

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA  
PÓS-GRADUAÇÃO EM ESTATÍSTICA

JÉSSICA PRISCILA RIVAS DOS SANTOS

MODELOS KUMARASWAMY INFLACIONADOS

RECIFE  
2018

**JÉSSICA PRISCILA RIVAS DOS SANTOS**

**MODELOS KUMARASWAMY INFLACIONADOS**

Tese apresentada ao programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Doutora em Estatística.

Orientadora: Prof. PhD. Francisco Cribari Neto

RECIFE

2018

Catálogo na fonte  
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S237m Santos, Jéssica Priscila Rivas dos  
Modelos Kumaraswamy inflacionados / Jéssica Priscila Rivas dos Santos. –  
2018.  
115 f.: il., fig., tab.

Orientador: Francisco Cribari Neto.  
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN,  
Estatística, Recife, 2018.  
Inclui referências e apêndices.

1. Estatística. 2. Máxima verossimilhança. I. Cribari Neto, Francisco  
(orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2018-038

**JÉSSICA PRISCILA RIVAS DOS SANTOS**

**MODELOS KUMARASWAMY INFLACIONADOS**

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Estatística.

Aprovada em: 20 de fevereiro de 2018.

**BANCA EXAMINADORA**

---

Prof. Francisco Cribari Neto  
UFPE

---

Prof.<sup>a</sup> Audrey Helen Mariz de Aquino Cysneiros  
UFPE

---

Prof.<sup>a</sup> Patrícia Leone Espinheira Ospina  
UFPE

---

Prof.<sup>a</sup> Tarciana Liberal Pereira de Araújo  
UFPB

---

Prof.<sup>a</sup> Hildete Prisco Pinheiro  
Unicamp

*Aos meus pais, Geraldo Santos e Sandra Rivas,  
e a minha irmã, Caroline Rivas.*

## Agradecimentos

Acima de tudo e de todos, quero agradecer ao meu Deus por sempre permanecer ao meu lado, me guardar, me dar paz e força para persistir e finalizar mais uma etapa em minha vida.

Aos meus pais, Sandra Rivas e Geraldo Santos, por serem minha base, por acreditarem em mim e torcerem sempre. Essa conquista também é de vocês.

A minha irmã, Caroline Rivas, e meu cunhado, Lucas Oliveira, por me apoiarem e incentivarem. Vocês são meus exemplos de persistência e determinação.

Ao professor Francisco Cribari-Neto, pela competente orientação, por seu profissionalismo e pela paciência que teve comigo durante esses quatro anos de doutorado.

Ao meu amigo Sérgio Oliveira por sua perseverança em acreditar em mim, pela força que me deu em todos os momentos ao longo desses anos. Acho que você foi a pessoa que mais me ouviu chorar (risos). Sua crença foi muito motivadora para mim, meu “co”.

Aos meus familiares que torceram pelo meu êxito e oraram por mim.

A Andrea, dona Rose e Arthurzinho por serem minha família pernambucana, por me acolherem e por me darem força.

A Jack, Rose, Eliss, Flor, Lu e todas as demais amigas que acreditaram em mim, me suportando nos períodos cruciais para finalização dessa tese. Vocês também são minha família pernambucana e são de extrema importância para meu equilíbrio emocional e espiritual. Amo vocês.

Aos meus colegas do IFPE *Campus* Paulista que torceram tanto por mim e foram muito compreensíveis com essa etapa tão árdua em minha vida. Desejo todo sucesso em dobro para vocês e para esse *Campus* que mora em meu coração.

Aos meus colegas da pós-graduação pelos momentos sofridos e divertidos que vivemos juntos, em especial a Daniele Trindade que está por perto desde 2008, na graduação. É muito tempo hein?! Obrigada pela força, principalmente na reta final.

A Valéria Bittencourt, pelo profissionalismo e por sua dedicação em ajudar sempre. Você é luz, Val.

Aos professores do Departamento de Estatística que compartilharam dos seus conhecimentos.

Aos membros da banca examinadora, pelas sugestões e correções.

A CAPES, pelo apoio financeiro nos meses iniciais do doutorado.

*“Consagre ao Senhor tudo o que você faz, e os seus planos serão bem sucedidos”.*  
Provérbios 16:3

# Resumo

Para investigar o comportamento de uma variável dado o conhecimento de variáveis explicativas é comum utilizar o modelo de regressão clássico ou os modelos lineares generalizados. Nenhum desses modelos, contudo, é adequado para modelar variáveis no intervalo  $(0, 1)$ . Para modelar variáveis que assumem valores em  $(0, 1)$  é bastante utilizado o modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004). Já para variáveis que assumem valores em  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$  é possível utilizar os modelos de regressão beta inflacionados propostos por Ospina e Ferrari (2012). Esta tese tem como objetivo introduzir distribuições Kumaraswamy inflacionadas, além de propor modelos de regressão Kumaraswamy inflacionados que permitem modelar dados nos intervalos  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$ , bem como abordar as respectivas inferências e avaliar os desempenhos dos estimadores de máxima verossimilhança em cada cenário. Simulações de Monte Carlo foram realizadas para verificar os desempenhos dos estimadores de máxima verossimilhança e de testes de hipóteses. Aplicações a dados reais também são apresentadas.

**Palavras chave:** Máxima verossimilhança. Modelo inflacionado. Regressão Kumaraswamy. Simulação de Monte Carlo. Teste da razão de verossimilhanças.



# Abstract

To investigate the behavior of a variable given the knowledge of explanatory variables it is common to use the classical regression model or the generalized linear models. However none of these models are suitable for modeling variables observed on  $(0, 1)$ . The beta regression model introduced by Ferrari and Cribari (2004) is useful when the response is restricted to the interval  $(0, 1)$ . Ospina and Ferrari (2012) developed an extension of such a model that allow practitioners to model fractional data observed on  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$ . Our goal in this thesis is to introduce inflated Kumaraswamy distributions in addition to proposing inflated Kumaraswamy regression models that allow one to model data that assume values in  $[0, 1)$ ,  $(0, 1]$  and  $[0, 1]$ . We develop parameter estimation, interval estimation and hypothesis testing inference. We present Monte Carlo simulation results to evaluate the performances of the maximum likelihood estimators and hypothesis tests. Empirical applications are also displayed.

**Keywords:**Inflated model. Kumaraswamy regression. Likelihood ratio test. Maximum likelihood. Monte Carlo simulation.

# Lista de Figuras

2.1	Densidades Kumaraswamy para diferentes valores de $(\alpha, \beta)$ . . . . .	20
2.2	Densidades Kumaraswamy inflacionadas em $c = 0$ com $\lambda = 0.5$ e diferentes valores de $(\alpha, \beta)$ . . . . .	23
2.3	Densidades Kumaraswamy inflacionadas em $c = 1$ com $\lambda = 0.5$ e diferentes valores de $(\alpha, \beta)$ . . . . .	24
2.4	Densidades Kumaraswamy inflacionada em um, para $\alpha = 1.5$ , $\beta = 3.0$ e diferentes valores de $\lambda$ , utilizadas nas simulações. . . . .	29
2.5	Poderes dos testes, distribuição Kumaraswamy inflacionada em um. . . . .	37
2.6	(a) Histograma e densidades estimadas e (b) funções distribuição acumuladas para a proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados. . . . .	39
2.7	(a) Histograma e densidades estimadas e (b) funções distribuição acumuladas para a proporção da população em domicílios com banheiro e água encanada. . . . .	41
3.1	Densidades Kumaraswamy inflacionadas em zero e um com $\lambda = 0.2$ , $p = 0.5$ e diferentes valores de $(\alpha, \beta)$ . . . . .	44
3.2	Poderes dos testes, distribuição KIZU. . . . .	62
3.3	Histograma da proporção de acumulação de água em reservatórios de Pernambuco. . . . .	64
3.4	(a) Histograma e densidades estimadas e (b) funções distribuição acumuladas para a proporção de acumulação de água em reservatórios de Pernambuco. . . . .	65
4.1	Histograma (painel esquerdo) e boxplot (painel direito) da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados. . . . .	102
4.2	Resíduos (painel esquerdo) e envelope simulado dos resíduos (painel direito) do modelo de regressão Kumaraswamy inflacionado em zero ajustado. . . . .	104
4.3	Histograma (painel esquerdo) e boxplot (painel direito) da proporção de pessoas de 10 a 14 anos desocupadas no Brasil em 2010. . . . .	105
4.4	Resíduos (painel esquerdo) e envelope simulado dos resíduos (painel direito) do modelo de regressão Kumaraswamy inflacionado em zero e um. . . . .	107

