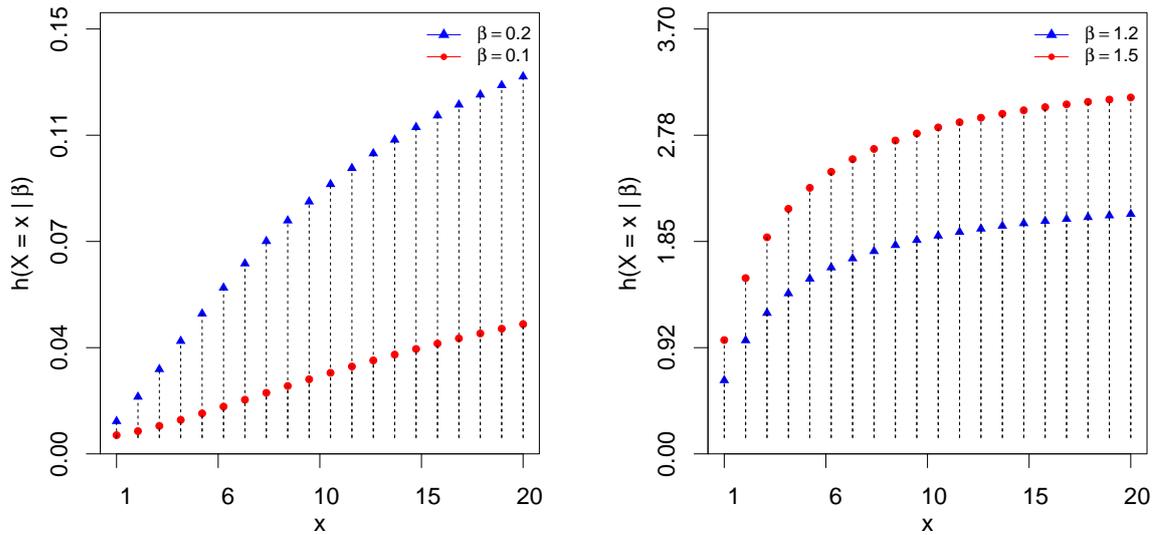


Figura 57 – Comportamento da função de risco da distribuição Lindley discreta ponderada (via série infinita) para alguns valores de β .

De (3.122), tem-se que a esperança e a variância são dadas, respectivamente, por:

$$\mathbb{E}(X) = \frac{e^\beta + 2}{e^\beta - 1}, \tag{3.126}$$

e,

$$\text{Var}(X) = \frac{-3e^\beta}{-e^{2\beta} + 2e^\beta - 1}. \tag{3.127}$$

De maneira análoga as distribuições anteriores, avaliando numericamente a razão entre $\mathbb{E}(X)$ e $\text{Var}(X)$, tem-se que $\mathbb{E}(X) < \text{Var}(X)$ para $0 < \beta < 1$. Para $\beta > 1$ tem-se $\mathbb{E}(X) > \text{Var}(X)$ e a Figura 58 ilustra esse resultado. Então, conclui-se que a versão ponderada suporta tanto a superdispersão quanto a subdispersão.

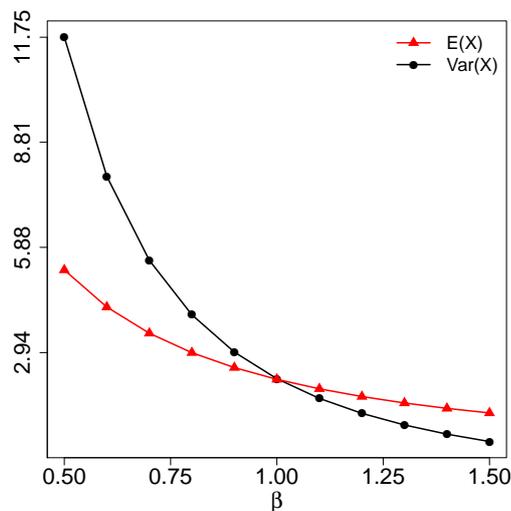


Figura 58 – Esperança e variância distribuição Lindley discreta ponderada (via série infinita) para alguns valores de β .

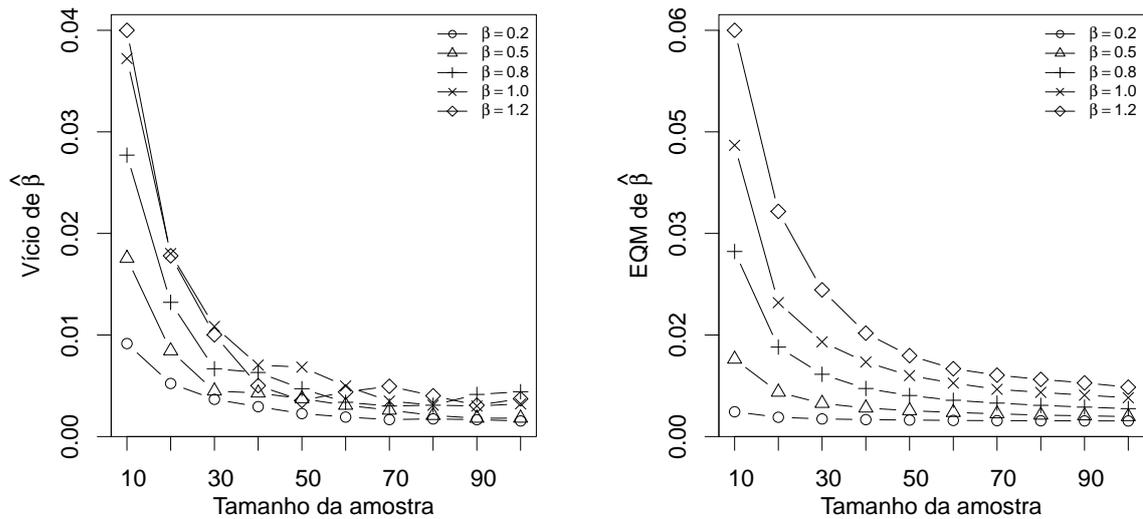


Figura 59 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta ponderada (via série infinita).

Quando a estimação dos parâmetros dessa distribuição, seja x_1, \dots, x_n uma amostra aleatória de (3.122), temos que o logaritmo da função de verossimilhança é proporcional a:

$$\ell(\beta | \mathbf{x}) \propto 3n \log(e^\beta - 1) - \beta(2n + n\bar{x}). \tag{3.128}$$

Assim, derivando-se (3.128) e resolvendo-se em β , obtemos que o estimador de máxima verossimilhança de β é descrito por:

$$\hat{\beta} = \log\left(\frac{\bar{x} + 2}{\bar{x} - 1}\right). \tag{3.129}$$

Analogamente as distribuições anteriores, foi realizado um estudo de simulação Monte Carlo utilizando os mesmos cenários anteriores para avaliar algumas propriedades, como o vício e o EQM de $\hat{\beta}$. Os resultados desse estudo estão dispostos na Tabela 20 e na Figura 59. Por meio da Figura 59 pode-se notar que a taxa de decaimento da magnitude tanto do vício e quanto do EQM de $\hat{\beta}$ tende a ser exponencial. Além disso, tanto o vício quanto o EQM são positivos em todos cenários.

Tabela 20 – Vício e EQM de $\hat{\beta}$ da distribuição Lindley discreta ponderada (via série infinita).

	n	Valores de β						n	Valores de β				
		0.2	0.5	0.8	1.0	1.2			0.2	0.5	0.8	1.0	1.2
Vício	10	0.0091	0.0183	0.0294	0.0397	0.0423	EQM	10	0.0016	0.0102	0.0278	0.0451	0.0632
	20	0.0049	0.0084	0.0136	0.0188	0.0186		20	0.0007	0.0048	0.0122	0.0194	0.0342
	30	0.0032	0.0041	0.0065	0.0111	0.0101		30	0.0005	0.0030	0.0077	0.0129	0.0215
	40	0.0024	0.0039	0.0061	0.0068	0.0047		40	0.0004	0.0023	0.0054	0.0098	0.0144
	50	0.0017	0.0033	0.0043	0.0067	0.0031		50	0.0003	0.0018	0.0043	0.0075	0.0108
	60	0.0013	0.0026	0.0029	0.0047	0.0040		60	0.0002	0.0015	0.0035	0.0063	0.0087
	70	0.0010	0.0021	0.0025	0.0031	0.0046		70	0.0002	0.0013	0.0031	0.0053	0.0076
	80	0.0011	0.0015	0.0026	0.0025	0.0036		80	0.0002	0.0011	0.0027	0.0048	0.0069
	90	0.0010	0.0012	0.0038	0.0026	0.0026		90	0.0002	0.0010	0.0024	0.0043	0.0063
	100	0.0009	0.0012	0.0040	0.0027	0.0033		100	0.0001	0.0009	0.0022	0.0039	0.0057

3.6 Resumo das Distribuições Propostas

Nesta seção, apresentamos uma tabela-resumo com as distribuições propostas neste trabalho. São expostas as características: o nome da distribuição, a função de probabilidade, parâmetros, o suporte (domínio), $E(X)$ e $Var(X)$.

Tabela 21 – Resumo das distribuições propostas.

Distribuição	Função de Probabilidade	Parâmetros	Suporte	$E(X)$	$Var(X)$
LDS ²	$(1 - e^{-\beta})^2(1+x)e^{-\beta x}$	$\beta > 0$	$x \geq 0$	$\frac{2e^{-\beta}}{1 - e^{-\beta}}$	$\frac{2e^{-\beta}}{(1 - e^{-\beta})^2}$
LDR ³	$\frac{\beta^2(1+x)^{x-1} \prod_{i=1}^{x-1} (1 - \beta^2(1+i))}{\beta(x+1) + 1 - \beta(i+1) + 1}$	$\beta > 0$	$x = 1, \dots, k$	—	—
LDTZ ¹	$\frac{e^{-\beta x} [\beta(1 - 2e^{-\beta}) + (1 - e^{-\beta})(1 + \beta x)]}{(1+x)e^{-\beta(x+2)}(e^{\beta} - 1)^2}$	$\beta > 0$	$x > 0$	$\frac{2\beta - \beta e^{-\beta} + 1 - e^{-\beta}}{(1+2\beta)(1 - e^{-\beta})^2}$	$\frac{2e^{2\beta}(3e^{\beta} - 2)}{(e^{\beta} - 1)^2(2e^{\beta} - 1)^2}$ (3.65)
LDSTZ ²	$\frac{e^{-\beta(1+2\beta)}}{(1+x)e^{-\beta(x+2)}(e^{\beta} - 1)^2}$	$\beta > 0$	$x > 0$	$\frac{(e^{\beta} - 1)(2e^{\beta} - 1)}{\beta - e^{-\beta} + 1}$	$\frac{(e^{\beta} - 1)^2(2e^{\beta} - 1)^2}{1 + e^{\beta}}$ (3.94)
SLD ¹	$\frac{e^{-\beta(x-1)}}{1 + \beta} [\beta(1 - 2e^{-\beta}) + (1 - e^{-\beta})(1 + \beta x - \beta)]$	$\beta > 0$	$x > 0$	$\frac{(e^{-\beta} - 1)^2(\beta + 1)}{1 + e^{\beta}}$	$\frac{e^{\beta} - 1}{e^{\beta} - 1}$ (3.101)
SLDS ²	$\frac{x e^{-\beta(1+x)}(e^{\beta} - 1)^2}{2\beta e^{\beta} + e^{\beta} - \beta - 1}$	$\beta > 0$	$x > 0$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$ (3.121)
LDP ¹	$\frac{x(e^{\beta} - 1)^x (e^{-\beta})^x (\beta(1 - 2e^{-\beta}) + (1 - e^{-\beta})(\beta x + 1))}{2\beta e^{\beta} + e^{\beta} - \beta - 1}$	$\beta > 0$	$x > 0$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$	$\frac{2e^{\beta} - e^{2\beta} - 1}{-3e^{\beta}}$ (3.41)
LDSP ²	$0.5x(e^{\beta} - 1)^3(1+x)e^{-\beta(x+2)}$	$\beta > 0$	$x > 0$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$ (3.80)
LPD ¹	$\left(1 + \frac{\beta x^{\alpha}}{\beta + 1}\right) \gamma^{x^{\alpha}} - \left(1 + \frac{\beta(x+1)^{\alpha}}{\beta + 1}\right) \gamma^{(x+1)^{\alpha}}$	$0 < \gamma < 1$ $\alpha, \beta > 0$	$x \geq 0$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$ (3.40)
LPDTZ ¹	$\left(1 + \frac{\beta x^{\alpha}}{\beta + 1}\right) \gamma^{x^{\alpha}} - \left(1 + \frac{\beta(x+1)^{\alpha}}{\beta + 1}\right) \gamma^{(x+1)^{\alpha}}$	$0 < \gamma < 1$ $\alpha, \beta > 0$	$x > 0$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$ (3.79)
SLPD ¹	$\frac{\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)}{\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)} + \frac{\theta^{\beta} e^{-\theta x} [x^{\beta} - (x+1)^{\beta} e^{-\theta}]}{\theta^{\beta} e^{-\theta x} [x^{\beta} - (x+1)^{\beta} e^{-\theta}]}$	$0 < \gamma < 1$ $\alpha, \beta > 0$	$x > 0$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$ (3.109)
LPonD ¹	$\frac{\Gamma(\beta)}{(\theta + \beta)\Gamma(\beta, \theta x) - \Gamma(\beta, \theta x + \theta)} + \frac{\theta^{\beta} e^{-\theta x} [x^{\beta} - (x+1)^{\beta} e^{-\theta}]}{\theta^{\beta} e^{-\theta x} [x^{\beta} - (x+1)^{\beta} e^{-\theta}]}$	$\theta, \beta > 0$	$x \geq 0$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$	$\frac{e^{\beta} + 2}{e^{\beta} - 1}$ (3.56)
LPonDTZ ¹	$\frac{\Gamma(\beta)}{(\theta + \beta)\Gamma(\beta, \theta) + \theta^{\beta} e^{-\theta}} + \frac{\theta^{\beta} e^{-\theta} \Gamma(\beta, \theta) + \theta^{\beta} e^{-\theta}}{\theta^{\beta} e^{-\theta} \Gamma(\beta, \theta) + \theta^{\beta} e^{-\theta}}$	$\theta, \beta > 0$	$x > 0$	—	—
SLPonD ¹	$\frac{\Gamma(\beta, \theta(x-1)) - \Gamma(\beta, \theta(x-1) + \theta)}{\Gamma(\beta)} + \frac{\theta^{\beta} e^{-\theta(x-1)} [(x-1)^{\beta} - x^{\beta} e^{-\theta}]}{(\theta + \beta)\Gamma(\beta)}$	$\theta, \beta > 0$	$x > 0$	—	—

1: via função de sobrevivência; 2: via série infinita; 3: via função de risco

3.7 Aplicação em Dados da Literatura

3.7.1 Aplicação 1

Nesta aplicação, consideramos os dados referentes ao número total de dentes cariados entre os quatro molares decíduos numa amostra de 100 crianças entre 10 e 11 anos de idade (PHYO, 1973). A média e a variancia são, respectivamente, $\bar{x} = 0.67$ dentes e $s^2 = 1.15$ dentes², o que evidencia a superdispersão. O ajuste das distribuições apresentadas neste trabalho (exceto as versões truncada, deslocada e ponderada) foi comparado ao ajuste das distribuições Burr XII discreta (KRISHNA; PUNDIR, 2009), Binomial Negativa, Binomial, Geométrica e Poisson. O conjunto de dados junto aos valores esperados por cada uma das distribuições consideradas na análise pode ser visto na Tabela 22.

Tabela 22 – N° de molares cariados observados e esperados pelas distribuições LPD, LPonD, LDS, LD, DBURR, NB, B, POIS e GEO

N° de Molares Cariados	Observado	Esperado LPD	Esperado LPonD	Esperado LDS	Esperado LD	Esperado DBURR	Esperado NB	Esperado B	Esperado POIS	Esperado GEO
0	64.00	62.59	63.50	60.34	57.13	63.77	62.86	48.03	51.17	59.88
1	17.00	21.27	19.81	26.05	26.88	21.60	21.05	38.66	34.28	24.02
2	10.00	8.91	9.09	9.27	10.45	7.14	8.79	11.67	11.49	9.64
3	6.00	3.91	4.17	3.01	3.71	3.04	3.91	1.56	2.57	3.87
4	3.00	3.33	3.43	1.32	1.84	4.45	3.39	0.08	0.49	2.59

Os parâmetros foram estimados pelo método da máxima verossimilhança e para critério de comparação dos ajustes foram utilizados os valores de $-\log L$, AIC, BIC e o teste χ^2 goodness-of-fit (ver Tabela 23).

Tabela 23 – Estimativas dos parâmetros e critérios de comparação das distribuições LD, LDS, LPD, LPonD, POIS, WEI, GD, BURR e NB.

Distribuição	MLE	E.P.	$-\log L$	χ^2	p -valor	G.L.	AIC	BIC
LD	$\hat{\beta} = 1.29$	0.10	113.67	6.63	0.084	3	229.3591	231.9643
LDS	$\hat{\beta} = 1.38$	0.10	114.50	8.52	0.036	3	231.0036	233.6088
LPD	$\hat{\alpha} = 0.80$ $\hat{\beta} = 1.44$	0.10 0.14	112.05	2.17	0.337	2	228.1019	233.3122
LPonD	$\hat{\theta} = 0.87$ $\hat{\beta} = 0.52$	0.18 0.17	111.52	1.35	0.508	2	227.0559	232.2662
POIS	$\hat{\lambda} = 0.67$	0.08	121.04	29.42	0.001	3	244.0964	246.7015
GEO	$\hat{\beta} = 0.40$	0.03	112.47	3.59	0.309	3	226.9471	229.5523
B	$\hat{p} = 0.16$	0.01	130.95	138.66	0.001	3	263.9152	266.5204
NB	$\hat{p} = 0.67$ $\hat{\mu} = 0.67$	0.26 0.11	112.03	2.12	0.345	2	228.0643	233.2747
DBURR	$\hat{\theta} = 0.23$ $\hat{\alpha} = 1.44$	0.04 0.24	114.73	5.49	0.064	2	233.4794	238.6897

LD: Lindley discreta (via função de sobrevivência); LDS: Lindley discreta (via série infinita); LPD: Lindley potência discreta; LPonD: Lindley ponderada discreta; POIS: Poisson; DBURR: Burr XII discreta; NB: Binomial negativa; B: Binomial; GEO: Geométrica.

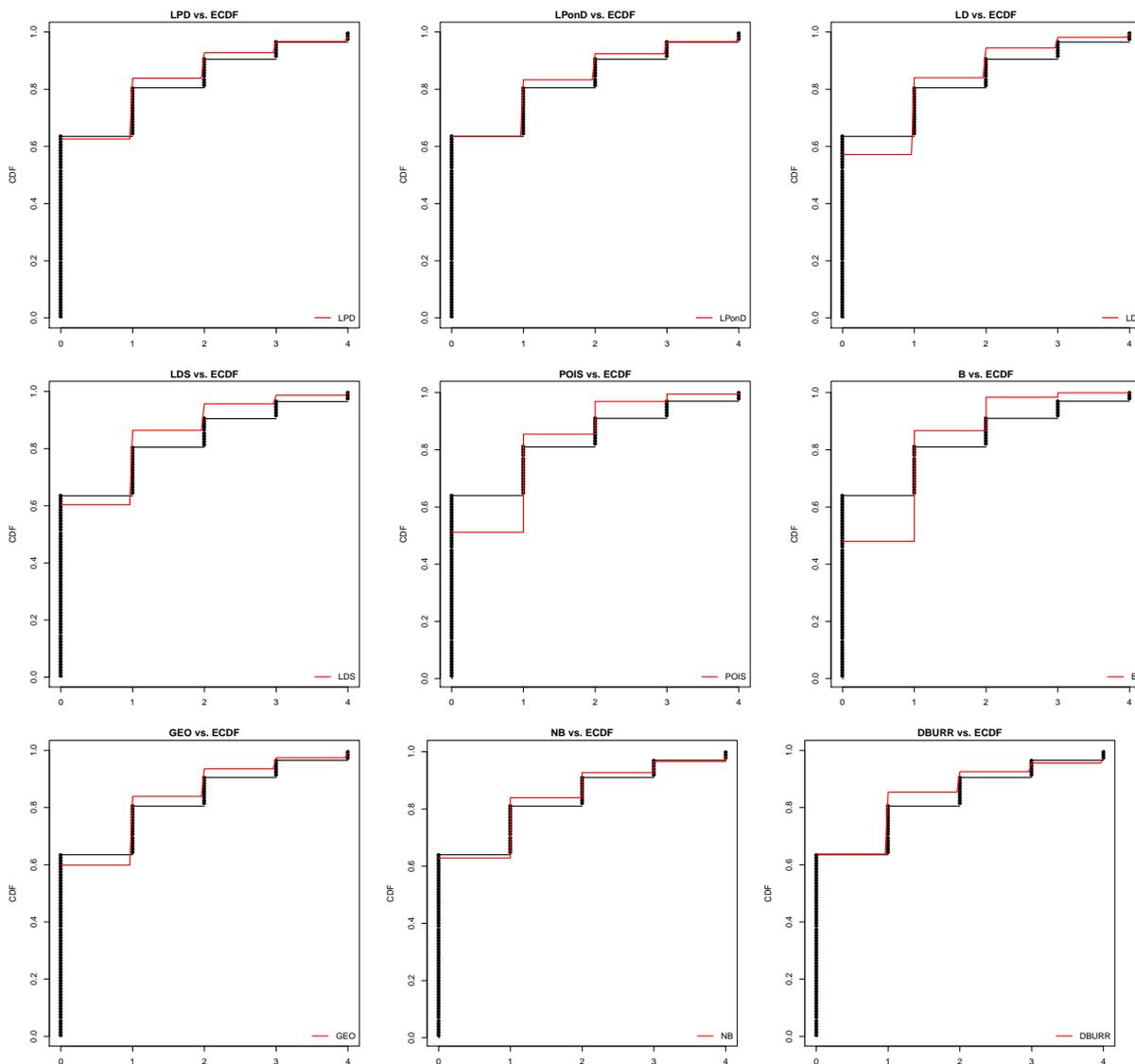


Figura 60 – CDF empírica vs CDF ajustada das distribuições LD, LDS, LPD, LPonD, POIS, GEO, B, DBURR e NB em relação ao número de molares cariados.

Comparando-se os valores descritos na Tabela 23, pode-se notar que, em termos de ajuste, as distribuições Lindley potência discreta e Lindley ponderada discreta obtidos pelo método de discretização baseado na função de sobrevivência (NAKAGAWA; OSAKI, 1975) se mostraram melhores ou equivalentes em relação as distribuições tradicionais (Binomial, Poisson, Geométrica e Binomial negativa) e a distribuição Burr XII discreta.

Além disso, observando-se os valores esperados na Tabela 22, concluímos que as melhores distribuições são a Lindley potência discreta, Lindley ponderada discreta e Binomial negativa. Porém, dentre as três, a melhor ajustada é a distribuição Lindley ponderada discreta. Na Figura 60 é ilustrado o gráfico da CDF empírica contra a CDF de cada uma das distribuições utilizadas na análise. Comparando-os, novamente, o melhor ajuste é da distribuição Lindley ponderada discreta. Portanto, essas novas distribuições se mostram como uma alternativa para modelagem de dados de natureza discreta.

3.7.2 Aplicação 2

Nesta aplicação, consideramos os dados referentes ao número de vezes em que um computador falha em cada uma das 128 semanas de operação (ver Chakraborty e Chakravarty (2012)). A média e a variância são, respectivamente, $\bar{x} = 4.023$ vezes e $s^2 = 14.464$ vezes², o que evidencia superdispersão. O ajuste das distribuições propostas neste trabalho ao ajuste das distribuições Burr XII discreta (KRISHNA; PUNDIR, 2009), Binomial Negativa, Binomial, Geométrica e Poisson. O conjunto de dados junto aos valores esperados por cada uma das distribuições consideradas na análise pode ser visto na Tabela 24.

Tabela 24 – N° de vezes que o computador falha observadas e esperados pelas distribuições LPD, LPonD, LDS, LD, DBURR, NB, B, POIS e GEO

N° de Vezes da Falha	Observado	Esperado LPD	Esperado LPonD	Esperado LDS	Esperado LD	Esperado DBURR	Esperado NB	Esperado B	Esperado POIS	Esperado GEO
0	15.00	16.95	16.19	17.96	16.52	19.89	16.31	2.16	2.31	25.52
1	19.00	19.04	18.94	20.17	18.95	36.60	19.46	8.96	9.27	20.43
2	23.00	18.08	18.25	18.92	18.16	19.05	18.45	18.44	18.61	16.36
3	14.00	15.81	16.08	16.27	15.96	10.70	15.99	25.09	24.91	13.10
4	15.00	13.17	13.43	13.30	13.33	6.83	13.19	25.41	25.01	10.49
6	18.00	19.00	19.35	18.59	19.23	8.28	18.86	33.97	33.52	15.12
9	12.00	15.22	15.35	14.07	15.32	6.66	15.01	13.03	13.31	13.14
9+	12.00	10.72	10.41	8.72	10.53	20.00	10.72	0.94	1.07	13.85

Os parâmetros foram estimados pelo método da máxima verossimilhança e para critério de comparação dos ajustes foram utilizados os valores de $-\log L$, AIC, BIC e o teste χ^2 goodness-of-fit (ver Tabela 25).

Tabela 25 – Estimativas dos parâmetros e critérios de comparação das distribuições LD, LDS, LPD, LPonD, POIS, WEI, GD, BURR e NB.

Distribuição	MLE	E.P.	$-\log L$	χ^2	p-valor	G.L.	AIC	BIC
LPD	$\hat{\alpha} = 0.98$ $\hat{\beta} = 0.38$	0.06 0.04	316.62	2.91	0.71	5	637.25	642.95
LPonD	$\hat{\theta} = 0.38$ $\hat{\beta} = 1.01$	0.04 0.17	316.63	2.84	0.72	5	637.26	642.97
LDS	$\hat{\beta} = 0.40$	0.02	316.80	3.53	0.74	6	635.59	638.45
LD	$\hat{\beta} = 0.38$	0.02	316.64	2.88	0.82	6	635.27	638.13
DBURR	$\hat{\theta} = 3.30$ $\hat{\alpha} = 0.78$	0.68 0.04	342.58	40.18	0.001	5	689.16	694.87
NB	$\hat{p} = 1.69$ $\hat{\mu} = 4.01$	0.31 0.32	316.42	2.53	0.77	5	636.84	642.54
B	$\hat{p} = 0.03$	0.001	390.48	235.95	0.001	6	782.95	785.80
POIS	$\hat{\lambda} = 4.01$	0.17	384.97	209.09	0.001	6	771.95	774.80
GEO	$\hat{\beta} = 0.80$	0.15	320.70	10.04	0.12	6	643.41	646.26

LD: Lindley discreta (via função de sobrevivência); LDS: Lindley discreta (via série infinita); LPD: Lindley potência discreta; LPonD: Lindley ponderada discreta; POIS: Poisson; DBURR: Burr XII discreta; NB: Binomial negativa; B: Binomial; GEO: Geométrica.

Comparando-se os valores ajustados descritos nas Tabelas 25 e 24, pode-se notar que, em termos de ajuste, as distribuições Lindley potência discreta, Lindley ponderada discreta,

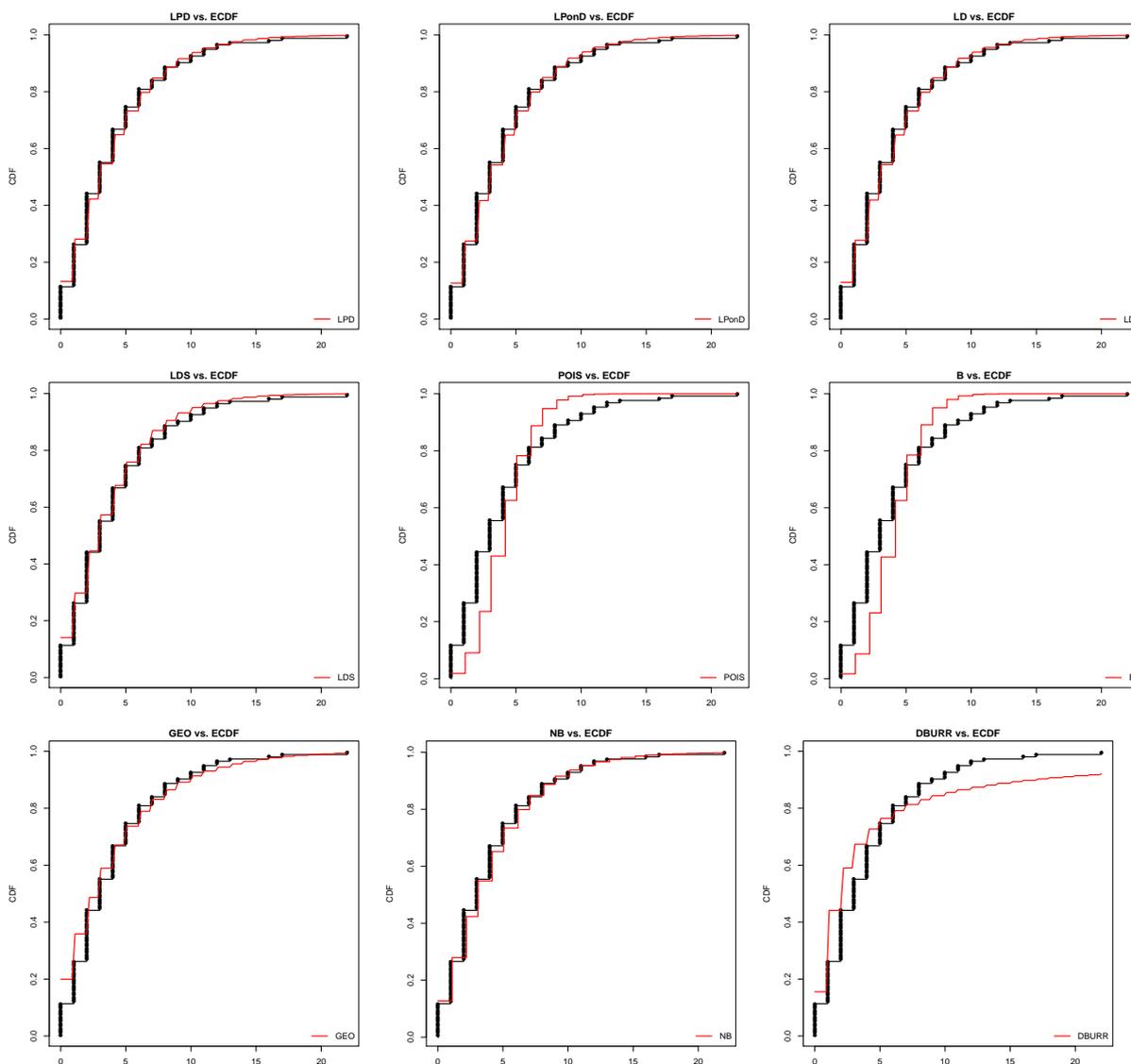


Figura 61 – CDF empírica vs CDF ajustada das distribuições LD, LDS, LPD, LPonD, POIS, GEO, B, DBURR e NB em relação ao número de vezes que o computador falhou.