# Lista de Tabelas

2.1	Médias, variâncias, vieses e EQMs dos estimadores dos parâmetros da distribuição Kumaraswamy inflacionada em um; $\alpha = 1.5, \beta = 3.0$ . . . . .	30
2.2	Amplitudes médias dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em um; $\alpha = 1.5$ e $\beta = 3.0$ . . . . .	31
2.3	Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em um; $\alpha = 1.5$ e $\beta = 3.0$ . . . . .	32
2.4	Taxas (%) de rejeição nulas dos testes, considerando $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em zero. . . . .	33
2.5	Taxas (%) de rejeição não nulas dos testes, considerando $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em zero. . . . .	34
2.6	Taxas (%) de rejeição nulas dos testes, considerando $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em um. . . . .	35
2.7	Taxas (%) de rejeição não nulas dos testes, considerando $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em um. . . . .	36
2.8	Medidas resumo da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados. . . . .	39
2.9	Medidas resumo da proporção da população em domicílios com banheiro e água encanada. . . . .	40
3.1	Médias, variâncias, vieses e EQMs dos estimadores dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de $\lambda$ ; $p = 0.5, \alpha = 1.5, \beta = 3.0$ . . . . .	51
3.2	Médias, variâncias, vieses e EQMs dos estimadores dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de $p$ ; $\lambda = 0.5, \alpha = 1.5, \beta = 3.0$ . . . . .	52
3.3	Amplitudes médias dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de $\lambda$ ; $p = 0.5, \alpha = 1.5$ e $\beta = 3.0$ . . . . .	53
3.4	Amplitudes médias dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de $p$ ; $\lambda = 0.3, \alpha = 1.5$ e $\beta = 3.0$ . . . . .	54
3.5	Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de $\lambda$ ; $p = 0.5, \alpha = 1.5$ e $\beta = 3.0$ . . . . .	55

3.6	Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de $p$ ; $\lambda = 0.5, \alpha = 1.5$ e $\beta = 3.0$ . . . . .	56
3.7	Taxas (%) de rejeição nulas dos testes, considerando $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em zero e um. . . . .	58
3.8	Taxas (%) de rejeição nulas dos testes, considerando $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$ , para a distribuição Kumaraswamy inflacionada em zero e um. . . . .	59
3.9	Taxas (%) de rejeição não nulas dos testes, considerando $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em zero e um. . . . .	60
3.10	Taxas (%) de rejeição não nulas dos testes, considerando $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$ , para a distribuição Kumaraswamy inflacionada em zero e um. . . . .	61
3.11	Medidas resumo da proporção de acumulação de água em reservatórios de Pernambuco. . . . .	64
4.1	Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 1; $\pi_0 = -0.5, \pi_1 = -1.5$ e $\tau = 2.0$ . . . . .	75
4.2	Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 1; $\pi_0 = -0.5, \pi_1 = -1.5$ e $\tau = 2.0$ . . . . .	76
4.3	Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 1; $\pi_0 = -0.5, \pi_1 = -1.5$ e $\tau = 2.0$ . . . . .	77
4.4	Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 2; $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0$ e $\gamma_1 = -2.0$ . . . . .	78
4.5	Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 2; $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0$ e $\gamma_1 = -2.0$ . . . . .	79
4.6	Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 2; $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0$ e $\gamma_1 = -2.0$ . . . . .	80
4.7	Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 3; $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$ e $\varsigma_1 = 1.5$ . . . . .	81
4.8	Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 3; $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$ e $\varsigma_1 = 1.5$ . . . . .	81
4.9	Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 3; $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$ e $\varsigma_1 = 1.5$ . . . . .	82

4.10	Taxas (%) de rejeição nulas dos testes, considerando $H_0 : \varsigma_1 = \varsigma_1^{(0)}$ vs. $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , para o modelo Kumaraswamy inflacionado em zero com dispersão variável.	83
4.11	Taxas (%) de rejeição não nulas dos testes, considerando $H_0 : \varsigma_1 = \varsigma_1^{(0)}$ vs. $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , para o modelo Kumaraswamy inflacionado em zero com dispersão variável.	84
4.12	Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 1; $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0$ e $\tau = 2.0$ .	91
4.13	Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 1; $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0$ e $\tau = 2.0$ .	92
4.14	Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 1; $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0$ e $\tau = 2.0$ .	93
4.15	Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 2; $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$ e $\varsigma_1 = 1.5$ .	94
4.16	Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 2; $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$ e $\varsigma_1 = 1.5$ .	95
4.17	Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 2; $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$ e $\varsigma_1 = 1.5$ .	96
4.18	Taxas (%) de rejeição nulas dos testes, considerando $H_0 : \varsigma_1 = \varsigma_1^{(0)}$ vs. $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , para o modelo Kumaraswamy inflacionada em zero e um com dispersão variável.	97
4.19	Taxas (%) de rejeição não nulas dos testes, considerando $H_0 : \varsigma_1 = \varsigma_1^{(0)}$ vs. $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , para o modelo Kumaraswamy inflacionada em zero e um com dispersão variável.	97
4.20	Medidas resumo da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados.	101
4.21	Estimativas, erros-padrão, estatística da razão de verossimilhanças e $p$ -valor - modelo KI em zero.	103
4.22	Medidas resumo da proporção de pessoas de 10 a 14 anos desocupadas no Brasil em 2010, mas que havia procurado trabalho ao longo do mês anterior à data dessa pesquisa.	104

4.23	Estimativas, erros-padrão, estatística da razão de verossimilhanças e $p$ -valor - modelo KIZU. . . . .	106
1	Descrição das variáveis regressoras do modelo Kumaraswamy inflacionado em zero, retiradas do Atlas do Desenvolvimento Humano no Brasil 2013. . . . .	114
2	Descrição das variáveis regressoras do modelo Kumaraswamy inflacionado em zero e um, retiradas do Atlas do Desenvolvimento Humano no Brasil 2013. . .	115

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>16</b>
1.1	Organização da tese . . . . .	16
1.2	Suporte computacional . . . . .	17
<b>2</b>	<b>DISTRIBUIÇÃO KUMARASWAMY INFLACIONADA EM ZERO OU UM</b>	<b>18</b>
2.1	Introdução . . . . .	18
2.2	Distribuição Kumaraswamy . . . . .	20
2.3	Distribuição Kumaraswamy inflacionada em zero ou em um . . . . .	21
2.4	Estimação por máxima verossimilhança . . . . .	23
2.5	Testes de hipóteses . . . . .	27
2.5.1	Teste da razão de verossimilhanças . . . . .	27
2.5.2	Teste Wald . . . . .	28
2.5.3	Teste escore de Rao . . . . .	28
2.6	Avaliação numérica . . . . .	28
2.6.1	Estimação pontual . . . . .	29
2.6.2	Estimação intervalar . . . . .	30
2.6.3	Testes de hipóteses . . . . .	32
2.7	Aplicação . . . . .	38
2.7.1	Aplicação da distribuição KIZ . . . . .	38
2.7.2	Aplicação da distribuição KIU . . . . .	40
2.8	Conclusões . . . . .	41
<b>3</b>	<b>DISTRIBUIÇÃO KUMARASWAMY INFLACIONADA EM ZERO E UM</b>	<b>42</b>
3.1	Introdução . . . . .	42
3.2	Distribuição Kumaraswamy inflacionada em zero e um . . . . .	43
3.3	Estimação por máxima verossimilhança . . . . .	45
3.4	Testes de Hipóteses . . . . .	49
3.5	Resultados numéricos . . . . .	49
3.5.1	Estimação pontual . . . . .	50
3.5.2	Estimação intervalar . . . . .	53
3.5.3	Testes de hipóteses . . . . .	57
3.6	Aplicação . . . . .	63
3.7	Conclusões . . . . .	65

<b>4</b>	<b>MODELOS DE REGRESSÃO</b>	
	<b>KUMARASWAMY INFLACIONADOS</b>	<b>67</b>
4.1	Introdução . . . . .	67
4.2	Modelo de regressão Kumaraswamy . . . . .	68
4.3	Modelo de regressão Kumaraswamy inflacionado em zero ou um . .	70
4.3.1	Dispersão fixa . . . . .	70
4.3.2	Dispersão variável . . . . .	72
4.3.3	Testes de hipóteses . . . . .	74
4.4	Resultados numéricos para o modelo $KI_c$ . . . . .	74
4.5	Modelo de regressão Kumaraswamy inflacionado em zero e um . . .	84
4.5.1	Dispersão fixa . . . . .	85
4.5.2	Dispersão variável . . . . .	88
4.6	Resultados numéricos para o modelo KIZU . . . . .	90
4.7	Seleção do modelo, resíduos e bondade do ajuste . . . . .	97
4.7.1	Resíduos quantis aleatorizados . . . . .	98
4.7.2	Critérios de seleção do modelo . . . . .	100
4.7.3	Coefficiente de determinação generalizado . . . . .	100
4.7.4	Teste RESET . . . . .	100
4.8	Aplicações dos modelos Kumaraswamy inflacionados . . . . .	101
4.8.1	Modelo de regressão Kumaraswamy inflacionado em zero . . . . .	101
4.8.2	Modelo de regressão Kumaraswamy inflacionado em zero e um . . . . .	104
4.9	Conclusões . . . . .	107
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>108</b>
	<b>REFERÊNCIAS</b>	<b>111</b>
	<b>APÊNDICE A - COVARIÁVEIS DO MODELO KUMARASWAMY</b>	
	<b>INFLACIONADO EM ZERO</b>	<b>114</b>
	<b>APÊNDICE B - COVARIÁVEIS DO MODELO KUMARASWAMY</b>	
	<b>INFLACIONADO EM ZERO E UM</b>	<b>115</b>