Lindley discreta obtidos pelo método de discretização baseado na função de sobrevivência e Lindley discreta obtida por série infinita se mostraram equivalentes em relação Binomial negativa. Todavia, dentre essas distribuições, as distribuições Lindley discreta (via função de sobrevivência) e Lindley discreta (via série infinita) são vencedores em termos de parsimonia. Além disso, por termos de equação mais simples, a melhor ajustada é a distribuição Lindley discreta (via série infinita).

Na Figura 61 é ilustrado o gráfico da CDF empírica contra a CDF de cada uma das distribuições utilizadas na análise e pode-se notar que as distribuições selecionadas acima possuem ajuste equivalente. Portanto, de acordo com as análises realizadas, essas novas distribuições se mostram como uma alternativa para modelagem de dados de natureza discreta.

MODELANDO O COMPRIMENTO DAS SEQUÊNCIAS DE DIAS CHUVOSOS

4.1 Introdução

Tão importante quanto a análise do volume da precipitação pluviométrica acumulado em um determinado período de tempo é a análise do comprimento das sequências de dias chuvosos. Segundo Shaw et al. (2010), um dia é considerado chuvoso se a precipitação acumulada em um período de 24 horas for simplesmente mensurável ou superior a algum limiar fixado a priori, por exemplo 0.1, 1.0, 5.0 ou 10.0 milímetros. O valor do limiar usado para definir um dia como chuvoso varia bastante na literatura. Por exemplo, a fim de detectar seus impactos no comprimento das sequências de dias chuvosos na Grécia, em Anagnostopoulou et al. (2003) são considerados limiares de 0.1, 1.0 e 10.0 milímetros/dia já em Barron et al. (2003) é adotado um limiar de 0.85 milímetros/dia. Em Zin e Jemain (2010) é considerado o valor 0.1 milímetro/dia como limiar. Os limiares de 0.1, 1.0 e 5.0 milímetros/dia foram considerados em Lana et al. (2006).

Independente do limiar adotado, a estimação da probabilidade de ocorrer sequências de dias chuvosos de comprimento k , $k \geq 1$, é de grande interesse e importância. Segundo Sen (2009), períodos longos de seca ou de chuva afetam de maneira significativa, por exemplo, produções agrícolas, projetos de gerenciamento de água e atividades industriais.

A análise do comprimento da sequência de uma variável climatológica em um determinado estado não se restringe apenas a precipitação e a descrição de seu comportamento tem uma série de implicações práticas. Por exemplo, longos períodos com baixa temperatura aumentam o consumo de energia, afetam a produção de leite, afetam a produção de leite, longos períodos de chuva em uma região também aumentam o risco de inundações, dentre outras implicações (MCCALLA; DAY; MILLWARD, 1978).

Na análise do comprimento das sequências dos dias chuvosos e secos leva em consideração duas condições (MATHUGAMA; PEIRIS, 2011). A primeira considera que o estado

- chuvoso ou seco - em um determinado dia, depende do que ocorreu no dia ou dias anteriores. A segunda, considerada neste trabalho, que os comprimentos das sequências num certo estado são independentes e identicamente distribuídos. Sob estas suposições a análise se inicia pela escolha e ajuste de uma ou mais distribuições discretas de probabilidade com suporte no intervalo $[1; k]$. O valor de k depende do período considerado, por exemplo, se o objetivo é analisar os comprimentos das sequências mensais, nos meses de janeiro, maio, julho, agosto e outubro, o comprimento máximo observado será $k = 30$ dias.

Naturalmente, antes da escolha e ajuste de uma distribuição de probabilidade, a hipótese de que os comprimentos das sequências são independentes e identicamente distribuídos deve ser testada. Esta suposição é necessária, por exemplo, para estimar os parâmetros das distribuições candidatas pelo método da máxima verossimilhança. Neste trabalho as suposições supracitadas são avaliadas a partir dos valores das estatísticas dos testes de Bartels (BARTELS, 1982), Wald-Wolfowitz (WALD; WOLFOWITZ, 1943) e Mann-Kendall (MANN, 1945).

Várias funções de probabilidade tem sido introduzidas e utilizadas na descrição do comportamento dos comprimentos das sequências dos dias chuvosos. Deni, Jemain e Ibrahim (2008) ajusta sete distribuições aos comprimentos das sequências de dias chuvosos e secos observados entre 1971 e 2005 em várias estações meteorológicas da Malaysia Peninsular. Deni e Jemain (2009a) e Deni, Jemain e Ibrahim (2010) analisaram o comportamento dos comprimentos das sequências utilizando, respectivamente, 12 e 13 funções de probabilidade, dentre elas a mistura de distribuições logarítmicas, a mistura da logarítmica e Poisson e a mistura da logarítmica e geométrica. A mistura de duas distribuições também foram consideradas por Dobi-Wantuch, Mika e Szeidl (2000), Deni e Jemain (2008b), Deni, Jemain e Ibrahim (2009), Deni e Jemain (2009b), Deni e Jemain (2012) dentre outros.

Neste trabalho, consideramos as versões truncada em zero, deslocada da origem e ponderada das distribuições propostas no Capítulo 3 como alternativa as distribuições Poisson truncada em zero e Binomial negativa truncada em zero que são as distribuições mais frequentemente utilizadas neste tipo de análise. Na Seção 4.2, apresentamos as distribuições Binomial negativa truncada em zero e Poisson truncada em zero. Os dados utilizados são descritos na Seção 4.3. Os resultados obtidos dos ajustes das distribuições consideradas na análise são apresentadas e discutidos na Seção 4.4. A Seção 4.5 finaliza essa aplicação com algumas conclusões.

4.2 Distribuições de Probabilidade

As próximas subseções apresentam alguns detalhes matemáticos sobre as distribuições de probabilidade Binomial negativa truncada em zero e Poisson truncada em zero que estamos considerando a critério de comparação com as distribuições propostas no capítulo anterior para modelar o comprimento de sequências de dias chuvosos.

4.2.1 Binomial Negativa Truncada em Zero

A distribuição Binomial negativa é uma das distribuições estatísticas mais frequentemente utilizada para modelar os dados da contagem quando não podemos garantir equidispersão e ela pode ser obtida de muitas maneiras diferentes, por exemplo, como uma mistura de distribuição Poisson-Gama (HILBE, 2011). Além disso, de acordo com Jain e Consul (1971), a distribuição Binomial negativa é talvez a primeira distribuição de probabilidade cuja variância é maior do que sua média. Sua função de probabilidade é escrita na forma:

$$P(X = x | \theta, \alpha) = \frac{\Gamma\left(x + \frac{1}{\alpha}\right)}{\Gamma(x + 1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{\theta\alpha}{1 + \theta\alpha}\right)^x \left(\frac{1}{1 + \theta\alpha}\right)^{\frac{1}{\alpha}} \quad (4.1)$$

para $x = 0, 1, \dots$, $\alpha > 0$ e $\theta > 0$. A variável aleatória X tem média $\mathbb{E}(X) = \theta$ e variância $Var(X) = \theta(1 + \alpha\theta)$ que excede a média. Na literatura, Cameron e Trivedi (1998), a refere como a distribuição NEGBIN2 e quando α tende para 0 observa-se a distribuição de Poisson e para $\alpha = 1$ a distribuição Geométrica.

Um fato importante a se destacar é de que a estimativa de máxima verossimilhança de α existe e é única se a variância amostral exceder a média amostral (ARAGÓN; EBERLY; EBERLEY, 1992). Wang (1996), Saha e Paul (2005) e Al-Khasawneh (2010) outras referências importantes sobre este assunto.

No entanto, a distribuição binomial negativa não pode ser aplicada diretamente para descrever a distribuição dos comprimentos dos períodos de chuva, porque estes períodos tem sempre um comprimento de pelo menos um dia. Esta situação é semelhante quando vamos modelar, por exemplo, a distribuição do comprimento de permanência hospitalar em dias, o número de artigos publicados por professores efetivos e assim por diante (BÖHNING; KUHNERT, 2006).

Para superar o fato de que o valor zero não pode ocorrer, definimos a distribuição binomial negativa truncada em zero (ZTNBD) (JOHNSON; KEMP; KOTZ, 2005). Uma discussão detalhada do modelo ZTNB pode ser encontrada em Grogger e Carson (1991).

A partir de (4.1) temos $P(X = 0 | \theta, \alpha) = \left(\frac{1}{1 + \theta\alpha}\right)^{\frac{1}{\alpha}}$, então uma distribuição ZTNB tem função de probabilidade escrita na forma:

$$P(X = x | X > 0, \theta, \alpha) = \frac{\Gamma\left(x + \frac{1}{\alpha}\right)}{\Gamma(x + 1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{\theta\alpha}{1 + \theta\alpha}\right)^x \left(\frac{1}{1 + \theta\alpha}\right)^{\frac{1}{\alpha}} \left[1 - \left(\frac{1}{1 + \theta\alpha}\right)^{\frac{1}{\alpha}}\right]^{-1} \quad (4.2)$$

para $x = 1, 2, \dots$ e 0 caso contrário. A média e a variância são escritas, respectivamente, como: $\mathbb{E}(X) = \theta \left[1 - \left(\frac{1}{1 + \theta\alpha}\right)^{\frac{1}{\alpha}}\right]^{-1}$ e $Var(X) = \mathbb{E}(X) [(1 + \theta + \theta\alpha) - \mathbb{E}(X)]$ que são um pouco maiores em relação a distribuição binomial negativa. Para $\theta > \frac{(\frac{\alpha}{1 + \alpha})^{-\alpha} - 1}{\alpha}$ a variável aleatória X com distribuição ZTNB tem $Var(X) > \mathbb{E}(X)$ e $Var(X) < \mathbb{E}(X)$ caso contrário. Isso

implica que a distribuição ZTNB pode ser utilizada para modelar dados de contagem com subdispersão e superdispersão.

4.2.2 Poisson Truncada em Zero

A distribuição de Poisson (P) é um modelo de probabilidade tradicional para modelar dados de contagem quando temos equidispersão já que para uma variável aleatória discreta X com distribuição de Poisson, é fácil verificar que $\mathbb{E}(X) = \text{Var}(X)$. A função de probabilidade da distribuição de Poisson é escrita na forma:

$$P(X = x | \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (4.3)$$

em que $\lambda > 0$ e $x \in \mathbb{Z}_+$. A variável aleatória X tem média descrita por $\mathbb{E}(X) = \lambda$ e variância $\text{Var}(X) = \lambda$ que é igual a média.

Assim como a distribuição Binomial Negativa, a distribuição de Poisson não pode ser aplicada diretamente para descrever o comprimento de sequências de dias chuvosos. Neste caso, definimos a distribuição Poisson truncada em zero (ZTP). Um estudo detalhado da mesma pode ser visto em (ZUUR et al., 2009). De (4.3), a função de probabilidade da distribuição ZTP é dada por:

$$P(X = x | X > 0, \lambda) = \frac{\lambda^x e^{-\lambda}}{x! [1 - e^{-\lambda}]} \quad (4.4)$$

em que $\lambda > 0$ e $x \in \mathbb{Z}_+^*$. A média é descrita por $\mathbb{E}(X) = \frac{\lambda}{[1 - e^{-\lambda}]}$ e a variância por $\text{Var}(X) = \frac{\lambda}{[1 - e^{-\lambda}]} - \frac{\lambda^2 e^{-\lambda}}{[1 - e^{-\lambda}]^2}$ que não excede a média. Isso implica que a distribuição ZTP só pode ser utilizada para modelar dados de contagem com subdispersão.

4.3 Os Dados

Utilizamos as séries formadas pelas precipitações diárias observadas em 33 estações meteorológicas localizadas nos estados do Rio Grande do Sul (18 estações), Paraná (8 estações) e Santa Catarina (7 estações). As séries foram obtidas no portal do Instituto Nacional de Meteorologia <<http://www.inmet.gov.br>> e abrangem período compreendido entre 1º de janeiro de 1995 e 31 de dezembro de 2015.

Os comprimentos das sequências foram contados dentro de cada mês e foi adotado como um dia chuvoso os dias com precipitação superior ou igual 5.0mm/dia. Isso se da ao fato de que, ao se estudar o clima regional do Sul do Brasil, observa-se que o mesmo possui uma pluviometria homogênea e é mesotérmico do tipo temperado. Ademais, essa região, possui um litoral em toda sua extensão o que implica um ativo processo de evaporação e este, por sua vez, leva à condensação ou formação de nuvens, que se convertem em chuvas. Além disso, as

posições latitudinal e marítima dessa região determinam uma intensa insolação e evaporação, além de forte concentração de núcleos de condensação que contribuem para o acréscimo de chuvas sempre que esta região é atingida por frentes frias (IBGE, 1977).

Uma observação levada em consideração é de que, possivelmente por algum problema operacional nas estações, o número de dias de medição não coincidem com o número de dias no período considerado para a compilação dos comprimentos das sequências. Além disso, meses em que o número de dias de medição não coincidiram com o número de dias do calendário foram descartados.

Na Tabela 26, para cada um das estações, é apresentado as características dos comprimentos das sequências dos dias chuvosos. A saber, a cidade onde se localiza a estação (Cidade), o código da estação segundo a Organização Mundial de Meteorologia (OMM), o período usado na consolidação da série (De - Até), número de dias no período (n_1), número efetivo de dias que a estação operou no período (n_2), a porcentagem de dias que a estação não operou no período (Perdidos), número de dias com precipitação superior a 5.0mm/dia (Precipitação) e a média e variância do comprimento da sequência de dias chuvosos (Média e Var).

Tabela 26 – Características do banco de dados para cada estação.

Cidade	OMM	De	Até	n_1	n_2	Perdidos	Precipitação	Média	Var
Londrina	83766	1995-01-01	2014-12-31	7305	6446	11.76%	1217	1.59	1.02
Maringa	83767	1995-01-01	2014-11-30	7274	6233	14.31%	1207	1.55	0.94
Campo Mourao	83783	1995-01-01	2014-12-31	7305	6666	8.75%	1321	1.52	0.81
Ivai	83811	1996-01-01	2014-12-31	6940	6175	11.02%	1263	1.54	0.74
Castro	83813	1996-01-01	2014-12-31	6940	5928	14.58%	1172	1.52	0.74
Irati	83836	1995-01-01	2014-12-31	7305	7057	3.39%	1480	1.56	0.85
Curitiba	83842	1995-01-01	2014-12-31	7305	6904	5.49%	1406	1.53	0.81
Paranagua	83844	1995-01-01	2014-12-31	7305	6416	12.17%	1724	1.76	1.37
Indaial	83872	1995-01-01	2014-12-31	7305	6298	13.79%	1409	1.56	0.87
Irai	83881	1995-01-01	2014-12-31	7305	6418	12.14%	1362	1.54	0.85
Chapeco	83883	1995-01-01	2014-12-31	7305	6815	6.71%	1468	1.52	0.85
Campos Novos	83887	1995-01-01	2014-12-31	7305	6420	12.11%	1520	1.58	0.90
Lages	83891	1995-01-01	2014-12-31	7305	7057	3.39%	1546	1.51	0.77
Florianopolis	83897	1995-01-01	2014-12-31	7305	5872	19.62%	1311	1.52	0.79
Sao Luiz Gonzaga	83907	1995-01-01	2014-12-31	7305	6663	8.79%	1280	1.52	0.72
Cruz Alta	83912	1995-01-01	2014-12-31	7305	6757	7.50%	1382	1.51	0.79
Passo Fundo	83914	1995-01-01	2014-12-31	7305	6329	13.36%	1364	1.51	0.73
Lagoa Vermelha	83916	2000-01-01	2014-12-31	5479	4319	21.17%	930	1.56	0.85
Bom Jesus	83919	1995-01-01	2014-12-31	7305	6634	9.19%	1500	1.54	0.75
Sao Joaquim	83920	1995-01-01	2014-12-31	7305	6057	17.08%	1489	1.59	0.88
Urussanga	83923	2001-09-01	2014-08-31	4748	4687	1.28%	980	1.61	0.96
Uruguaiana	83927	1995-01-01	2014-12-31	7305	6414	12.20%	983	1.44	0.58
Santa Maria	83936	1995-01-01	2014-12-31	7305	7026	3.82%	1333	1.43	0.61
Bento Gonçalves	83941	1995-01-01	2013-12-31	6940	5536	20.23%	1144	1.52	0.69
Caxias Do Sul	83942	1995-01-01	2014-12-31	7305	6698	8.31%	1460	1.51	0.77
Torres	83948	1995-01-01	2014-12-31	7305	6082	16.74%	1223	1.53	0.81
Santana Do Livramento	83953	1998-06-01	2013-11-30	5662	4350	23.17%	685	1.42	0.49
Encruzilhada Do Sul	83964	1995-01-01	2014-12-31	7305	7026	3.82%	1344	1.46	0.58
Porto Alegre	83967	1995-01-01	2014-12-31	7305	6964	4.67%	1288	1.47	0.59
Bage	83980	1995-01-01	2014-12-31	7305	6660	8.83%	1152	1.45	0.59
Pelotas	83985	1996-01-01	2014-08-31	6818	6113	10.34%	995	1.40	0.52
Rio Grande	83995	1995-01-01	2014-12-31	7305	6965	4.65%	1099	1.37	0.47
Santa Vitoria Do Palmar	83997	1995-01-01	2014-12-31	7305	5996	17.92%	861	1.35	0.46

De acordo com a Tabela 26, nota-se que a maior taxa de valores perdidos é proveniente da estação RS-83941 com 20.23% e a menor proveniente das estações PR-83836 e SC-83891 com 3.39% em ambas. Ainda, a estação em que se teve o maior número de dias

com precipitação superior a 5.0mm/dia é a estação PR-83844 com 1724 dias e a que teve o menor número de dias com precipitação superior a 5.0mm é a estação RS-83953 com 685 dias. Por fim, pode-se perceber que os dados possuem subdispersão $\mathbb{E}(X) > \text{Var}(X)$, o que induz o uso de distribuições que comportam subdispersão para a análise dos mesmos. Além disso, de distribuições com subdispersão serem candidatas, as mesmas devem possuir suporte $x > 0$ já que a resposta $x = 0$ não é observada neste caso.

4.4 Resultados e Discussões

Segundo o teste de Bartels, em 3 das 33 séries a hipótese de comprimentos i.i.d foi rejeitada em nível de significância de 1%. Estas séries são provenientes das estações PR-83783, RS-83912 e SC-83887. As séries provenientes das estações PR-83783 e RS-83912 também não passaram no teste de Wald-Wolfwitz adotando-se o mesmo nível de significância. Todas as séries passaram no teste de Mann-Kendall. Os p-valores associados aos valores das estatísticas, de ambos os testes, foram calculados a partir de $B = 100.000$ permutações (GOOD, 2006) das séries originais. As bibliotecas lawstat (GASTWIRTH et al., 2015), randtests (CAEIRO; MATEUS; CAEIRO, 2014) e (AI, 2014), do sistema R (R Core Team, 2015), foram utilizadas, respectivamente, na aplicação dos testes Bartels e Wald-Wolfwitz, respectivamente.

Para as séries no período considerado, o ajuste das distribuições Lindley e suas variantes Lindley potência e Lindley ponderada truncadas em zero, deslocadas da origem e ponderadas propostas neste trabalho foi comparado ao ajuste das distribuições Poisson truncada em zero e Binomial negativa truncada em zero que são modelos frequentemente usados para modelar as sequências de dias chuvosos.

Os parâmetros foram estimados pelo método da máxima verossimilhança e um resumo do número de séries ajustadas por cada distribuição, considerando os níveis de significância de 1% e 5%, AIC, AICc e BIC são apresentados na [Tabela 27](#). O ajuste foi avaliado pelo teste qui-quadrado de aderência e os valores AIC, AICc e BIC foram usados como critérios de discriminação. A estimação foi feita pela biblioteca "fitdistplus" do software R (R Core Team, 2015).

Por meio da Tabela 27, concluímos que, exceto pelas distribuições Lindley discreta ponderada (via função de sobrevivência e via série infinita), todas as outras foram selecionadas em pelo menos 20 séries ao nível de significância de 1% e 5%. Além disso, o melhor ajuste foi da distribuição Lindley discreta truncada em zero (via função de sobrevivência) tendo o menor AIC em 17 das 33 séries consideradas, o menor AICc em 18 séries e o menor BIC em 21 séries. Por outro lado, a distribuição Lindley discreta ponderada (via série infinita) se ajustou em metade das séries a 1% e a 8 séries a 5%, e, de todas as séries que a mesma se ajustou, não houve nenhuma série em que a mesma teve o menor AIC, AICc e o menor BIC. Já os modelos tradicionais, obtivemos que a distribuição Poisson truncada em zero se ajustou em 7

Tabela 27 – Número de séries ajustadas para cada distribuição em relação ao comprimento da sequência de dias chuvosos.

Distribuição	Nível de significância		AIC	AICc	BIC
	$\alpha = 1\%$	$\alpha = 5\%$			
Lindley Discreta Truncada em Zero ¹	32 (96.97%)	30 (90.91%)	17 (51.51%)	18 (54.54%)	21 (63.63%)
Lindley Discreta Deslocada da Origem ¹	27 (81.82%)	24 (72.72%)	5 (15.15%)	5 (15.15%)	5 (15.15%)
Lindley Discreta Ponderada ¹	18 (54.54%)	12 (36.37%)	1 (3.03%)	1 (3.03%)	1 (3.03%)
Lindley Discreta Truncada em Zero ²	32 (96.97%)	28 (84.84%)	2 (6.06%)	2 (6.06%)	2 (6.06%)
Lindley Discreta Deslocada da Origem ²	21 (63.63%)	20 (60.61%)	2 (6.06%)	2 (6.06%)	2 (6.06%)
Lindley Discreta Ponderada ²	16 (48.48%)	8 (24.25%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Lindley Potência Discreta Truncada em Zero ¹	33 (100.0%)	31 (93.94%)	2 (6.06%)	1 (3.03%)	1 (3.03%)
Lindley Potência Discreta Deslocada da Origem ¹	33 (100.0%)	31 (93.94%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Lindley Ponderada Discreta Truncada em Zero ¹	33 (100.0%)	31 (93.94%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Lindley Ponderada Discreta Deslocada da Origem ¹	33 (100.0%)	31 (93.94%)	4 (12.12%)	4 (12.12%)	1 (6.06%)
Poisson Truncada em Zero	7 (21.21%)	5 (15.15%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Binomial Negativa Truncada em Zero	33 (100.0%)	31 (93.94%)	0 (0.00%)	0 (0.00%)	0 (0.00%)

¹: Via função de sobrevivência.²: Via série infinita.

séries a 1% e 5 séries a 5%, porém não obteve o menor AIC, AICc e BIC em nenhuma das séries. Para a Binomial negativa truncada em zero o ajuste foi em todas as séries a 1% e 31 séries a 5%, porém, para o AIC, AICc e BIC, ocorreu o mesmo com a distribuição Poisson truncada em zero.

Diante dos resultados acima, apresentamos os valores observados e esperados pelas distribuições Lindley discreta truncada em zero (via função de sobrevivência) (LDTZ¹), Lindley discreta deslocada da origem (via função de sobrevivência) (LDD¹), Lindley ponderada discreta deslocada da origem (LPondD¹) e binomial negativa truncada em zero (ZTNB) para cada uma das estações consideradas na análise. Tais valores são descritos nas tabelas abaixo.

Tabela 28 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83923.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	374.00	371.79	364.81	375.50	377.04
2	154.00	151.62	159.85	146.67	146.13
3	54.00	56.48	57.77	56.46	54.89
4	12.00	19.95	19.06	20.52	20.29
4+	16.00	10.16	8.51	10.85	11.66

Tabela 29 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83766.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	498.00	472.51	463.90	498.23	496.22
2	161.00	188.87	199.06	157.95	164.26
3	60.00	68.95	70.43	65.43	62.19
4	29.00	23.87	22.75	26.72	25.02
4+	18.00	11.80	9.85	17.67	18.31

Tabela 30 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83767.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	510.00	495.13	487.07	512.11	513.49
2	181.00	188.19	198.09	166.67	168.18
3	49.00	65.32	66.39	63.95	60.69
4	23.00	21.50	20.30	23.83	22.91
4+	17.00	9.87	8.15	13.43	14.73

Tabela 31 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83783.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	573.00	559.68	551.18	573.03	571.66
2	189.00	206.17	216.83	188.72	192.61
3	71.00	69.34	70.26	68.81	66.61
4	16.00	22.11	20.77	24.13	23.33
4+	18.00	9.70	7.96	12.30	12.78

Tabela 32 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83811.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	510.00	521.91	513.69	511.69	515.05
2	218.00	196.96	207.24	209.90	205.11
3	63.00	67.88	68.88	68.94	69.18
4	18.00	22.18	20.90	20.55	21.23
4+	10.00	10.08	8.29	7.92	8.43

Tabela 33 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83813.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	491.00	497.66	490.24	491.90	494.21
2	192.00	182.90	192.32	190.10	187.09
3	66.00	61.37	62.14	62.10	61.92
4	14.00	19.53	18.32	18.61	19.02
4+	7.00	8.54	6.99	7.29	7.75

Tabela 34 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83836.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	592.00	597.41	587.50	594.36	598.09
2	245.00	230.50	242.68	233.81	229.89
3	74.00	81.22	82.62	82.17	80.90
4	25.00	27.13	25.67	26.77	27.15
4+	13.00	12.73	10.53	11.89	12.97

Tabela 35 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83842.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	596.00	587.96	578.81	596.81	596.58
2	215.00	219.37	230.79	207.59	209.67
3	61.00	74.74	75.81	74.64	72.69
4	33.00	24.14	22.72	25.53	25.03
4+	12.00	10.79	8.87	12.43	13.03

Tabela 36 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação PR83844.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	562.00	538.20	524.14	562.32	559.52
2	233.00	254.94	269.52	228.27	235.88
3	94.00	110.38	114.57	105.22	102.12
4	54.00	45.33	44.49	46.92	44.79
4+	35.00	29.15	25.28	35.28	35.69

Tabela 37 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83881.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	574.00	566.48	557.57	575.79	577.23
2	211.00	212.48	223.56	200.14	200.45
3	64.00	72.77	73.85	72.58	70.19
4	18.00	23.63	22.26	25.08	24.68
4+	19.00	10.64	8.76	12.41	13.45

Tabela 38 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83907.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	533.00	546.28	538.24	535.51	540.31
2	226.00	199.76	210.00	213.71	207.00
3	57.00	66.69	67.48	67.48	67.67
4	18.00	21.11	19.78	19.29	20.22
4+	9.00	9.17	7.49	7.01	7.79

Tabela 39 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83912.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	602.00	597.63	588.98	604.33	606.90
2	221.00	215.67	226.69	206.43	204.99
3	54.00	71.05	71.84	71.29	69.08
4	29.00	22.19	20.77	23.33	23.25
4+	10.00	9.46	7.72	10.61	11.78

Tabela 40 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83914.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	591.00	592.49	584.03	592.40	594.54
2	219.00	212.83	223.67	212.48	210.51
3	66.00	69.80	70.52	70.38	69.28
4	21.00	21.70	20.29	21.74	21.94
4+	9.00	9.19	7.49	9.00	9.73

Tabela 41 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83916.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	374.00	377.59	371.40	376.14	379.11
2	158.00	144.90	152.54	146.39	143.27
3	40.00	50.78	51.63	51.33	50.32
4	18.00	16.87	15.95	16.72	17.00
4+	8.00	7.85	6.49	7.43	8.29

Tabela 42 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83919.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	611.00	618.92	609.12	612.44	614.46
2	251.00	233.91	246.13	241.77	239.24
3	70.00	80.72	81.95	81.80	81.53
4	28.00	26.42	24.90	25.46	25.80
4+	12.00	12.03	9.90	10.53	10.97

Tabela 43 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83927.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	472.00	470.18	464.81	471.30	469.75
2	146.00	151.93	159.20	150.12	152.55
3	49.00	44.80	44.83	45.19	44.76
4	14.00	12.52	11.51	12.74	12.43
4+	3.00	4.58	3.65	4.66	4.51

Tabela 44 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83936.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	664.00	648.19	641.01	663.13	660.42
2	180.00	205.38	215.13	184.71	190.18
3	62.00	59.37	59.33	60.01	57.87
4	23.00	16.27	14.92	18.88	18.08
4+	6.00	5.79	4.61	8.26	8.46

Tabela 45 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83941.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	478.00	487.66	480.46	479.01	480.57
2	197.00	178.53	187.70	189.65	187.14
3	49.00	59.67	60.39	60.40	60.86
4	21.00	18.91	17.73	17.47	17.84
4+	8.00	8.23	6.73	6.47	6.59

Tabela 46 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83942.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	637.00	629.06	619.87	637.87	638.44
2	219.00	227.85	239.52	215.92	217.07
3	77.00	75.35	76.21	75.45	73.32
4	20.00	23.62	22.13	25.07	24.69
4+	13.00	10.12	8.27	11.69	12.48

Tabela 47 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83948.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	523.00	513.23	505.30	523.42	522.87
2	181.00	190.84	200.76	177.45	179.99
3	58.00	64.79	65.70	64.48	62.54
4	27.00	20.86	19.62	22.42	21.83
4+	10.00	9.28	7.62	11.23	11.77

Tabela 48 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83953.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	329.00	338.20	334.69	329.49	330.75
2	121.00	105.16	110.05	117.75	115.26
3	24.00	29.83	29.73	28.95	30.01
4	8.00	8.02	7.32	6.25	6.49
4+	2.00	2.78	2.20	1.55	1.49

Tabela 49 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83964.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	606.00	620.06	612.42	606.48	607.82
2	231.00	208.82	219.04	227.45	224.59
3	58.00	64.18	64.44	64.01	65.37
4	19.00	18.70	17.29	16.13	16.46
4+	5.00	7.23	5.81	4.93	4.76

Tabela 50 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83967.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	578.00	589.05	581.63	578.02	578.43
2	216.00	200.31	210.17	215.24	213.88
3	60.00	62.17	62.48	62.28	63.31
4	18.00	18.29	16.94	16.25	16.40
4+	5.00	7.17	5.78	5.22	4.98

Tabela 51 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83980.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	535.00	540.28	533.84	535.32	535.94
2	187.00	178.61	187.27	185.19	184.22
3	53.00	53.89	54.02	54.11	54.26
4	12.00	15.41	14.21	14.50	14.65
4+	7.00	5.81	4.66	4.88	4.94

Tabela 52 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83985.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	497.00	501.32	496.34	497.93	499.53
2	162.00	151.95	158.93	156.62	154.39
3	37.00	42.02	41.80	41.99	42.03
4	8.00	11.01	10.02	10.34	10.67
4+	6.00	3.69	2.91	3.12	3.38

Tabela 53 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83995.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	579.00	576.94	571.79	578.39	576.90
2	160.00	165.60	172.95	163.26	165.71
3	46.00	43.35	42.93	43.85	43.28
4	13.00	10.76	9.71	11.03	10.74
4+	2.00	3.36	2.62	3.47	3.37

Tabela 54 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação RS83997.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	466.00	465.83	462.02	466.30	466.69
2	129.00	128.14	133.66	127.27	126.99
3	31.00	32.14	31.71	32.45	32.10
4	7.00	7.64	6.86	7.72	7.81
4+	3.00	2.26	1.75	2.26	2.42

Tabela 55 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83872.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	567.00	568.12	558.66	568.74	571.62
2	224.00	219.43	231.03	218.09	215.65
3	78.00	77.40	78.74	78.00	76.41
4	23.00	25.89	24.50	26.08	26.19
4+	11.00	12.16	10.07	12.09	13.13

Tabela 56 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83883.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	621.00	621.36	611.99	621.35	621.76
2	229.00	229.09	240.92	228.57	228.76
3	81.00	77.12	78.12	77.77	76.89
4	21.00	24.62	23.11	24.69	24.62
4+	11.00	10.81	8.86	10.62	10.97

Tabela 57 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83887.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	602.00	595.75	585.23	603.06	603.68
2	231.00	236.12	248.77	226.46	227.62
3	86.00	85.47	87.19	85.33	83.23
4	23.00	29.34	27.89	30.45	29.95
4+	19.00	14.32	11.92	15.71	16.52

Tabela 58 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83891.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	667.00	663.55	653.81	668.27	669.60
2	241.00	241.28	253.65	234.59	234.35
3	76.00	80.10	81.04	80.52	78.74
4	25.00	25.21	23.62	26.04	25.91
4+	12.00	10.86	8.88	11.59	12.40

Tabela 59 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83897.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	566.00	561.60	553.31	567.45	569.01
2	204.00	204.61	215.12	196.54	196.12
3	64.00	68.06	68.87	68.28	66.42
4	17.00	21.46	20.12	22.45	22.30
4+	14.00	9.27	7.58	10.27	11.15

Tabela 60 – Valores observados e esperados dos comprimentos das sequências de dias chuvosos para a estação SC83920.