# 1 INTRODUÇÃO

A ferramenta mais usual para investigar o comportamento de uma variável dado o conhecimento de variáveis explicativas é o modelo de regressão clássico. Apesar de ser muito utilizado, esse modelo possui limitações, como por exemplo a suposição de normalidade dos erros do modelo ou a relação linear entre a variável resposta e as variáveis regressoras, o que nem sempre é observado na prática. McCullagh e Nelder (1989) propuseram uma classe de modelos mais flexíveis, quando comparados com o modelo de regressão linear clássico: a classe de modelos lineares generalizados (MLG). Com essa nova estrutura é possível modelar tanto a média quanto o parâmetro de dispersão em função de covariáveis. Todavia, nem o modelo de regressão linear clássico nem os MLGs são adequados para modelar variáveis no intervalo  $(0, 1)$ .

Na prática, muitos dados podem ser observados no intervalo  $(0, 1)$ , como por exemplo índices de desenvolvimento humano, indicador de qualidade de vida, etc. Nesse caso, as distribuições beta e Kumaraswamy são apropriadas para ajustar os dados. No contexto de modelagem de dados no intervalo  $(0, 1)$ , Ferrari e Cribari (2004) propuseram o modelo de regressão beta, que permite modelar a média e a dispersão de uma variável aleatória que segue distribuição beta. Contudo, se o conjunto de dados apresentar valores zeros e/ou uns essas distribuições e modelo não são mais adequados.

Quando o interesse reside em modelar variáveis que assumem valores nos intervalos  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$ , é possível utilizar os modelos de regressão beta inflacionados propostos por Ospina e Ferrari (2012), que são baseados em distribuições beta inflacionadas (Ospina e Ferrari, 2010).

Visando explorar algumas vantagens da distribuição Kumaraswamy, que foi desenvolvida por Kumaraswamy (1976), esta tese tem como objetivo introduzir distribuições Kumaraswamy inflacionadas, além de propor modelos de regressão Kumaraswamy inflacionados que permitem modelar dados nos intervalos  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$ , bem como abordar as respectivas inferências.

## 1.1 Organização da tese

A presente tese está dividida como segue: no Capítulo 2 está apresentada a distribuição Kumaraswamy inflacionada em zero ou um e descritas as inferências relacionadas. Também foram realizadas simulações de Monte Carlo para verificar os desempenhos dos estimadores de máxima verossimilhança e dos testes de hipóteses, e é apresentada uma aplicação a dados reais; no Capítulo 3 é introduzida a distribuição Kumaraswamy inflacionada em zero e um, a devida inferência, bem como resultados de simulações e aplicação a dados reais. O modelo de regressão Kumaraswamy inflacionado em zero ou um e o modelo Kumaraswamy inflacionado em zero e um são introduzidos no Capítulo 4, juntamente

com resultados numéricos oriundos de simulações de Monte Carlo. Por fim, o Capítulo 5 apresenta as considerações finais desta tese.

## 1.2 Suporte computacional

A linguagem de programação OX, versão 7.0, para o sistema operacional Windows foi utilizada para realizar as simulações aqui apresentadas. OX é uma linguagem de programação matricial orientada a objetos que, utilizando uma sintaxe muito parecida com as de C e de C++, oferece uma gama de recursos matemáticos e estatísticos (Frery e Cribari-Neto, 2005) e está disponível gratuitamente para uso acadêmico em <http://www.doornik.com>. Para mais informes sobre essa ferramenta, ver Doornik e Ooms (2007).

O ambiente de programação R, versão 3.1.0, foi utilizado para produzir os gráficos. R é uma linguagem e um ambiente para computação estatística e para preparação de gráficos de alta qualidade (Frery e Cribari-Neto, 2005). Maiores informações sobre o R podem ser encontradas em <http://www.r-project.org>.

A tipografia desta tese foi realizada usando o sistema L<sup>A</sup>T<sub>E</sub>X, que foi criado por Leslie Lamport em meados da década de 80. Informações sobre o L<sup>A</sup>T<sub>E</sub>X podem ser encontradas em <http://www.latex-project.org>.

## 2 DISTRIBUIÇÃO KUMARASWAMY INFLACIONADA EM ZERO OU UM

### 2.1 Introdução

A modelagem de variáveis distribuídas de forma contínua no intervalo  $(0, 1)$  vem conquistando grande espaço na literatura estatística. A distribuição beta é bastante utilizada para modelar taxas e proporções, bem como outras variáveis que possuem suporte em  $(0, 1)$ , pois é bem flexível para moldar esse tipo de dados, uma vez que sua densidade pode apresentar diferentes formas dependendo dos valores dos dois parâmetros que indexam a distribuição. Entretanto, Kumaraswamy (1976) relata que muitos trabalhos empíricos notam que a distribuição beta não ajusta satisfatoriamente dados no contexto de estudos hidrológicos, especialmente se o período da série hídrica for pequeno, por exemplo uma série semanal. Diante disso, uma nova distribuição foi proposta por Kumaraswamy (1976) para ajustar dados que são delimitados em ambos os limites (inferior e superior). Esta nova proposta ficou conhecida na literatura estatística como distribuição Kumaraswamy e a partir dela muitos trabalhos foram desenvolvidos.

Garg (2009) derivou a distribuição conjunta, a distribuição do produto e a distribuição do quociente de duas estatísticas de ordem generalizadas provenientes da distribuição Kumaraswamy. Cordeiro e Castro (2009) descreveram uma nova família de distribuições generalizadas com base na distribuição Kumaraswamy a fim de estender as distribuições normal, Weibull, gamma, Gumbel e gaussiana inversa. Os autores expressaram os momentos ordinários de qualquer distribuição Kumaraswamy generalizada como funções lineares dos momentos ponderados da distribuição base. Carrasco et al. (2010) propuseram uma nova distribuição contínua com cinco parâmetros que generaliza as distribuições Kumaraswamy e beta, bem como algumas outras distribuições bastante conhecidas. Lemonte (2011) desenvolveu estimadores quase não viesados para a distribuição Kumaraswamy, derivou estimadores de máxima verossimilhança modificados que são livres de viés de segunda ordem e como uma alternativa para a correção de viés analítica considera um mecanismo de correção de viés baseado em bootstrap paramétrico. Barreto-Souza e Lemonte (2013) introduziram a distribuição Kumaraswamy bivariada, cujas distribuições marginais são Kumaraswamy e apresentaram algumas propriedades desta distribuição. Segundo Mitnik e Baek (2013), a distribuição Kumaraswamy tem uma importante vantagem sobre a distribuição beta que consiste no fato da inversa da função de distribuição acumulada ter forma fechada, ou simplesmente não depender de funções especiais, o que possibilita gerar ocorrências por inversão, conforme Jones (2009). Bridam e Nekoukhou (2013) introduziram uma nova classe de distribuições obtida pela composição da distribuição Kumaraswamy e da família de distribuições de série de potências. Silva e Barreto-Souza (2014) propuseram

um método de seleção entre as distribuições beta e Kumaraswamy. Além destes trabalhos citados, muitos outros utilizam a distribuição Kumaraswamy, ficando assim explícita a utilidade desta distribuição.

Ademais, em aplicações práticas quando se deseja trabalhar com variáveis que representam taxas e proporções, que em muitas situações são medidas no intervalo  $(0, 1)$ , por exemplo, proporção de pessoas analfabetas, precipitação diária, etc., a distribuição Kumaraswamy, assim como a distribuição beta, é adequada como já foi exposto. Todavia, é comum que dados de taxas e proporções assumam os valores zeros ou uns. Entretanto, o suporte da distribuição Kumaraswamy não inclui os valores extremos do intervalo  $(0, 1)$ , o que pode ser um limitante em aplicações práticas. Para acomodar os extremos do intervalo é preciso desenvolver uma nova distribuição que permita modelar dados estatísticos com suporte em  $(0, 1)$ , mas que permita a existência de observações com valor zero ou um. Uma forma de se aproximar é definir uma nova distribuição que resulta da mistura entre uma distribuição degenerada em zero ou em um e uma distribuição contínua.