Comprimento das sequências	Observado	Esperado LDTZ ¹	Esperado LDD ¹	Esperado LPondD ¹	Esperado ZTNB
1	575.00	574.08	563.65	575.40	575.72
2	229.00	230.86	243.30	228.60	229.24
3	89.00	84.80	86.61	85.37	84.19
4	26.00	29.54	28.14	29.84	29.61
4+	15.00	14.73	12.30	14.79	15.24

De acordo com os valores esperados para a região do Paraná, podemos fazer algumas conclusões a respeito de das sequências de dias chuvosos. Por exemplo, na cidade de Paranaguá (PR83844), observa-se uma frequência de 35 períodos em que se houve mais do que 4 dias consecutivos de chuva o que é natural dado que a mesma se localiza numa região litorânea. De acordo com os modelos expostos na tabela, a distribuição Lindley ponderada discreta deslocada da origem (LPondD¹) foi a que melhor descreveu esse comprimento visto que a mesma proporciona frequência esperada de 35.28. Por outro lado, na cidade de Ivai (PR83811), a distribuição Lindley discreta truncada em zero (LDTZ¹) se mostra melhor para modelar períodos de 3 dias e de mais de 4 dias consecutivos de chuva perante as outras distribuições. Isso se faz importante na prática uma vez que longos períodos de chuva afetam de maneira significativa plantações, por exemplo.

Em relação as regiões de Santa Catarina e Riso Grande do Sul, a maior frequência em que se houve mais do que 4 dias consecutivos de chuva é de 19 períodos provenientes das cidades Campos Novos-SC (SC83887) e Irai-RS (RS83881). De acordo com os modelos, novamente, o modelo que melhor prediz esses períodos é a distribuição LPondD¹. Entretanto, ao considerar os valores esperados de cada um dos modelos considerados para a cidade de Passo Fundo (RS83914), pode-se notar que a distribuição LDTZ¹ oferece uma melhor previsão para períodos de mais de um dia consecutivos de chuva. Portanto, em termos de ajuste, previsão e parsimonia, o melhor modelo para descrever os comprimentos de sequências de dias chuvosos é a distribuição Lindley discreta truncada em zero (via função de sobrevivência).

Nas Figuras 62, 63, 64 e 65 são ilustrados os ajustes em termos da função de probabilidade das distriuições que obtiveram a melhor seleção dos critérios de discriminação adotados na Tabela 27 para cada uma das estações. Analisando-se os gráficos, pode-se notar que as distribuições Lindley discreta truncada em zero (via função de sobrevivência), Lindley discreta deslocada da origem (via função de sobrevivência) e Lindley ponderada deslocada da origem possuem praticamente um ajuste equivalente para todas as estações. Por outro lado, a distribuição binomial negativa truncada em zero se mostra com ajuste equivalente as outras distribuições para as estações RS83985, RS83995, RS83914, RS83916 e RS83927 do Rio Grande do Sul e um ajuste inferior para as estações PR83766, PR83767 e PR83783 do Paraná e para a estação SC83897 de Santa Catarina.

4.5 Conclusões

Em suma, concluímos que a distribuição Lindley discreta truncada em zero (via função de sobrevivência) obteve um melhor ajuste em termos de parsimonia em comparação a todas as distribuições consideradas na análise. Além disso, para algumas estações, o valor esperado por esta distribuição foi o mais próximo do valor observado dos dados perante as distribuições usuais neste tipo de modelagem. Isso implica que essa distribuição pode ser utilizada na

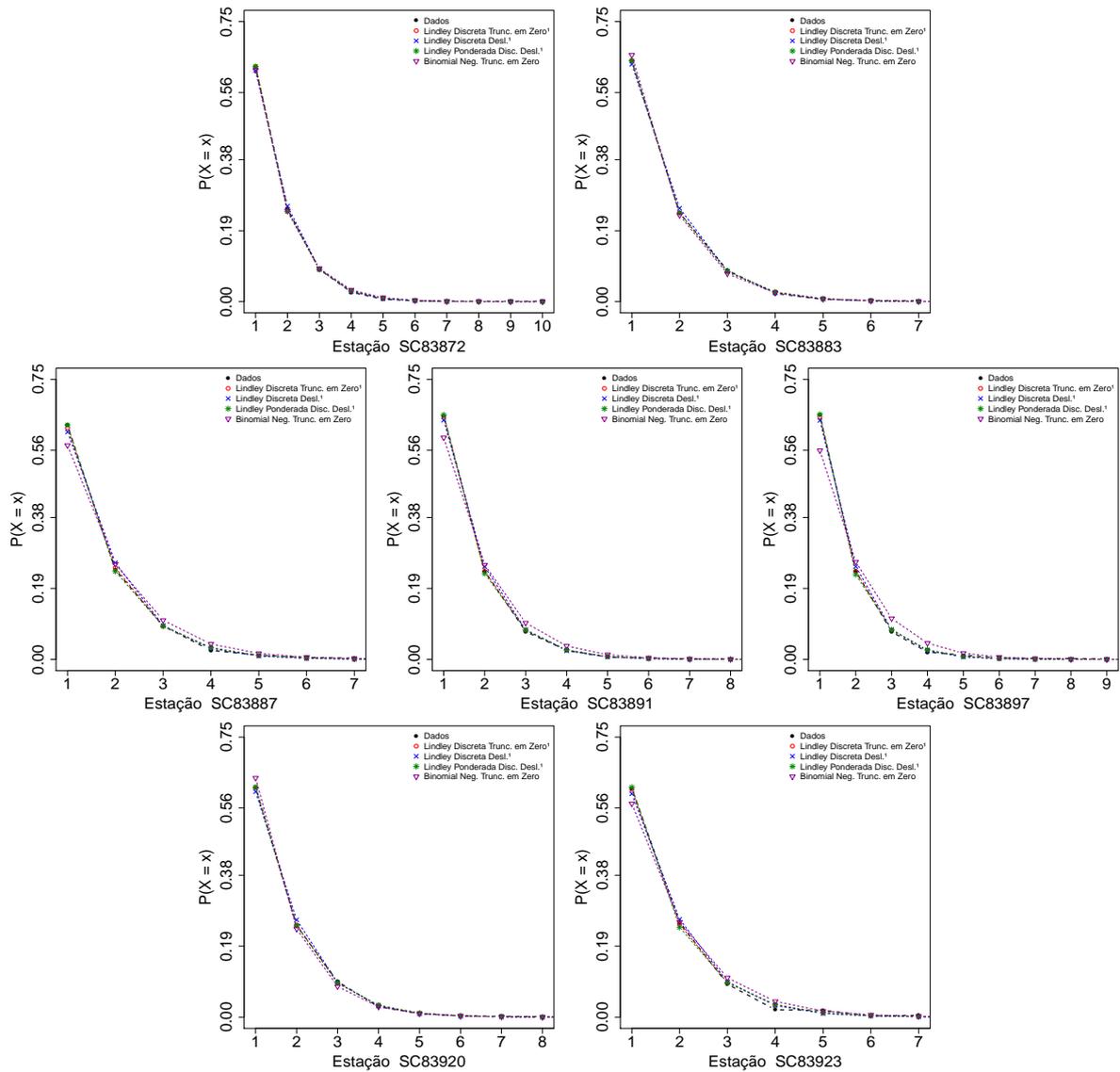


Figura 62 – Função de probabilidade empírica vs função de probabilidade ajustada das distribuições $LDTZ^1$, LDD^1 , $LPondD^1$ e $ZTNB$ para as estações de Santa Catarina.

análise de sequências de dias chuvosos como uma alternativa as distribuições tradicionais, como Poisson truncada em zero e binomial truncada em zero em termos de parsimonia e equação mais simples. Por outro lado, as distribuições Lindley discreta deslocada da origem (via função de sobrevivência) e Lindley ponderada deslocada da origem também se mostram boas em termo de ajuste para esta modelagem superando as distribuições clássicas consideradas.

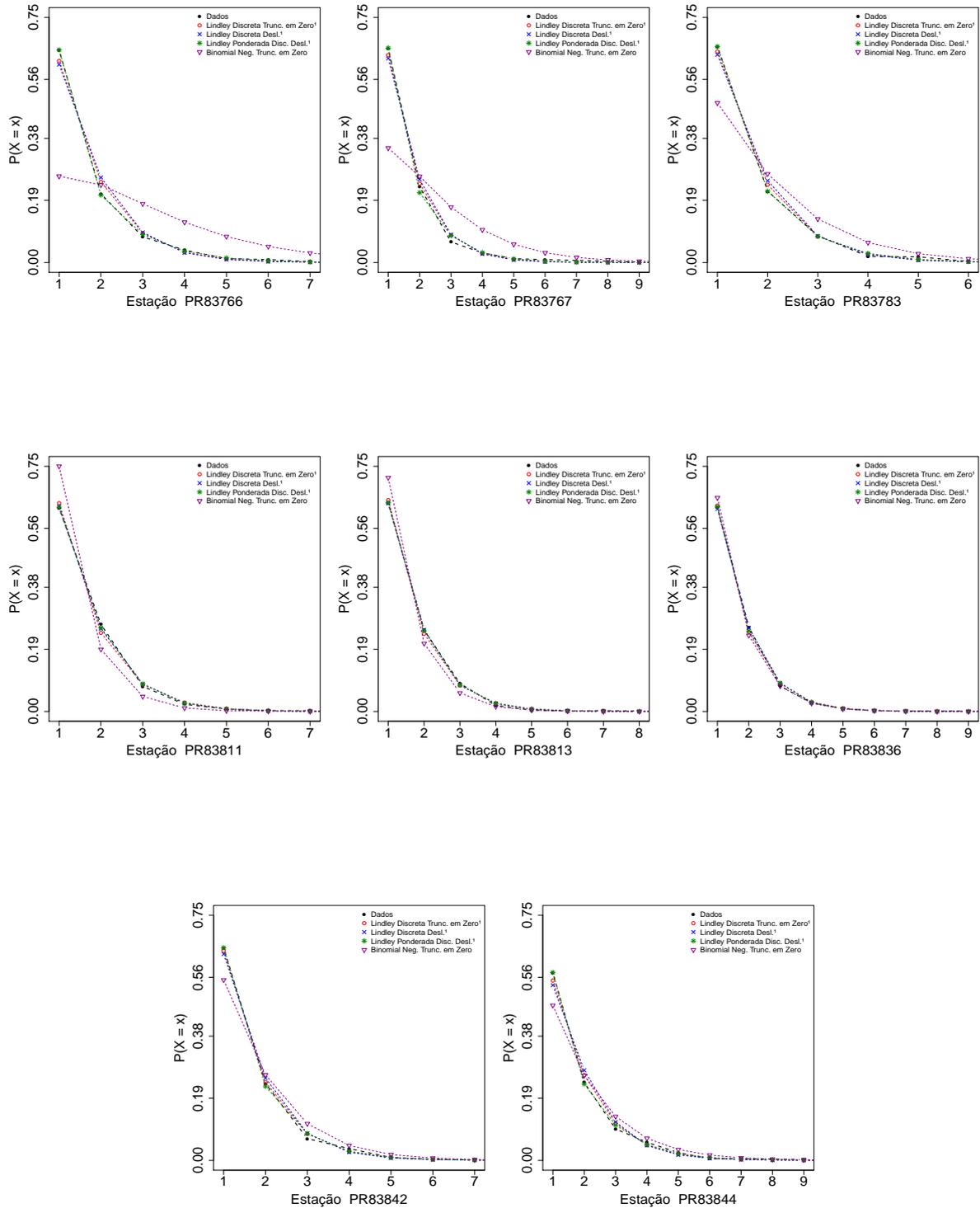


Figura 63 – Função de probabilidade empírica vs função de probabilidade ajustada das distribuições $LDTZ^1$, LDD^1 , $LPondD^1$ e $ZTNB$ para as estações do Paraná.

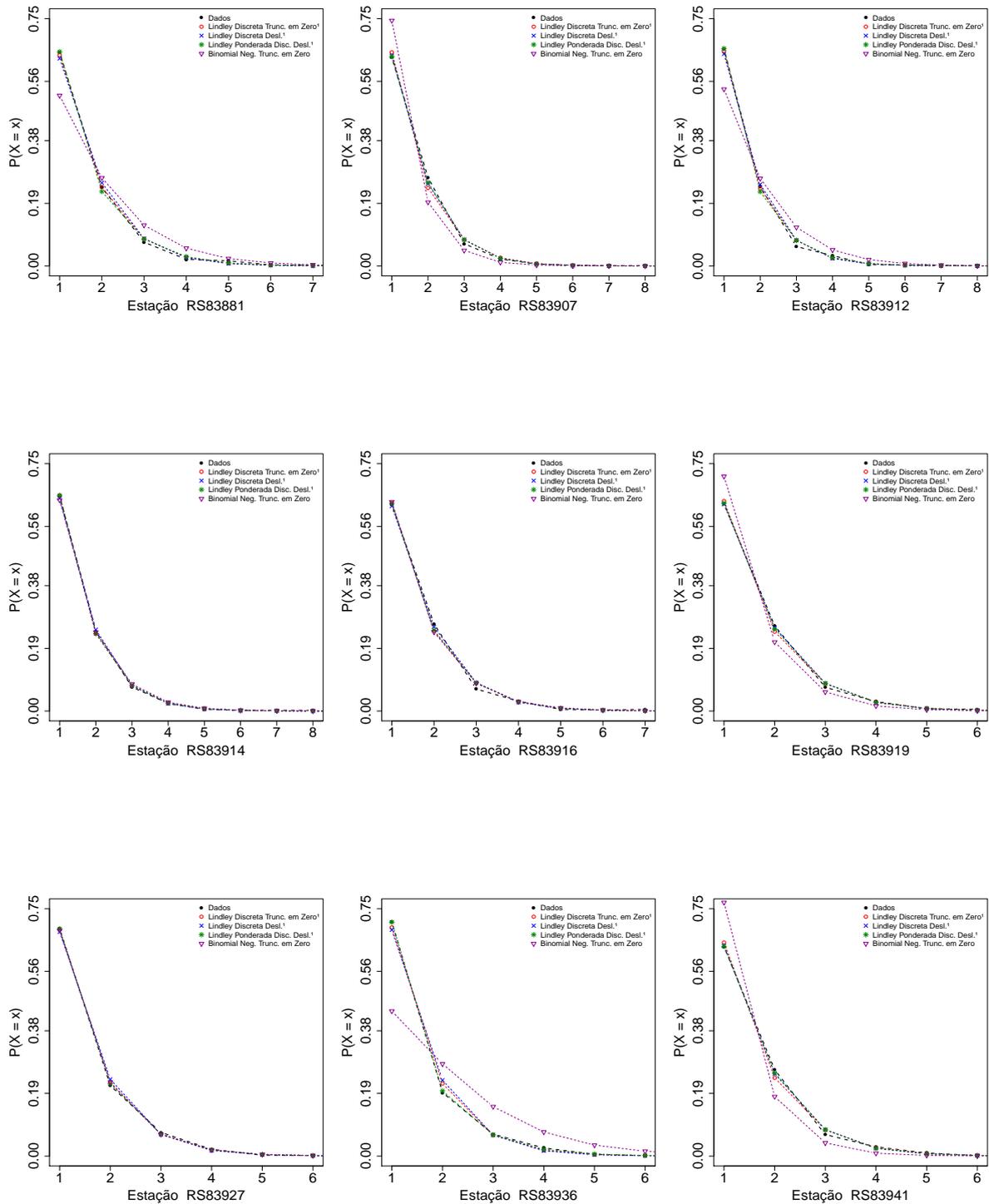


Figura 64 – Função de probabilidade empírica vs função de probabilidade ajustada das distribuições LDTZ¹, LDD¹, LPondD¹ e ZTNB para as estações do Rio Grande do Sul.

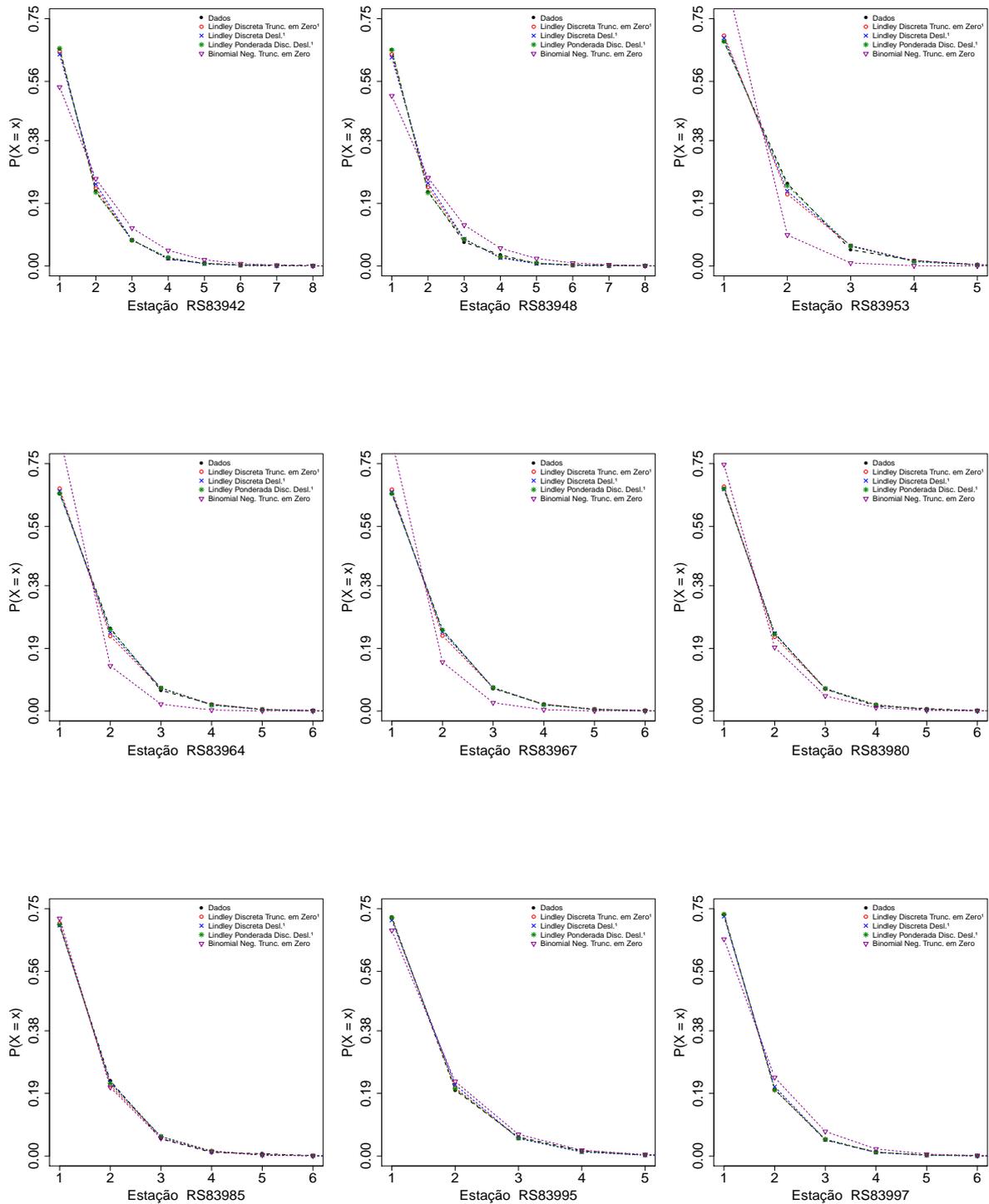


Figura 65 – Função de probabilidade empírica vs função de probabilidade ajustada das distribuições $LDTZ^1$, LDD^1 , $LPondD^1$ e $ZTNB$ para as estações do Rio Grande do Sul.

MODELANDO O COMPRIMENTO DAS SEQUÊNCIAS POSITIVAS

5.1 Introdução

Nas últimas décadas, estudos empíricos e a estrutura probabilística dos movimentos dos preços das ações do mercado financeiro tem atraído o interesse de muitos pesquisadores como Şensoy (2012), Adrian e Rosenberg (2008), Stockbridge (2008), Li e Gao (2006), Andersen et al. (2001), Cont (2001), Longin (1996), Tucker e Pond (1988), Akgiray e Booth (1987), entre outros. Conforme Şensoy (2012), esses estudos indicam que, de um ponto de vista estatístico, as variações aleatórias dos movimentos dos preços das ações partilham de propriedades estatísticas não-triviais.

Segundo Akgiray e Booth (1987), a evidência empírica indica que a distribuição empírica dos retornos das ações mostram graves desvios da normalidade e, entre as muitas distribuições de probabilidades de cauda longa são um “bom ajuste” aos dados e que estão implícitas em cenários econômicos razoáveis, duas hipóteses concorrentes se destacam: a primeira, conforme Mandelbrot (1997) e Fama (1965), é que o preço das ações seguem leis estáveis com expoentes característicos entre um e dois e a segunda vê as distribuições empíricas do retorno de ações como distribuições de cauda longa com variâncias finitas.

Ademais, de acordo com Cont (2001), tais estudos empíricos apresentaram propriedades comum em toda uma vasta gama de instrumentos, mercados e períodos de tempo. Essas propriedades são chamadas de fatos empíricos estilizados. Segundo Şensoy (2012), os pesquisadores têm acordos de vários fatos estilizados sobre os mercados financeiros, são eles: caudas pesadas nas distribuições de retorno de ativos, ausência de autocorrelações em retornos de ativos, cluster de volatilidade e assimetria entre subida e descida.

Todavia, conforme Ohira et al. (2002) e Li e Gao (2006), os estudos sobre os retornos das ações da bolsa de valores centram-se principalmente na análise com um período de tempo estabelecido como, por exemplo, hora, diário e semanal. Além disso, um outro indicador para

o movimento do preço das ações do mercado financeiro é o sinal do retorno, isto é, a subida ou descida dos preços das ações.

De acordo com Şensoy (2012) e Li e Gao (2006), o termo usado para medir o movimento dos preços referentes ao retorno de ações do mercado é “corrida”. Uma corrida pode ser definida como uma série consecutiva de movimentos de preços sem uma reversão de sinal, portanto, uma corrida positiva (negativa) é uma seqüência ininterrupta de retornos positivos (negativos) e esta corrida continua até que um retorno negativo (positivo) aparece. Matematicamente, podemos definir o retorno diário de um ativo por meio da expressão:

$$r_t = s_t - s_{t-1} \quad (5.1)$$

em que s_t indica o preço no instante t em que t pode representar o dia atual e s_{t-1} indica o preço no instante $t-1$ em que $t-1$ pode representar o dia anterior. Por exemplo, considerando, em particular, os preços de abertura diários da empresa Oracle no período de 2 de março de 2015 até 16 de março de 2015 (desconsiderando sábados e domingos) observa-se os valores 42.21, 42.94, 42.97, 42.50, 43.26, 42.63, 43.79 e 43.96. Tais valores geram os seguintes retornos diários 0.73, 0.03, -0.47, 0.75, -0.62, 1.16, 0.16 e -0.47. Os sinais desses retornos geram a seqüência “+ + - + - + +” que contém três corridas positivas e duas negativas. Assim, o comprimento das seqüências de dias em que as corridas foram positivas são 2, 1 e 2, respectivamente.

A construção de corridas é relativamente simples, porém, no melhor do nosso conhecimento, pouca investigação foi feita sobre elas em finanças. Além disso, para a análise das mesmas, a suposição de que os comprimentos das seqüências de dias em que as corridas foram positivas (negativas) para uma certa empresa são independentes e identicamente distribuídos. Por fim, autores como Fama (1965) investigou as corridas de várias unidades populacionais, e encontrou poucas evidências de violações de eficiência com base na dependência de série em retornos. Grafton (1981) realizou uma pesquisa semelhante para testar a hipótese de eficiência do mercado. Já Easley, Kiefer e O’Hara (1997) usaram corridas para examinar a dependência em dados intraday.

5.2 Distribuições de Probabilidade

Como a análise considerada aqui representa uma visão sobre o mercado financeiro, consideramos as distribuições Binomial negativa truncada em zero e Poisson truncada em zero por serem modelos tradicionais frequentemente utilizados na análise de dados de contagem truncados em zero e utilizamos também os modelos truncados em zero, deslocados da origem e ponderados do Capítulo 3 que no melhor do nosso conhecimento não foram utilizadas para este tipo de análise.

5.3 Os Dados

Assim como na Aplicação 1, a suposição de que as sequências são independentes e identicamente distribuídos foram avaliadas a partir dos valores das estatísticas dos testes de Bartels (BARTELS, 1982) e Mann-Kendall (MANN, 1945). Foram utilizadas as séries formadas pelo retorno diário de abertura das ações da bolsa de valores observadas em trinta empresas do ranque *The World's Biggest Public Companies* da Forbes de 2015 pois a mesma é baseada em uma pontuação composta de medidas igualmente ponderada de receitas, lucros, ativos e valor de mercado. As séries foram obtidas no portal de finanças do Yahoo, <<http://finance.yahoo.com/>>, e abrangem o período entre 1 de janeiro de 2005 a 31 de dezembro de 2015.

Na Tabela 61, para cada uma das empresas, são apresentadas o nome da empresa (Empresa), sigla no portal de finanças do Yahoo (Sigla), país de origem da empresa (País), número de sequências com corridas positivas (n_1), o maior, a média e variância comprimento da sequência das corridas positivas (Max, Média e Var).

Tabela 61 – Características do banco de dados para cada empresa.

Empresa	Sigla	País	n_1	Max	Média	Var
IBM	IBM	Estados Unidos	690	12	2.05	2.22
Oracle Corporation	ORCL	Estados Unidos	722	13	1.92	1.73
Petróleo Brasileiro S.A.	PBR	Brasil	693	10	2.01	2.17
Banco Bradesco S.A.	BBD	Brasil	689	15	2.04	2.22
Berkshire Hathaway	BRK-B	Estados Unidos	705	12	1.91	1.87
JPMorgan Chase	JPM	Estados Unidos	724	12	1.93	1.73
Exxon Mobil	XOM	Estados Unidos	714	10	2.01	2.04
PetroChina	0857.HK	China	720	8	1.85	1.60
General Electric	GE	Estados Unidos	709	13	1.89	1.70
Wells Fargo	WFC	Estados Unidos	720	10	1.97	1.80
Toyota Motor	TM	Japão	666	11	2.07	2.13
Apple	AAPL	Estados Unidos	714	11	2.08	2.63
Royal Dutch Shell	RDSB.L	Holanda	721	12	1.92	1.93
Alphabet Inc.	GOOGL	Estados Unidos	689	9	2.07	2.17
HSBC Holdings	HSBA.L	Reino Unido	726	10	1.90	1.72
Chevron	CVX	Estados Unidos	686	14	2.09	2.58
Wal-Mart Stores	WMT	Estados Unidos	724	10	1.92	1.72
Samsung Electronics	005930.KS	Coréia do Sul	733	7	1.85	1.36
Citigroup	C	Estados Unidos	718	11	1.83	1.60
China Mobile	0941.HK	China	791	11	1.77	1.41
Allianz	ALV.DE	Alemanha	724	10	2.05	2.01
Verizon Communications	VZ	Estados Unidos	699	10	2.00	2.12
Bank of America	BAC	Estados Unidos	733	9	1.83	1.41
Sinopec	0386.HK	China	718	8	1.81	1.40
Microsoft	MSFT	Estados Unidos	694	12	1.99	1.89

Intel	INTC	Estados Unidos	690	9	2.04	2.05
Volkswagen AG	VLKAY	Alemanha	671	13	2.03	2.34
Honda Motor Co., Ltd.	HMC	Japão	675	12	2.04	2.10
McDonald's Corp.	MCD	Estados Unidos	726	13	1.97	1.89
Johnson & Johnson	JNJ	Estados Unidos	718	12	1.94	1.85

De acordo com a [Tabela 61](#), nota-se a empresa em que se teve o maior número de sequências com corridas positivas é a empresa China Mobile com 791 sequências e a empresa que teve o menor número de sequências com corridas positivas é a empresa Toyota Motor com 666 sequências. Quanto ao maior comprimento de sequência, o destaque vai para uma empresa brasileira, Banco Bradesco S.A., com uma sequência de tamanho 15. Por fim, pode-se perceber que os dados possuem superdispersão e também a subdispersão, o que induz o uso de distribuições que comportam tanto superdispersão quanto a subdispersão para a análise dos mesmos.

5.4 Resultados e Discussões

Segundo o teste de Bartels, em nenhuma das séries a hipótese de comprimentos i.i.d. foi rejeitada em nível de significância de 1%. Já no teste de Mann-Kendall, em 2 das 30 séries a mesma hipótese foi rejeitada adotando o mesmo nível de significância dos testes anteriores. Estas séries são provenientes das empresas Petróleo Brasileiro S.A. e Wells Fargo.

Para as séries no período considerado, o ajuste das distribuições Lindley e suas variantes truncadas em zero, deslocadas da origem e ponderadas propostas neste trabalho foi comparado ao ajuste das distribuições Poisson truncada em zero e Binomial negativa truncada em zero. Os parâmetros foram estimados pelo método da máxima verossimilhança e um resumo do número de séries ajustadas para cada distribuição, considerando os níveis de significância de 1% e 5%, o AIC, AICc e o BIC são apresentados na [62](#). O ajuste foi avaliado pelo teste qui-quadrado de aderência e os valores AIC, AICc e BIC foram usados como critérios de discriminação.