Este conceito já é trabalhado na literatura para dados discretos e contínuos. No caso discreto podem ser encontradas as distribuições Poisson inflacionada em zero, binomial negativa inflacionada em zero; no caso contínuo, e mais especificamente para dados no intervalo  $[0, 1)$  ou no intervalo  $(0, 1]$ , tem-se a distribuição beta inflacionada em zero ou em um, proposta por Ospina e Ferrari (2010). Nesse último cenário, há também a distribuição simplex inflacionada em zero ou um e a distribuição simplex inflacionada em zero e um, proposta por Lucena (2017) em sua tese de doutorado.

O objetivo do presente capítulo é propor uma nova distribuição para modelar dados nos suportes  $[0, 1)$  ou  $(0, 1]$ . A proposta consiste em utilizar a distribuição Kumaraswamy para modelar a parte contínua dos dados e uma distribuição degenerada para modelar a parte discreta. A distribuição Kumaraswamy foi escolhida porque apresenta vantagens sobre a distribuição beta, como segue: a função de distribuição acumulada beta envolve a função beta incompleta, o que dificulta a geração de ocorrências de variáveis aleatórias beta através do método da inversão, mas isso não acontece na distribuição Kumaraswamy porque sua função de distribuição acumulada é relativamente simples e possui forma algébrica fechada, consequentemente a função quantílica não depende de funções especiais. Sendo assim, para gerar ocorrências de variáveis aleatórias a partir da distribuição beta é preciso utilizar o método da rejeição, que é computacionalmente mais intensivo, a depender da implementação que seja realizada. Isto também dá à distribuição Kumaraswamy uma vantagem no que tange à modelagem dos quantis; a densidade Kumaraswamy também pode assumir diversas formas: unimodal, ‘U’, crescente, decrescente, constante, dependendo dos valores dos parâmetros, ainda apresenta formato simétrico e assimétrico; além da distribuição Kumaraswamy ajustar melhor dados hidrológicos, como já foi mencionado.

Diante do exposto, será apresentada uma breve descrição sobre a distribuição Kumaraswamy na Seção 2.2; a distribuição proposta é discutida na Seção 2.3; já na Seção 2.4 é apresentada a estimação dos parâmetros da nova distribuição bem como a forma da sua matriz de informação de Fisher; na Seção 2.5 são descritas três estatísticas de

teste de hipóteses que serão utilizadas nas simulações feitas posteriormente; os resultados numéricos das simulações de Monte Carlo são exibidos na Seção 2.6.

## 2.2 Distribuição Kumaraswamy

Suponha que  $Y$  segue distribuição Kumaraswamy, cuja notação é  $Y \sim \text{Kum}(\alpha, \beta)$ , com função distribuição acumulada (fda) dada por

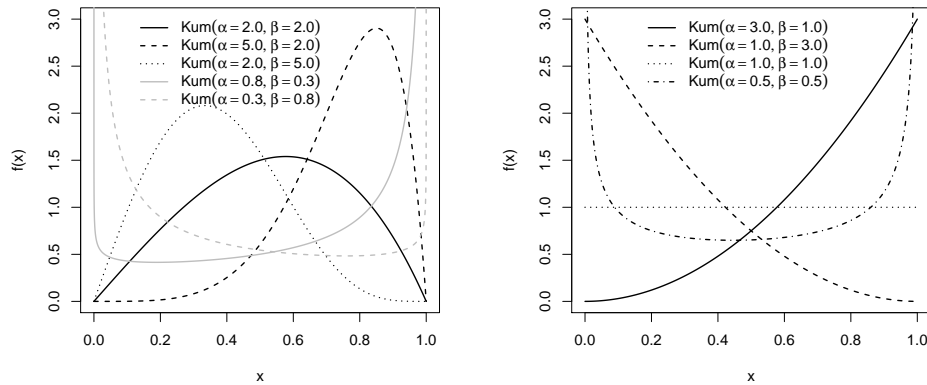
$$G(y; \alpha, \beta) = 1 - (1 - y^\alpha)^\beta, \quad (2.1)$$

com  $y \in (0, 1)$ , em que  $\alpha > 0$  e  $\beta > 0$  são os parâmetros de forma da distribuição. A função densidade de probabilidade (fdp) de  $Y$  é

$$g(y; \alpha, \beta) = \alpha\beta y^{\alpha-1}(1 - y^\alpha)^{\beta-1}, \quad y \in (0, 1). \quad (2.2)$$

Assim como a densidade beta, a densidade Kumaraswamy também pode assumir diversas formas, dependendo dos valores assumidos por seus parâmetros. A densidade da Kumaraswamy pode ser unimodal, ter forma de “banheira”, ser crescente, decrescente ou até mesmo constante. Segundo Lemonte (2011), é possível mostrar que a distribuição Kumaraswamy tem as mesmas propriedades de forma básicas da distribuição beta: se  $\alpha > 1$  e  $\beta > 1$  a densidade Kumaraswamy apresenta comportamento unimodal; se  $\alpha < 1$  e  $\beta < 1$  a densidade tem forma de ‘U’; se  $\alpha > 1$  e  $\beta \leq 1$  a densidade é crescente; se  $\alpha \leq 1$  e  $\beta > 1$  a densidade é decrescente; por fim, se  $\alpha = \beta = 1$  a densidade assume forma constante.

A Figura 2.1 apresenta diferentes densidades Kumaraswamy. Nota-se a diversidade de formatos que  $g(y; \alpha, \beta)$  assume dependendo dos valores de  $\alpha$  e  $\beta$ .



**Figura 2.1:** Densidades Kumaraswamy para diferentes valores de  $(\alpha, \beta)$ .

A média e a variância de  $Y$  são dadas, respectivamente, por

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

e

$$\text{Var}(Y) = \mu_2 - \mu_1^2, \quad (2.4)$$

em que

$$\mu_r = \mathbb{E}(Y^r) = \frac{\beta \Gamma\left(1 + \frac{r}{\alpha}\right) \Gamma(\beta)}{\Gamma\left(1 + \frac{r}{\alpha} + \beta\right)} = \beta B\left(1 + \frac{r}{\alpha}, \beta\right), \quad r = 1, 2, \dots, \quad (2.5)$$

é o  $r$ -ésimo momento da distribuição Kumaraswamy,  $\Gamma(\cdot)$  e  $B(\cdot, \cdot)$  sendo as funções gama e beta, respectivamente.

## 2.3 Distribuição Kumaraswamy inflacionada em zero ou em um

Dados de taxas e proporções podem conter zeros e/ou uns. Quando as observações incluem zeros ou uns, a distribuição Kumaraswamy não é adequada uma vez que seu suporte é  $(0, 1)$ . Um modelo apropriado deve adicionar alguma probabilidade de massa em zero ou um. Esse foi o enfoque adotado no contexto da distribuição beta por Ospina e Ferrari (2010), que propuseram uma distribuição para modelar proporções quando os dados contêm zeros ou uns. Ospina e Ferrari (2010) assumem que a variável aleatória segue distribuição mistura contínua-discreta com probabilidade de massa em zero ou um. Seguindo tal enfoque, é possível utilizar uma mistura de duas distribuições: a distribuição Kumaraswamy para modelar o componente contínuo, visto que esta pode assumir diversas formas que possibilitam um bom ajuste de dados em  $(0, 1)$ , e uma distribuição degenerada em um valor conhecido  $c$ , em que  $c = 0$  ou  $c = 1$ . Introduzimos assim a distribuição Kumaraswamy inflacionada em  $c$  ( $\text{KI}_c$ ), cuja fda é dada por

$$\text{KI}_c(y; \lambda, \alpha, \beta) = \lambda \mathbb{I}_{[c, 1]}(y) + (1 - \lambda)G(y; \alpha, \beta), \quad (2.6)$$

em que  $\mathbb{I}_A(y)$  é uma função indicadora, assumindo o valor 1 quando  $y \in A$  e 0 quando  $y \notin A$ ,  $0 < \lambda < 1$  é o parâmetro de mistura e  $G(y; \alpha, \beta)$  é a função distribuição acumulada Kumaraswamy. É possível observar neste caso que, com probabilidade  $1 - \lambda$ , os valores da variável aleatória  $Y$  seguirão distribuição Kumaraswamy com parâmetros  $(\alpha, \beta)$  e com probabilidade  $\lambda$  estes valores serão oriundos de uma distribuição degenerada em  $c$ .

Ocorrências de uma variável aleatória Kumaraswamy inflacionada em  $c$ , com função distribuição acumulada dada em (2.6), podem ser geradas com base no método da inversão, utilizando a inversa da distribuição Kumaraswamy. Um algoritmo para gerar valores de  $y$  é:

1. Gerar ocorrência  $u$  de uma variável aleatória uniforme padrão, isto é,  $U \sim \mathcal{U}(0, 1)$ ;
2. Obter valores para a subamostra contínua de  $y$ , a partir da função quantílica da distribuição Kumaraswamy;

3. Obter valores (zeros ou uns) para a subamostra discreta de  $y$ , com base no valor do parâmetro de mistura adotado.

Note que  $KI_c^{-1}(u)$  sempre será definida pois  $0 \leq u \leq 1$  e a imagem de  $KI_c$  é o intervalo  $(0, 1)$ . Além disso, observe que  $Y$  como definido no algoritmo de fato tem função distribuição acumulada  $KI_c$ , pois

$$\Pr(Y \leq y) = \Pr[KI_c^{-1}(U) \leq y] = \Pr\{KI_c[KI_c^{-1}(U)] \leq KI_c(y)\} = \Pr[U \leq KI_c(y)] = KI_c(y),$$

uma vez que  $U \sim \mathcal{U}(0, 1)$ .

A função densidade de probabilidade de  $Y$  referente à medida gerada pela mistura é dada por

$$ki_c(y; \lambda, \alpha, \beta) = \begin{cases} \lambda, & \text{se } y = c, \\ (1 - \lambda)g(y; \alpha, \beta), & \text{se } y \in (0, 1), \end{cases} \quad (2.7)$$

em que  $0 < \lambda < 1$ ,  $\alpha > 0$  e  $\beta > 0$  são os parâmetros da distribuição Kumaraswamy e  $g(y; \alpha, \beta)$  é a densidade expressa em (2.2). Note que  $\lambda$  representa a probabilidade de  $Y$  ser igual a  $c$  que pode assumir os valores zero ou um, ou seja,  $\lambda = \Pr(Y = 0)$  ou  $\lambda = \Pr(Y = 1)$ .

Se  $Y$  é uma variável aleatória com função densidade de probabilidade dada por (2.7) utiliza-se a notação  $Y \sim KI_c(\lambda, \alpha, \beta)$ .

As Figuras 2.2 e 2.3 apresentam diferentes densidades Kumaraswamy inflacionadas em  $c = 0$  e  $c = 1$ , respectivamente, variando os valores de  $\alpha$  e  $\beta$ , sendo o parâmetro de mistura fixado em  $\lambda = 0.5$ . Note que a função densidade de probabilidade da distribuição Kumaraswamy inflacionada em  $c$  expressa em (2.7), pode assumir diversas formas como por exemplo, unimodal, ‘U’, crescente, decrescente e até mesmo constante assim como também é observado nos diversos formatos da distribuição Kumaraswamy sem inflacionamento. Em todos os gráficos observa-se um comportamento assimétrico explicado pela existência da massa de probabilidade no ponto  $c$ , representada no gráfico pela linha vertical com um ponto em cima.

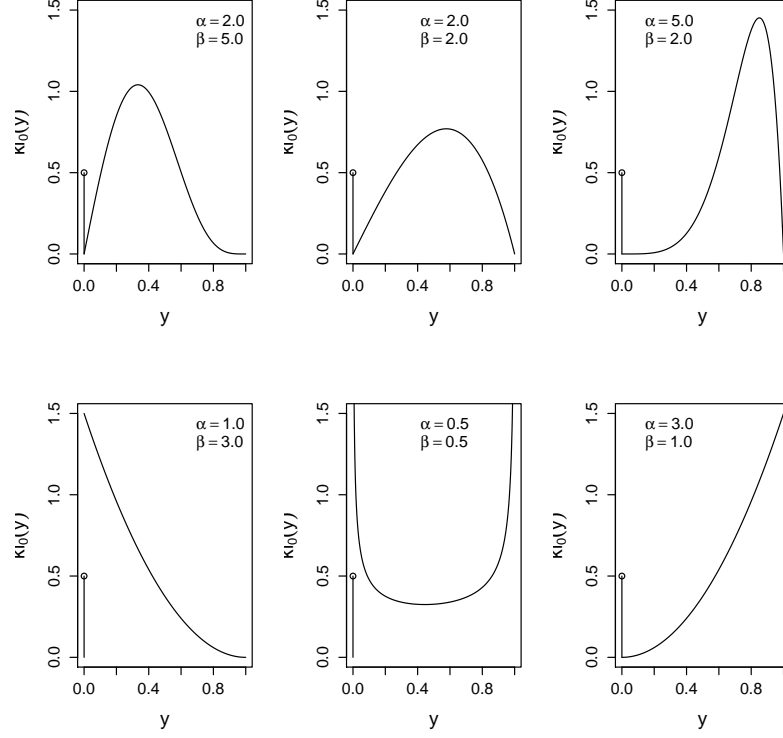
O  $r$ -ésimo momento de  $Y$  é dado por

$$\mathbb{E}(Y^r) = \lambda c + (1 - \lambda)\mu_r, \quad r = 1, 2, \dots, \quad (2.8)$$

em que  $\mu_r$  é o  $r$ -ésimo momento da distribuição Kumaraswamy expresso em (2.5).

A partir das expressões (2.4) e (2.8) é possível calcular a média e a variância da distribuição mistura que são dadas, respectivamente, por

$$\begin{aligned} \mathbb{E}(Y) &= \lambda c + (1 - \lambda)\mu_1 \\ &= \lambda c + \beta(1 - \lambda)B\left(1 + \frac{1}{\alpha}, \beta\right) \quad \text{e} \end{aligned}$$



**Figura 2.2:** Densidades Kumaraswamy inflacionadas em  $c = 0$  com  $\lambda = 0.5$  e diferentes valores de  $(\alpha, \beta)$ .

$$\begin{aligned}
 \text{Var}(Y) &= \lambda c + (1 - \lambda)\mu_2 - [\lambda c + (1 - \lambda)\mu_1]^2 \\
 &= \lambda c + \beta(1 - \lambda)B\left(1 + \frac{2}{\alpha}, \beta\right) - \left[\lambda c + \beta(1 - \lambda)B\left(1 + \frac{1}{\alpha}, \beta\right)\right]^2 \\
 &= \lambda c(1 - \lambda c) + (1 - \lambda)\beta \left\{ B\left(1 + \frac{2}{\alpha}, \beta\right) - B\left(1 + \frac{1}{\alpha}, \beta\right) [2\lambda c + \right. \\
 &\quad \left. + \beta(1 - \lambda)B\left(1 + \frac{1}{\alpha}, \beta\right)] \right\},
 \end{aligned}$$

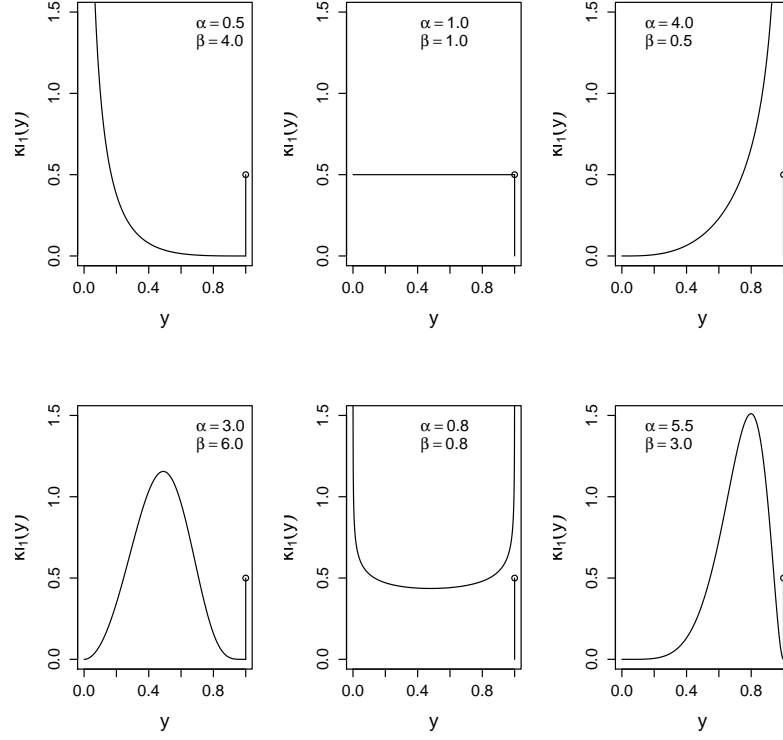
em que  $B(\cdot, \cdot)$  é a função beta.

## 2.4 Estimação por máxima verossimilhança

É possível reescrever a função densidade em (2.7) da seguinte maneira:

$$\text{ki}_c(y; \lambda, \alpha, \beta) = [\lambda^{\mathbb{I}_{\{c\}}(y)}(1 - \lambda)^{1 - \mathbb{I}_{\{c\}}(y)}] \times [g(y; \alpha, \beta)^{1 - \mathbb{I}_{\{c\}}(y)}]. \quad (2.9)$$





**Figura 2.3:** Densidades Kumaraswamy inflacionadas em  $c = 1$  com  $\lambda = 0.5$  e diferentes valores de  $(\alpha, \beta)$ .