De acordo com a [Tabela 62](#), concluímos que, dentre todas as distribuições propostas neste trabalho, os melhores ajustes foram da distribuição Lindley discreta truncada em zero (via função de sobrevivência) tendo o menor AIC em 11 das 30 séries consideradas, o menor AICc em 11 séries e o menor BIC em 20 séries, o da distribuição Lindley ponderada deslocada da origem que obteve o menor AIC e AICc em 8 séries e o menor BIC em 5 séries e da distribuição Lindley potência discreta truncada em zero que obteve o menor AIC em 7 séries, o menor AICc em 1 séries e o menor BIC em 4 das séries. Além disso, essas distribuições se ajustaram entre 90% e 97% das séries em nível de significância de 1% e entre 70% a 97% em nível de significância de 5%. É importante destacar também que a distribuição Lindley discreta truncada em zero é uni-paramétrica o que favorece seu uso perante a distribuição Lindley potência discreta.

Tabela 62 – Número de séries ajustadas para cada distribuição em relação ao comprimento da sequência de corridas positivas.

Distribuição	Nível de significância		AIC	AICc	BIC
	$\alpha = 1\%$	$\alpha = 5\%$			
Lindley Discreta Truncada em Zero ¹	27 (90.00%)	21 (70.00%)	11 (36.67%)	11 (36.67%)	20 (66.67%)
Lindley Discreta Deslocada da Origem ¹	15 (50.00%)	13 (43.3%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Lindley Discreta Ponderada ¹	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Lindley Discreta Truncada em Zero ²	22 (73.33%)	19 (63.33%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Lindley Discreta Deslocada da Origem ²	4 (13.33%)	1 (3.33%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Lindley Discreta Ponderada ²	4 (13.33%)	1 (3.33%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Lindley Potência Discreta Truncada em Zero ¹	29 (96.67%)	28 (93.33%)	7 (23.33%)	1 (3.33%)	4 (13.33%)
Lindley Potência Discreta Deslocada da Origem ¹	29 (96.67%)	29 (96.67%)	1 (3.33%)	0 (0.00%)	0 (0.00%)
Lindley Ponderada Discreta Truncada em Zero ¹	29 (96.67%)	29 (96.67%)	1 (3.33%)	0 (0.00%)	1 (3.33%)
Lindley Ponderada Discreta Deslocada da Origem ¹	29 (96.67%)	29 (96.67%)	8 (26.67%)	8 (26.67%)	5 (16.67%)
Poisson Truncada em Zero	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Binomial Negativa Truncada em Zero	30 (100.0%)	28 (93.33%)	2 (6.67%)	12 (40.00%)	0 (0.00%)

¹: Via função de sobrevivência.

²: Via séries.

Já os modelos tradicionais, obtivemos que a distribuição Poisson truncada em zero não se ajustou em nenhuma das séries tanto a 1% quanto a 5% e também não obteve o menor AIC, AICc e BIC em nenhuma das séries. Para a Binomial negativa truncada em zero o ajuste foi em todas as séries a 1% e 28 séries a 5% e teve o menor AIC em 2 séries, o menor AICc em 12 séries e nenhuma das séries obteve o menor BIC. Dentre ambos os modelos, apenas a Binomial negativa truncada em zero oferece condições para a concorrência contra as distribuições Lindley discreta truncada em zero (via função de sobrevivência), Lindley ponderada deslocada da origem e Lindley potência discreta truncada em zero.

Por outro lado, a distribuição Lindley discreta ponderada (via sobrevivência) não se ajustou em nenhuma das séries a 1% e a 5%. Além disso, as distribuições Lindley discreta deslocada da origem (via série infinita) e Lindley discreta ponderada (via séries) também obtiveram péssimos ajustes sendo ajustadas em apenas 4 séries das 30 séries consideradas a 1% e uma série a 5%.

CONCLUSÕES E PROPOSTAS FUTURAS

6.1 Conclusões

Neste trabalho, concluímos que os métodos de discretização estudados, principalmente o método baseado na função de sobrevivência, se mostraram favoráveis para a construção de novos modelos discretos, baseados nas distribuições Lindley, Lindley potência e Lindley ponderada, como alternativas para as distribuições discretas tradicionais presente na literatura. Estes novos modelos foram estudadas em detalhes e algumas de suas propriedades discutidas.

Para cada um dos modelos verificou-se que comportamento da função de probabilidade é unimodal e o comportamento da função de risco é crescente em alguns modelos e decrescente em outros. Além disso, nos modelos uni-paramétricos, a esperança e a variância possuem uma expressão explícita o que é uma vantagem perante os modelos bi-paramétricos que não possuem expressão explícita para as mesmas.

Os estudos de simulação conduzidos para avaliar o vício e o erro-quadrático-médio dos estimadores de máxima verossimilhança mostraram que, em quase todos os modelos, os estimadores de máxima verossimilhança são positivamente viesados. Nos modelos Lindley potência discreto e Lindley potência discreto truncado em zero observa-se um viés positivo-negativo para o parâmetro β , e no modelo Lindley potência discreto deslocado da origem observa-se que o viés é negativo para o parâmetro β .

Embora não tenha sido ilustrado no presente trabalho, a informação de Fisher para o modelo Lindley discreto baseado em uma série infinita possui expressão analítica e foi ilustrada no trabalho intitulado *A Comparative Study Between Two Discrete Lindley Distributions by Josmar Mazucheli and Oliveira, R. P.* que está em processo de submissão. Para todos os outros modelos podemos obter a mesma por meio das aproximações:

- No caso uni-paramétrico:

$$\mathbb{E} \left(-\frac{\partial^2 \ell}{\partial \beta^2} \right) = -\frac{\partial^2 \ell}{\partial \beta^2} \Big|_{(\hat{\beta})} \quad (6.1)$$

em que β é o parâmetro do modelo.

- No caso bi-paramétrico:

$$\mathbb{E} \left(-\frac{\partial^2 \ell}{\partial \alpha^2} \right) = -\frac{\partial^2 \ell}{\partial \alpha^2} \Big|_{(\hat{\alpha}, \hat{\beta})}, \mathbb{E} \left(-\frac{\partial^2 \ell}{\partial \alpha \partial \beta} \right) = -\frac{\partial^2 \ell}{\partial \alpha \partial \beta} \Big|_{(\hat{\alpha}, \hat{\beta})} \quad \text{e} \quad \mathbb{E} \left(-\frac{\partial^2 \ell}{\partial \beta^2} \right) = -\frac{\partial^2 \ell}{\partial \beta^2} \Big|_{(\hat{\alpha}, \hat{\beta})}$$

em que α e β são os parâmetros do modelo.

Nas aplicações envolvendo dados da literatura, obtemos um ajuste superior (Aplicação 1) e/ou equivalente (Aplicação 2) as distribuições discretas clássicas equivalente em termos de qualidade de ajuste e parsimonia ilustrando que os modelos aqui propostos podem ser utilizados como alternativa a análise de dados de natureza discreta.

Em relação aos dados climatológicos, a distribuição proposta que mais se adaptou a modelagem das sequências de dias chuvosos da região sul do Brasil foi a Lindley ponderada discreta deslocada da origem em que teve seu ajuste superior/equivalente, em termos do valor esperado, a todas as outras consideradas na análise. Todavia, a distribuição Lindley discreta truncada em zero (via função de sobrevivência), em termos de qualidade de ajuste e parsimonia, ofereceu um ajuste equivalente em alguns casos e superior em outros.

Por exemplo, considerando a cidade de Maringá, obtemos que a probabilidade empírica de se observar 4 dias consecutivos de chuva é de, aproximadamente, 0.029 ou 2.9%. Essa probabilidade de acordo com a distribuição Lindley discreta truncada em zero (via função de sobrevivência) é de, aproximadamente, 0.028 ou 2.8% e de acordo com a distribuição Lindley ponderada discreta deslocada da origem é de, aproximadamente, 0.031 ou 3.1%. Todavia, para a distribuição binomial negativa truncada em zero, a mesma probabilidade é de 0.1 ou 10% e para a distribuição Poisson truncada em zero a probabilidade é de 0.021 ou 2.1%. Desta forma, concluímos que os modelos propostos oferecem uma qualidade de ajuste superior perante os modelos clássicos de Poisson truncado em zero e binomial negativo truncado em zero considerados na análise. E, dentre os propostos, a distribuição Lindley discreta truncada em zero (via função de sobrevivência) é a “vencedora” em termos de parsimonia.

Por fim, em relação aos dados financeiros, concluímos que os modelos propostos, de acordo com os critérios de discriminação considerados, foram selecionados mais vezes em comparação os clássicos o que induz que os mesmos oferecem uma modelagem com melhor qualidade para esta estrutura de dados e também uma melhor previsão a respeito da sequência de dias em que o retorno foi positivo para um ativo.

6.2 Propostas Futuras

Com o objetivo de dar continuidade ao estudo sobre a discretização de uma variável aleatória contínua, estabelecemos algumas algumas propostas futuras. A primeira delas se refere a estudar outros métodos de discretização como, por exemplo, o método de duas fases compostas (CHAKRABORTY, 2015), o método via função de risco reversa (GHOSH; ROY;

CHANDRA, 2013) e o método de equações diferenciais de Pearson (PEARSON, 1895). E também realizar uma comparação entre os principais métodos de discretização ilustrando qual possui uma melhor aplicabilidade e flexibilidade.

A segunda proposta se refere a discretização de outras variantes da distribuição Lindley como, por exemplo, a Gamma-Lindley (ZEGHDOUDI; NEDJAR, 2015), a Marshall-Olkin Lindley estendida (GHITANY et al., 2012), a Lindley de dois parâmetros (SHANKER; SHARMA; SHANKER, 2013), a Lindley inversa potência (BARCO; MAZUCHELI; JANEIRO, 2016), a Lindley inversa (SHARMA et al., 2015), a Lindley transmutada (MEROVCI, 2013), a Quasi-Lindley (SHANKER; MISHRA, 2013), entre outras. Além disso, investigar algumas propriedades matemáticas importantes para estes novos modelos e utilizar os mesmos na modelagem de dados de sobrevivência.

Por fim, trabalhar com a inflação de zeros para estes novos modelos como uma alternativa aos modelos inflacionados de zero usuais como, por exemplo, o modelo de Poisson inflacionado de zeros, e também com os chamados “hurdle models” já que uma das limitações dos modelos convencionais de contagem é que os zeros e os “non-zeros” (positivos) são, em geral, assumidos como parte do mesmo processo de geração de dados.

REFERÊNCIAS

- ADRIAN, T.; ROSENBERG, J. Stock returns and volatility: Pricing the short-run and long-run components of market risk. *The Journal of Finance*, Wiley Online Library, v. 63, n. 6, p. 2997–3030, 2008.
- AI, M. *Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2*. 2014.
- AKGIRAY, V.; BOOTH, G. G. Compound distribution models of stock returns: an empirical comparison. *Journal of Financial Research*, Wiley Online Library, v. 10, n. 3, p. 269–280, 1987.
- AL-KHASAWNEH, M. F. Estimating the negative binomial dispersion parameter. *Asian Journal of Mathematics & Statistics*, v. 3, n. 1, p. 1–15, 2010.
- ALMALKI, S. J.; NADARAJAH, S. Modifications of the weibull distribution: a review. *Reliability Engineering & System Safety*, Elsevier, v. 124, p. 32–55, 2014.
- ALMALKI, S. J.; NADARAJAH, S. A new discrete modified weibull distribution. *Reliability, IEEE Transactions on*, IEEE, v. 63, n. 1, p. 68–80, 2014.
- ANAGNOSTOPOULOU, C. et al. Spatial and temporal analysis of dry spells in greece. *Theoretical and Applied Climatology*, Springer, v. 74, n. 1-2, p. 77–91, 2003.
- ANDERSEN, T. G. et al. The distribution of realized stock return volatility. *Journal of financial economics*, Elsevier, v. 61, n. 1, p. 43–76, 2001.
- ARAGÓN, J.; EBERLY, D.; EBERLEY, S. Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution. *Statistics & Probability Letters*, v. 15, n. 5, p. 375–379, 1992.
- BAKOUCH, H. S.; JAZI, M. A.; NADARAJAH, S. A new discrete distribution. *Statistics*, Taylor & Francis, v. 48, n. 1, p. 200–240, 2014.
- BARBIERO, A. Parameter estimation for type iii discrete weibull distribution: a comparative study. *Journal of Probability and Statistics*, Hindawi Publishing Corporation, v. 2013, 2013.
- BARCO, K. V. P.; MAZUCHELI, J.; JANEIRO, V. The inverse power lindley distribution. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, 2016.

- BARRON, J. et al. Dry spell analysis and maize yields for two semi-arid locations in east africa. *Agricultural and forest meteorology*, Elsevier, v. 117, n. 1, p. 23–37, 2003.
- BARTELS, R. The rank version of von neumann's ratio test for randomness. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 77, n. 377, p. 40–46, 1982.
- BI, Z.; FALOUTSOS, C.; KORN, F. The dgx distribution for mining massive, skewed data. In: ACM. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.], 2001. p. 17–26.
- BÖHNING, D.; KUHNERT, R. Equivalence of truncated count mixture distributions and mixtures of truncated count distributions. *Biometrics*, v. 62, n. 4, p. 1207–1215, 2006.
- BOULLE, M. Khiops: A statistical discretization method of continuous attributes. *Machine learning*, Springer, v. 55, n. 1, p. 53–69, 2004.
- BRACQUEMOND, C.; GAUDOIN, O. A survey on discrete lifetime distributions. *International Journal of Reliability, Quality and Safety Engineering*, World Scientific, v. 10, n. 01, p. 69–98, 2003.
- BRUNELLO, G.; NAKANO, E. Inferência bayesiana no modelo Weibull discreto em dados com presença de censura. *TEMA (São Carlos)*, SciELO Brasil, v. 16, n. 2, p. 97–110, 2015.
- CAEIRO, F.; MATEUS, A.; CAEIRO, M. F. Package 'randtests'. 2014.
- CAMERON, A. C.; TRIVEDI, P. K. *Regression Analysis of Count Data*. [S.l.]: Cambridge University Press, 1998. v. 30. (Econometric Society Monographs, v. 30).
- CHAKRABORTY, S. Generating discrete analogues of continuous probability distributions—a survey of methods and constructions. *JOURNAL of Statistical Distributions and Applications*, Springer, v. 2, n. 1, p. 1–30, 2015.
- CHAKRABORTY, S.; CHAKRAVARTY, D. Discrete gamma distributions: properties and parameter estimations. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 41, n. 18, p. 3301–3324, 2012.
- COLLETT, D. *Modelling Survival Data in Medical Research*. second. New York: Chapman and Hall, 2003.
- CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. Taylor & Francis, 2001.
- DEKA, S.; BORAH, M.; KAKATY, S. Statistical modeling of wet and dry spell frequencies over north-east india. *Journal of Applied and Natural Science*, v. 2, n. 1, p. 42–47, 2010.
- DENI, S. M.; JEMAIN, A. A. Mixed geometric truncated poisson model for sequences of wet days. *Journal of Applied Sciences*, v. 8, n. 21, p. 3975–3980, 2008.
- DENI, S. M.; JEMAIN, A. A. Mixed geometric truncated Poisson model for sequences of wet days. *Journal of Applied Sciences*, v. 8, n. 21, p. 3975–3980, 2008.
- DENI, S. M.; JEMAIN, A. A. Fitting the distribution of dry and wet spells with alternative probability models. *Meteorology and Atmospheric Physics*, v. 104, n. 1-2, p. 13–27, 2009. ISSN 0177-7971.

- DENI, S. M.; JEMAIN, A. A. Mixed log series geometric distribution for sequences of dry days. *Atmospheric Research*, v. 92, n. 2, p. 236 – 243, 2009.
- DENI, S. M.; JEMAIN, A. A. Comparison between mixed probability models and Markov chain models for weekly dry and wet spells in Peninsular Malaysia. *Proceedings of the World Congress on Engineering*, v. 1, 2012.
- DENI, S. M.; JEMAIN, A. A.; IBRAHIM, K. The spatial distribution of wet and dry spells over Peninsular Malaysia. *Theoretical and Applied Climatology*, v. 94, n. 3-4, p. 163–173, 2008.
- DENI, S. M.; JEMAIN, A. A.; IBRAHIM, K. Mixed probability models for dry and wet spells. *Statistical Methodology*, v. 6, n. 3, p. 290 – 303, 2009.
- DENI, S. M.; JEMAIN, A. A.; IBRAHIM, K. The best probability models for dry and wet spells in Peninsular Malaysia during monsoon seasons. *International Journal of Climatology*, v. 30, n. 8, p. 1194–1205, 2010. ISSN 1097-0088.
- DOBI-WANTUCH, I.; MIKA, J.; SZEIDL, L. Modelling wet and dry spells with mixture distributions. *Meteorology and Atmospheric Physics*, v. 73, n. 3-4, p. 245–256, 2000. ISSN 0177-7971.
- DORAY, L. G.; LUONG, A. Efficient estimators for the good family. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 26, n. 3, p. 1075–1088, 1997.
- EASLEY, D.; KIEFER, N. M.; O'HARA, M. One day in the life of a very common stock. *Review of Financial Studies*, Soc Financial Studies, v. 10, n. 3, p. 805–835, 1997.
- ENGLEHARDT, J. D.; LI, R. The discrete Weibull distribution: an alternative for correlated counts with confirmation for microbial counts in water. *Risk Analysis*, Wiley Online Library, v. 31, n. 3, p. 370–381, 2011.
- FAMA, E. F. The behavior of stock-market prices. *The journal of Business*, JSTOR, v. 38, n. 1, p. 34–105, 1965.
- GASTWIRTH, J. et al. Noguchi, k. lawstat: Tools for biostatistics. *Public Policy, and Law*, 2015.
- GHITANY, M. et al. Marshall-olkin extended lindley distribution and its application. *International Journal of Applied Mathematics*, v. 25, n. 5, p. 709–721, 2012.
- GHITANY, M.; AL-MUTAIRI, D.; NADARAJAH, S. Zero-truncated poisson–lindley distribution and its application. *Mathematics and Computers in Simulation*, Elsevier, v. 79, n. 3, p. 279–287, 2008.
- GHITANY, M. et al. A two-parameter weighted lindley distribution and its applications to survival data. *Mathematics and Computers in Simulation*, Elsevier, v. 81, n. 6, p. 1190–1201, 2011.
- GHITANY, M. E. et al. Power lindley distribution and associated inference. *Computational Statistics & Data Analysis*, Elsevier, v. 64, p. 20–33, 2013.

- GHOSH, T.; ROY, D.; CHANDRA, N. K. Reliability approximation through the discretization of random variables using reversed hazard rate function. *International Journal of Mathematical, Computational, Statistical, Natural and Physical Engineering*, International Science Index, v. 7, n. 4, p. 96 – 100, 2013.
- GÓMEZ-DÉNIZ, E. Another generalization of the geometric distribution. *Test*, Springer, v. 19, n. 2, p. 399–415, 2010.
- GÓMEZ-DÉNIZ, E.; CALDERÍN-OJEDA, E. The discrete lindley distribution: properties and applications. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 81, n. 11, p. 1405–1416, 2011.
- GOOD, I. J. The population frequencies of species and the estimation of population parameters. *Biometrika*, Biometrika Trust, v. 40, n. 3-4, p. 237–264, 1953.
- GOOD, P. I. *Permutation, parametric, and bootstrap tests of hypotheses*. [S.l.]: Springer Science & Business Media, 2006.
- GRAFTON, R. Algorithm as 157: The runs-up and runs-down tests. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, JSTOR, v. 30, n. 1, p. 81–85, 1981.
- GROGGER, J. T.; CARSON, R. T. Models for truncated counts. *Journal of Applied Econometrics*, v. 6, n. 3, p. 225–238, 1991.
- HAIGHT, F. A. Queueing with balking. *Biometrika*, JSTOR, v. 44, n. 3/4, p. 360–369, 1957.
- HAMADA, M. S. et al. *Bayesian reliability*. [S.l.]: Springer, New York, 2008. (Springer Series in Statistics).
- HILBE, J. M. *Negative binomial regression*. 2th. ed. [S.l.]: Cambridge University Press, Cambridge, 2011.
- HUSSAIN, T.; AHMAD, M. Discrete inverse rayleigh distribution. *Pak. J. Statist*, v. 30, n. 2, p. 203–222, 2014.
- IBGE. *Geografia do Brasil*. [S.l.]: IBGE, 1977. v. 5. 35–36 p.
- INUSAH, S.; KOZUBOWSKI, T. J. A discrete analogue of the laplace distribution. *Journal of Statistical Planning and Inference*, Elsevier Science, v. 136, 2006.
- JAIN, G. C.; CONSUL, P. C. A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics*, v. 21, n. 4, p. 501–513, 1971.
- JAZI, M. A.; LAI, C.-D.; ALAMATSAZ, M. H. A discrete inverse weibull distribution and estimation of its parameters. *Statistical Methodology*, Elsevier Science, v. 7, p. 121–132, 2010.
- JODRA, P. Computer generation of random variables with lindley or poisson–lindley distribution via the lambert w function. *Mathematics and Computers in Simulation*, v. 81, n. 4, p. 851 – 859, 2010.
- JOHNSON, N. L.; KEMP, A. W.; KOTZ, S. *Univariate discrete distributions*. Third. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2005. xx+646 p. (Wiley Series in Probability and Statistics).

- KALBFLEISCH, J. D.; PRENTICE, R. L. *The Statistical Analysis of Failure Time Data*. 2nd. ed. New York, NY: Wiley, 2002.
- KEILSON, J.; GERBER, H. Some results for discrete unimodality. *Journal of the American Statistical Association*, American Statistical Association, v. 66, 1971.
- KEMP, A. W. Characterizations of a discrete normal distribution. *Journal of Statistical Planning and Inference*, v. 63, n. 2, p. 223 – 229, 1997. In Honor of C.R. Rao.
- KEMP, A. W. *Classes of discrete lifetime distributions*. Taylor & Francis, 2004.
- KEMP, A. W. The discrete half-normal distribution. In: *Advances in mathematical and statistical modeling*. [S.l.]: Springer, 2008. p. 353–360.
- KHAN, M. A.; KHALIQUE, A.; ABOUAMMOH, A. On estimating parameters in a discrete Weibull distribution. *Reliability, IEEE Transactions on*, IEEE, v. 38, n. 3, p. 348–350, 1989.
- KLEIN, J. P.; MOESCHBERGER, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag, 1997.
- KOZUBOWSKI, T. J.; INUSAH, S. A skew laplace distribution on integers. *Annals of the Institute of Statistical Mathematics*, Springer, v. 58, n. 3, p. 555–571, 2006.
- KRISHNA, H.; PUNDIR, P. S. Discrete maxwell distribution. *Interstat*, 2007.
- KRISHNA, H.; PUNDIR, P. S. Discrete burr and discrete pareto distributions. *Statistical Methodology*, Elsevier, v. 6, n. 2, p. 177–188, 2009.
- KULASEKERA, K. Approximate mle's of the parameters of a discrete Weibull distribution with type i censored data. *Microelectronics Reliability*, Elsevier, v. 34, n. 7, p. 1185–1188, 1994.
- KULASEKERA, K.; TONKYN, D. W. A new discrete distrlbution, with applications to survival, dispersal and dispersion. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 21, n. 2, p. 499–518, 1992.
- LAI, C. Issues concerning constructions of discrete lifetime models. *Quality Technology and Quantitative Management*, v. 10, n. 2, p. 251–262, 2013.
- LANA, X. et al. Statistical distributions and sampling strategies for the analysis of extreme dry spells in catalonia (ne spain). *Journal of hydrology*, Elsevier, v. 324, n. 1, p. 94–114, 2006.
- LAWLESS, J. F. *Statistical models and methods for lifetime data*. second. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2003. xxii+630 p. (Wiley Series in Probability and Statistics).
- LEE, E. T.; WANG, J. W. *Statistical methods for survival data analysis*. Third. Hoboken, NJ: [s.n.], 2003. (Wiley Series in Probability and Statistics).
- LI, H.; GAO, Y. Statistical distribution of stock returns runs. In: *Econophysics of Stock and other Markets*. [S.l.]: Springer, 2006. p. 59–66.
- LINDLEY, D. V. Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 20, p. 102–107, 1958.

- LISMAN, J.; ZUYLEN, M. V. Note on the generation of most probable frequency distributions. *Statistica Neerlandica*, Wiley Online Library, v. 26, n. 1, p. 19–23, 1972.
- LONGIN, F. M. The asymptotic distribution of extreme stock market returns. *Journal of business*, JSTOR, p. 383–408, 1996.
- MANDELBROT, B. B. *The variation of certain speculative prices*. [S.l.]: Springer, 1997.
- MANN, H. B. Nonparametric tests against trend. *Econometrica*, v. 13, n. 3, p. 245–259, 1945.
- MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, Biometrika Trust, v. 84, n. 3, p. 641–652, 1997.
- MATHUGAMA, S.; PEIRIS, T. Critical evaluation of dry spell research. *Int J Basic Appl Sci*, Citeseer, v. 6, n. 11, p. 153–160, 2011.
- MAZUCHELI, J.; ACHCAR, J. A. The lindley distribution applied to competing risks lifetime data. *Comput. Methods Prog. Biomed.*, Elsevier North-Holland, Inc., v. 104, n. 2, p. 188–192, 2011.
- MCCALLA, R.; DAY, E.; MILLWARD, H. The relative concept of warm and cold spells of temperature: methodology and application. *Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B*, Springer, v. 25, n. 4, p. 323–336, 1978.
- MEEKER, W. Q.; ESCOBAR, L. A. *Statistical Methods for Reliability Data*. New York: John Wiley & Sons, 1998.
- MEROVCI, F. Transmuted lindley distribution. *International Journal of Open Problems in Computer Science & Mathematics*, v. 6, 2013.
- MURTHY, D. P.; XIE, M.; JIANG, R. *Weibull models*. [S.l.]: John Wiley & Sons, 2004. v. 505.
- NAKAGAWA, T. Discrete extreme distributions. *IEEE Transactions on Reliability*, v. 5, p. 367–368, 1978.
- NAKAGAWA, T.; OSAKI, S. The discrete Weibull distribution. *IEEE Transactions on Reliability*, v. 5, p. 300–301, 1975.
- NEKOUKHO, V. et al. Discrete beta-exponential distribution. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 44, n. 10, p. 2079–2091, 2015.
- NEKOUKHO, V.; ALAMATSAZ, M. H.; BIDRAM, H. A discrete analog of the generalized exponential distribution. *Communication in Statistics- Theory and Methods*, Taylor and Francis Group, v. 41, p. 2000–2013, 2012.
- NEKOUKHO, V.; ALAMATSAZ, M. H.; BIDRAM, H. Discrete generalized exponential distribution of a second type. *Statistics*, Taylor and Francis Group, v. 47, p. 876–887, 2013.
- OHIRA, T. et al. Predictability of currency market exchange. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 308, n. 1, p. 368–374, 2002.

- PADGETT, W.; SPURRIER, J. D. On discrete failure models. *Reliability, IEEE Transactions on*, IEEE, v. 34, n. 3, p. 253–256, 1985.
- PATIL, G. P.; RAO, C. R. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, JSTOR, p. 179–189, 1978.
- PEARSON, K. Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, JSTOR, p. 343–414, 1895.
- PHYO, I. Use of a chain binomial in the epidemiology of caries. *Journal of Dental Research*, International and American Association of Dental Research, 1973.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015.
- RAO, C. R. On discrete distributions arising out of methods of ascertainment. *Sankhyā: The Indian Journal of Statistics, Series A*, JSTOR, p. 311–324, 1965.
- ROY, D. Discretization of continuous distributions with an application to stress-strength reliability. *Bulletin of the Calcutta Statistical Association*, Calcutta: The Association, [1947]-, v. 52, n. 205-208, p. 295–314, 2002.
- ROY, D. The discrete normal distribution. *Communication in Statistics- Theory and Methods*, Taylor and Francis Group, v. 32, p. 1871–1883, 2003.
- ROY, D. Discrete Rayleigh distribution. *Reliability, IEEE Transactions on*, IEEE, v. 53, n. 2, p. 255–260, 2004.
- ROY, D.; DASGUPTA, T. A discretizing approach for evaluating reliability of complex systems under stress-strength model. *IEEE Transactions on Reliability*, IEEE, v. 50, 2001.
- ROY, D.; GHOSH, T. A new discretization approach with application in reliability estimation. *IEEE Transactions on Reliability*, IEEE, v. 58, 2009.
- RUPERT, M. *Survival Analysis*. [S.l.]: John Wiley & Sons, 2011. v. 66.
- SAHA, K.; PAUL, S. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics*, v. 61, n. 1, p. 179–185, 2005.
- SANKARAN, M. The discrete Poisson-Lindley distribution. *Biometrics*, v. 26, p. 145–149, 1970.
- SATO, H. et al. A new defect distribution metrology with a consistent discrete exponential formula and its applications. *Semiconductor Manufacturing, IEEE Transactions on*, IEEE, v. 12, n. 4, p. 409–418, 1999.
- SEN, Z. *Spatial modeling principles in earth sciences*. [S.l.]: Springer Science & Business Media, 2009.
- ŞENSOY, A. Analysis on runs of daily returns in istanbul stock exchange. *Journal of Advanced Studies in Finance (JASF)*, n. 2 (III), p. 151–161, 2012.
- SHANKER, R.; MISHRA, A. A quasi lindley distribution. *African Journal of Mathematics and Computer Science Research*, Academic Journals, v. 6, n. 4, p. pp–64, 2013.

- SHANKER, R.; SHARMA, S.; SHANKER, R. A two-parameter lindley distribution for modeling waiting and survival times data. *Applied Mathematics*, Scientific Research Publishing, v. 4, n. 02, p. 363, 2013.
- SHARMA, V. K. et al. The inverse lindley distribution: A stress-strength reliability model. 2015.
- SHAW, E. M. et al. *Hydrology in practice*. [S.l.]: CRC Press, 2010.
- SIROMONEY, G. The general dirichlet's series distribution. *Journal of the Indian Statistical Association*, v. 2, 1964.
- STEIN, W. E.; DATTERO, R. A new discrete weibull distribution. *IEEE Transactions on Reliability*, Institute of Electrical and Electronics Engineers, v. 33, n. 2, p. 196–197, 1984.
- STOCKBRIDGE, R. The discrete binomial model for option pricing. *Program in Applied Mathematics*, 2008.
- TUCKER, A. L.; POND, L. The probability distribution of foreign exchange price changes: tests of candidate processes. *The Review of Economics and Statistics*, JSTOR, p. 638–647, 1988.
- WALD, A.; WOLFOWITZ, J. An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics*, JSTOR, v. 14, n. 4, p. 378–388, 1943.
- WANG, Y. Estimation problems for the two-parameter negative binomial distribution. *Statistics & Probability Letters*, v. 26, n. 2, p. 113–114, 1996.
- YANG, Z. Maximum likelihood phylogenetic estimation from dna sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution*, Springer, v. 39, p. 306–314, 1994.
- ZEGHDOUDI, H.; NEDJAR, S. Gamma lindley distribution and its application. *J. Appl. Probab. Stat*, v. 11, 2015.
- ZIN, W. Z. W.; JEMAIN, A. A. Statistical distributions of extreme dry spell in peninsular malaysia. *Theoretical and applied climatology*, Springer, v. 102, n. 3-4, p. 253–264, 2010.
- ZUUR, A. F. et al. Zero-truncated and zero-inflated models for count data. In: *Mixed effects models and extensions in ecology with R*. [S.l.]: Springer, 2009. p. 261–293.