Note que a densidade expressa em (2.9) fatora em dois termos: o primeiro dependendo apenas de  $\lambda$  e o segundo termo em função somente de  $\alpha$  e  $\beta$ . Isso facilita a construção da função de verossimilhança, que é apresentada a seguir.

A função de verossimilhança para  $\boldsymbol{\theta} = (\lambda, \alpha, \beta)^\top$  com base na amostra aleatória  $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$  proveniente da distribuição  $KI_c$  é

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^n ki_c(y_i; \lambda, \alpha, \beta) = L_1(\lambda; \mathbf{y}) \times L_2(\alpha, \beta; \mathbf{y}),$$

em que

$$L_1(\lambda; \mathbf{y}) = \prod_{i=1}^n \lambda^{\mathbb{I}_{\{c\}}(y_i)} (1 - \lambda)^{1 - \mathbb{I}_{\{c\}}(y_i)} = \lambda^{\sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i)} (1 - \lambda)^{n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i)}$$

e

$$L_2(\alpha, \beta; \mathbf{y}) = \prod_{i=1}^n g(y_i; \alpha, \beta)^{1 - \mathbb{I}_{\{c\}}(y_i)}.$$

Assim como a função densidade de probabilidade da distribuição inflacionada, a função de verossimilhança  $L(\boldsymbol{\theta})$  pode ser fatorada em dois termos. O primeiro termo depende

apenas de  $\lambda$  e o segundo depende apenas de  $(\alpha, \beta)$ .

O logaritmo da função de verossimilhança da distribuição Kumaraswamy inflacionada é dada por

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \ell_1(\lambda; \mathbf{y}) + \ell_2(\alpha, \beta; \mathbf{y}),$$

em que

$$\ell_1(\lambda; \mathbf{y}) = \log(\lambda) \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) + \ln(1 - \lambda) \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right]$$

e

$$\begin{aligned} \ell_2(\alpha, \beta; \mathbf{y}) = & \ln(\alpha\beta) \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right] + (\alpha - 1) \sum_{i=1}^n \ln(y_i) [1 - \mathbb{I}_{\{c\}}(y_i)] + \\ & (\beta - 1) \sum_{i=1}^n \ln(1 - y_i^\alpha) [1 - \mathbb{I}_{\{c\}}(y_i)]. \end{aligned}$$

Note que os termos  $\ln(y) [1 - \mathbb{I}_{\{c\}}(y)]$  e  $\ln(1 - y^\alpha) [1 - \mathbb{I}_{\{c\}}(y)]$  serão iguais a zero sempre que  $y = c$  e para isso foi usada a convenção  $\ln(0) \times 0 = 0$ . Essa convenção será utilizada em todos os cálculos de estimação por máxima verossimilhança presentes neste capítulo.

A função escore obtida diferenciando a função de log-verossimilhança com respeito aos parâmetros desconhecidos é  $U(\boldsymbol{\theta}) = [U_\lambda(\lambda), U_\alpha(\alpha, \beta), U_\beta(\alpha, \beta)]$ , em que

$$\begin{aligned} U_\lambda(\lambda) &= \frac{\partial \ell_1(\lambda; \mathbf{y})}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) - \frac{1}{1 - \lambda} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right], \\ U_\alpha(\alpha, \beta) &= \frac{\partial \ell_2(\alpha, \beta; \mathbf{y})}{\partial \alpha} = \frac{1}{\alpha} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right] + \sum_{i=1}^n \ln(y_i) [1 - \mathbb{I}_{\{c\}}(y_i)] + \\ & (\beta - 1) \sum_{i=1}^n \frac{y_i^\alpha}{y_i^\alpha - 1} \ln(y_i) [1 - \mathbb{I}_{\{c\}}(y_i)] \end{aligned}$$

e

$$U_\beta(\alpha, \beta) = \frac{\partial \ell_2(\alpha, \beta; \mathbf{y})}{\partial \beta} = \frac{1}{\beta} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right] + \sum_{i=1}^n \ln(1 - y_i^\alpha) [1 - \mathbb{I}_{\{c\}}(y_i)].$$

Resolvendo a equação  $U_\lambda(\lambda) = 0$  tem-se que o estimador de máxima verossimilhança (EMV) de  $\lambda$  é dado por  $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i)$ , que representa a proporção de valores iguais a  $c$  na amostra. Os estimadores de máxima verossimilhança de  $\alpha$  e  $\beta$  não têm forma fechada, mas podem ser obtidos maximizando numericamente a função de log-verossimilhança por meio de um algoritmo iterativo de otimização não-linear.

Para a obtenção da matriz de informação de Fisher foram calculadas as segundas derivadas da função de log-verossimilhança, que são dadas por

$$\begin{aligned}
\frac{\partial^2 \ell_1(\lambda; \mathbf{y})}{\partial \lambda \partial \lambda} &= -\frac{1}{\lambda^2} \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) - \frac{1}{(1-\lambda)^2} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right], \\
\frac{\partial^2 \ell_1(\lambda; \mathbf{y})}{\partial \lambda \partial \alpha} &= \frac{\partial^2 \ell_1(\lambda; \mathbf{y})}{\partial \lambda \partial \beta} = \frac{\partial^2 \ell_2(\alpha, \beta; \mathbf{y})}{\partial \alpha \partial \lambda} = \frac{\partial^2 \ell_2(\alpha, \beta; \mathbf{y})}{\partial \beta \partial \lambda} = 0, \\
\frac{\partial^2 \ell_2(\alpha, \beta; \mathbf{y})}{\partial \alpha \partial \alpha} &= -\frac{1}{\alpha^2} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right] - (\beta - 1) \sum_{i=1}^n \frac{y_i^\alpha}{(y_i^\alpha - 1)^2} \ln^2(y_i) [1 - \mathbb{I}_{\{c\}}(y_i)], \\
\frac{\partial^2 \ell_2(\alpha, \beta; \mathbf{y})}{\partial \alpha \partial \beta} &= \frac{\partial^2 \ell_2(\alpha, \beta; \mathbf{y})}{\partial \beta \partial \alpha} = \sum_{i=1}^n \frac{y_i^\alpha}{y_i^\alpha - 1} \ln(y_i) [1 - \mathbb{I}_{\{c\}}(y_i)] \quad \text{e} \\
\frac{\partial^2 \ell_2(\alpha, \beta; \mathbf{y})}{\partial \beta \partial \beta} &= -\frac{1}{\beta^2} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right].
\end{aligned}$$

Calculando os valores esperados das segundas derivadas acima multiplicados por  $-1$ , a matriz de informação de Fisher é

$$K(\boldsymbol{\theta}) = \begin{pmatrix} k_{\lambda\lambda} & 0 & 0 \\ 0 & k_{\alpha\alpha} & k_{\alpha\beta} \\ 0 & k_{\beta\alpha} & k_{\beta\beta} \end{pmatrix}, \quad (2.10)$$

em que

$$\begin{aligned}
k_{\lambda\lambda} &= \frac{n}{\lambda(1-\lambda)}, \\
k_{\alpha\alpha} &= \frac{n(1-\lambda)}{\alpha^2} + \frac{n\beta(1-\lambda)}{\alpha^2(\beta-2)} \{[\psi(\beta) - \psi(2)]^2 - [\psi'(\beta) - \psi'(2)]\}, \\
k_{\alpha\beta} &= k_{\beta\alpha} = -\frac{n(1-\lambda)}{\alpha(\beta-1)} \{[\psi(\beta+1) - \psi(2)]\}, \\
k_{\beta\beta} &= \frac{n(1-\lambda)}{\beta^2},
\end{aligned}$$

$\psi(\cdot)$  denotando a função digama, ou seja,  $\psi(z) = \partial \ln \Gamma(z) / \partial z$ , sendo  $\Gamma(\cdot)$  a função gama. Adicionalmente,  $\psi'(\cdot)$  é a derivada de primeira ordem da função  $\psi(\cdot)$ , conhecida como função trigama.

A distribuição assintótica dos estimadores de máxima verossimilhança é normal, ou seja,  $\hat{\boldsymbol{\theta}} \stackrel{a}{\sim} \mathcal{N}_3(\boldsymbol{\theta}, K(\boldsymbol{\theta})^{-1})$ , em que  $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\alpha}, \hat{\beta})^\top$  é o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ ,  $\stackrel{a}{\sim}$  denota assintoticamente distribuído,  $\mathcal{N}_3$  denota a distribuição normal trivariada e  $K(\boldsymbol{\theta})^{-1}$  é a inversa da matriz de informação de Fisher. A distribuição normal  $\mathcal{N}_3(0, K(\boldsymbol{\theta})^{-1})$  pode ser usada para construir intervalos de confiança assintóticos para  $\lambda, \alpha$  e  $\beta$  dados, respectivamente, por

$$\hat{\lambda} \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\lambda}), \quad \hat{\alpha} \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\alpha}) \quad \text{e} \quad \hat{\beta} \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\beta}),$$

em que  $\text{ep}(\cdot)$  é o erro-padrão de cada estimador dado, por exemplo, por  $\text{ep}(\hat{\lambda}) = \sqrt{\hat{k}_{\lambda\lambda}}$ , com  $\hat{k}_{\lambda\lambda}$  representando o elemento  $k_{\lambda\lambda}$  avaliado em  $\hat{\lambda}$ , ou seja, os erros-padrão são calculados pela raiz quadrada das variâncias estimadas dos estimadores de máxima verossimilhança que compõem a diagonal principal de  $K(\hat{\boldsymbol{\theta}})^{-1}$ , sendo  $K(\hat{\boldsymbol{\theta}})^{-1}$  a inversa da matriz de informação de Fisher com os parâmetros substituídos por estimativas;  $z_{(1-\frac{\delta}{2})}$  é o quantil  $(1 - \frac{\delta}{2})$  da distribuição normal padrão e  $1 - \delta$  é o nível de confiança adotado.

## 2.5 Testes de hipóteses

### 2.5.1 Teste da razão de verossimilhanças

Sejam  $\boldsymbol{\Theta}$  o espaço paramétrico,  $H_0 : \boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0$  a hipótese nula e  $H_1 : \boldsymbol{\vartheta} \in \boldsymbol{\Theta}_1$  a hipótese alternativa, em que  $\boldsymbol{\Theta}_0$  é o espaço paramétrico restrito a  $H_0$ , sendo  $\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$ ,  $\boldsymbol{\Theta}_1$  denotando o espaço paramétrico sob  $H_1$ , dos quais tem-se que  $\boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}$  e  $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1 = \emptyset$ .

O teste da razão de verossimilhanças ( $RV$ ) baseia-se na comparação de ajustes de dois modelos, o modelo restrito que é obtido impondo-se a restrição da hipótese nula e o modelo irrestrito que não embute restrições. A estatística de teste utiliza a razão entre as verossimilhanças irrestrita e restrita, isto é,  $L(\hat{\boldsymbol{\vartheta}}; y)/L(\boldsymbol{\vartheta}^{(0)}; y)$ , em que  $L(\hat{\boldsymbol{\vartheta}}; y)$  é a função de verossimilhança irrestrita avaliada em  $\hat{\boldsymbol{\vartheta}}$  e na amostra e  $L(\boldsymbol{\vartheta}^{(0)}; y)$  é a função de verossimilhança restrita avaliada em  $\boldsymbol{\vartheta}^{(0)} \in \boldsymbol{\Theta}_0$  e na amostra. Com base nessa estatística é possível decidir de qual modelo é mais verossímil que os dados tenham vindo.

Para este caso, a estatística da razão de verossimilhanças é  $RV = 2[\ell(\hat{\boldsymbol{\vartheta}}) - \ell(\boldsymbol{\vartheta}^{(0)})]$ , em que  $\ell(\cdot)$  é o logaritmo natural da função de verossimilhança,  $\hat{\boldsymbol{\vartheta}}$  é o estimador de máxima verossimilhança de  $\boldsymbol{\vartheta}$  e  $\boldsymbol{\vartheta}^{(0)}$ , neste cenário, é um escalar imposto. Sob  $H_0$  e sob certas condições de regularidade,  $RV \xrightarrow{d} \chi^2_1$ , em que  $\xrightarrow{d}$  denota convergência em distribuição. Assim, a hipótese nula será rejeitada se  $RV > \chi^2_{1,1-\delta}$ , em que  $\delta$  é o nível de significância do teste e  $\chi^2_{1,1-\delta}$  é o valor crítico assintótico obtido da distribuição  $\chi^2_1$ .

Agora seja  $\boldsymbol{\vartheta}$  um vetor de dimensão  $\nu$ . Podemos testar  $H_0 : \boldsymbol{\vartheta} = \boldsymbol{\vartheta}^{(0)}$  vs.  $H_1 : \boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}^{(0)}$ , em que  $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_\nu)^\top$  e  $\boldsymbol{\vartheta}^{(0)} = (\vartheta_1^{(0)}, \dots, \vartheta_\nu^{(0)})^\top$ ,  $\boldsymbol{\vartheta}^{(0)}$  sendo um vetor  $\nu \times 1$  dado. A estatística da razão de verossimilhanças é  $RV = 2[\ell(\hat{\boldsymbol{\vartheta}}) - \ell(\boldsymbol{\vartheta}^{(0)})]$ , em que  $\hat{\boldsymbol{\vartheta}}$  é o estimador de máxima verossimilhança de  $\boldsymbol{\vartheta}$ . Neste caso,  $RV \xrightarrow{d} \chi^2_\nu$  e  $H_0$  é rejeitada se  $RV > \chi^2_{\nu,1-\delta}$ .

Suponha, por outro lado, que o interesse reside em testar um subconjunto de parâmetros, ou seja,  $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^\top, \boldsymbol{\vartheta}_2^\top)^\top$ ,  $\boldsymbol{\vartheta}_1$  é um vetor  $r \times 1$  de parâmetros de interesse e  $\boldsymbol{\vartheta}_2$  é um vetor  $(\nu - r) \times 1$  de parâmetros de incômodo e  $H_0 : \boldsymbol{\vartheta}_1 = \boldsymbol{\vartheta}_1^{(0)}$  vs.  $H_1 : \boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_1^{(0)}$ . A estatística da razão de verossimilhanças é

$$RV = 2[\ell(\hat{\boldsymbol{\vartheta}}) - \ell(\tilde{\boldsymbol{\vartheta}})],$$

em que  $\hat{\boldsymbol{\vartheta}}$  é o estimador de máxima verossimilhança irrestrito de  $\boldsymbol{\vartheta}$  e  $\tilde{\boldsymbol{\vartheta}} = (\boldsymbol{\vartheta}_1^{(0)\top}, \tilde{\boldsymbol{\vartheta}}_2^\top)^\top$  é o estimador de máxima verossimilhança restrito de  $\boldsymbol{\vartheta}$ , obtido pela imposição de  $H_0$ . De maneira análoga aos cenários anteriores, sob  $H_0$  e sob certas condições de regularidade,

$RV \xrightarrow{d} \chi_r^2$  sendo  $r$  o número de restrições impostas na hipótese nula. Assim,  $H_0$  é rejeitada se  $RV > \chi_{r,1-\delta}^2$ .

### 2.5.2 Teste Wald

O teste Wald também pode ser utilizado para testar restrições sobre o vetor de parâmetros. Considerando agora apenas o caso em que o interesse é testar restrições para um subconjunto de parâmetros de  $\boldsymbol{\vartheta}$ , suponha  $H_0 : \boldsymbol{\vartheta}_1 = \boldsymbol{\vartheta}_1^{(0)}$  vs.  $H_1 : \boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_1^{(0)}$ , em que  $\boldsymbol{\vartheta}_1$  é o vetor  $r \times 1$  de parâmetros de interesse. A estatística de teste Wald é dada por

$$W = (\hat{\boldsymbol{\vartheta}}_1 - \boldsymbol{\vartheta}_1^{(0)})^\top [K^{rr}(\hat{\boldsymbol{\vartheta}})]^{-1} (\hat{\boldsymbol{\vartheta}}_1 - \boldsymbol{\vartheta}_1^{(0)}),$$

em que  $\hat{\boldsymbol{\vartheta}}_1$  é o vetor de estimadores de máxima verossimilhança irrestritos de  $\boldsymbol{\vartheta}_1$  e  $K^{rr}(\hat{\boldsymbol{\vartheta}})$  é o bloco  $r \times r$  da inversa da matriz de informação de Fisher referente a  $\boldsymbol{\vartheta}_1$  avaliado no estimador de máxima verossimilhança irrestrito.

De maneira similar ao teste  $RV$ ,  $W \xrightarrow{d} \chi_r^2$  sob  $H_0$  e certas condições de regularidade. Portanto,  $H_0$  é rejeitada se  $W > \chi_{r,1-\delta}^2$ .

### 2.5.3 Teste escore de Rao

Por fim, Rao (1948) propôs um teste que também é bastante utilizado: o teste escore. Para o caso em que  $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^\top, \boldsymbol{\vartheta}_2^\top)^\top$  e deseja-se testar  $H_0 : \boldsymbol{\vartheta}_1 = \boldsymbol{\vartheta}_1^{(0)}$  vs.  $H_1 : \boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_1^{(0)}$ , a estatística escore é

$$S = U_r(\tilde{\boldsymbol{\vartheta}})^\top K^{rr}(\tilde{\boldsymbol{\vartheta}}) U_r(\tilde{\boldsymbol{\vartheta}}),$$

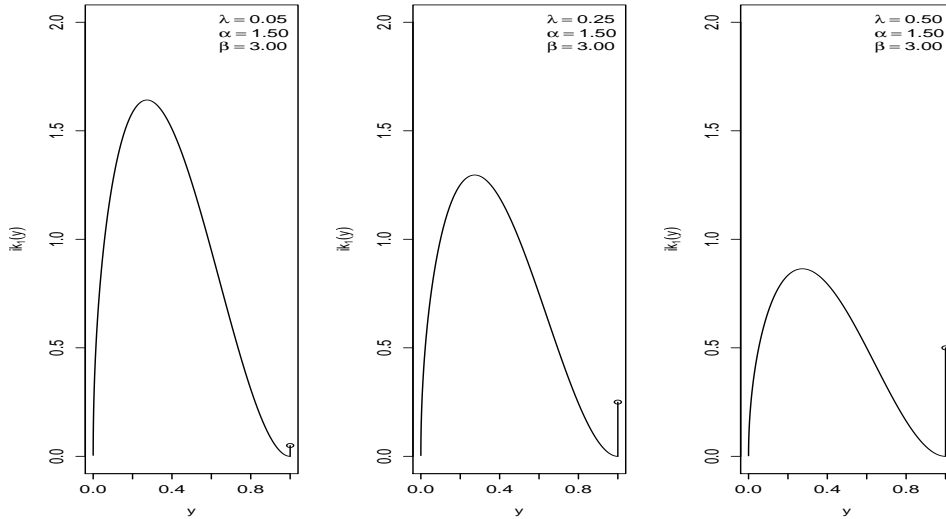
em que  $U_r(\tilde{\boldsymbol{\vartheta}})$  denota o vetor  $r \times 1$  que contém os  $r$  elementos da função escore referentes aos parâmetros que estão sob restrição na hipótese nula e  $K^{rr}(\tilde{\boldsymbol{\vartheta}})$  é a matriz  $r \times r$  que contém os elementos da inversa da informação de Fisher referentes aos  $r$  parâmetros testados, ambos avaliados em  $\tilde{\boldsymbol{\vartheta}} = \left( \boldsymbol{\vartheta}_1^{(0)\top}, \tilde{\boldsymbol{\vartheta}}_2^\top \right)^\top$ , sendo  $\boldsymbol{\vartheta}_1^{(0)}$  dado e  $\tilde{\boldsymbol{\vartheta}}_2$  o estimador restrito do vetor de parâmetros de incômodo. Aqui,  $S \xrightarrow{d} \chi_r^2$  sob  $H_0$  e certas condições de regularidade.  $H_0$  é rejeitada se  $S > \chi_{r,1-\delta}^2$ .

Para maiores detalhes sobre as condições de regularidade necessárias para a convergência em distribuição das estatísticas de teste consideradas nesse capítulo, ver o capítulo 4 de Serfling (2009).

## 2.6 Avaliação numérica

Foram realizadas simulações de Monte Carlo (MC) para avaliar os desempenhos dos estimadores de máxima verossimilhança dos parâmetros da distribuição Kumaraswamy inflacionada em um, ou seja  $c = 1$ , através de medidas como a média, variância, viés e erro quadrático médio (EQM).

Também foram avaliados os desempenhos de testes de  $H_0 : \lambda = \lambda_0$  *vs.*  $H_1 : \lambda \neq \lambda_0$ , com  $\lambda_0 = (0.05, 0.10, 0.20, 0.50)$ . Os testes utilizados nas simulações foram o teste da razão de verossimilhanças, o teste Wald e o teste escore de Rao, que foram apresentados na Seção 2.5. Os valores dos parâmetros utilizados nas simulações foram  $\alpha = 1.5$  e  $\beta = 3.0$ . Na Figura 2.4 é possível observar as formas assumidas pela função de densidade da distribuição Kumaraswamy inflacionada em um, utilizada nas simulações.



**Figura 2.4:** Densidades Kumaraswamy inflacionada em um, para  $\alpha = 1.5$ ,  $\beta = 3.0$  e diferentes valores de  $\lambda$ , utilizadas nas simulações.

Foram utilizadas 10000 réplicas de Monte Carlo e os tamanhos amostrais usados nas simulações foram  $n = 20, 30, 50, 100, 200$  e  $500$ .

Para a realização das simulações foi utilizada a linguagem matricial de programação Ox, (Doornik e Ooms, 2007). Maximizações da função de log-verossimilhança foram realizadas utilizando o método quasi-Newton BFGS (Nocedal, 1980) com primeiras derivadas analíticas. Os chutes iniciais utilizados nas simulações foram escolhidos arbitrariamente.

### 2.6.1 Estimação pontual

A Tabela 2.1 apresenta as médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros da distribuição Kumaraswamy inflacionada em um para os diferentes valores de  $n$ , considerando  $\lambda = (0.05, 0.25, 0.50)$ . É possível notar que à medida em que o tamanho amostral aumenta a média das estimativas de máxima verossimilhança aproximam-se dos valores impostos na simulação. Adicionalmente, a variância, o viés e o erro quadrático médio dos estimadores diminuem com o aumento de  $n$ , como esperado. Contudo, isso ocorre de forma mais lenta à medida em que o valor do parâmetro de mistura ( $\lambda$ ) aumenta, como é possível verificar a partir da análise da Tabela 2.1.

De maneira geral, os estimadores de máxima verossimilhança dos parâmetros da distribuição Kumaraswamy inflacionada em um apresentam comportamento assintótico esperado, evidenciando bom desempenho.

**Tabela 2.1:** Médias, variâncias, vieses e EQMs dos estimadores dos parâmetros da distribuição Kumaraswamy inflacionada em um;  $\alpha = 1.5, \beta = 3.0$ .

$\lambda$	Medida	Estimador	$n$				
			30	50	100	200	500
0.05	Média	$\hat{\lambda}$	0.0636	0.0543	0.0502	0.0501	0.0500
		$\hat{\alpha}$	1.5960	1.5562	1.5289	1.5142	1.5065
		$\hat{\beta}$	3.4949	3.2798	3.1347	3.0669	3.0298
	Variância	$\hat{\lambda}$	0.0011	0.0008	0.0005	0.0002	0.0001
		$\hat{\alpha}$	0.1061	0.0579	0.0266	0.0130	0.0051
		$\hat{\beta}$	1.8132	0.8120	0.3045	0.1358	0.0524
	Viés	$\hat{\lambda}$	0.0136	0.0043	0.0002	0.0001	0.0000
		$\hat{\alpha}$	0.0960	0.0562	0.0289	0.0142	0.0065
		$\hat{\beta}$	0.4949	0.2798	0.1347	0.0669	0.0298
	EQM	$\hat{\lambda}$	0.0013	0.0008	0.0005	0.0002	0.0001
		$\hat{\alpha}$	0.1153	0.0610	0.0274	0.0132	0.0052
		$\hat{\beta}$	2.0582	0.8903	0.3226	0.1403	0.0533
0.25	Média	$\hat{\lambda}$	0.2505	0.2497	0.2497	0.2500	0.2498
		$\hat{\alpha}$	1.6302	1.5739	1.5363	1.5170	1.5083
		$\hat{\beta}$	3.6793	3.3553	3.1665	3.0796	3.0351
	Variância	$\hat{\lambda}$	0.0061	0.0037	0.0018	0.0009	0.0004
		$\hat{\alpha}$	0.1423	0.0755	0.0343	0.0166	0.0064
		$\hat{\beta}$	2.9465	1.0743	0.4045	0.1790	0.0663
	Viés	$\hat{\lambda}$	0.0005	-0.0003	-0.0003	0.0000	-0.0002
		$\hat{\alpha}$	0.1302	0.0739	0.0363	0.0170	0.0083
		$\hat{\beta}$	0.6793	0.3553	0.1665	0.0796	0.0351
	EQM	$\hat{\lambda}$	0.0061	0.0037	0.0018	0.0009	0.0004
		$\hat{\alpha}$	0.1592	0.0810	0.0356	0.0169	0.0065
		$\hat{\beta}$	3.4079	1.2006	0.4323	0.1853	0.0675
0.50	Média	$\hat{\lambda}$	0.4999	0.4999	0.4998	0.4999	0.4998
		$\hat{\alpha}$	1.7126	1.6204	1.5572	1.5261	1.5119
		$\hat{\beta}$	4.3808	3.6131	3.2624	3.1233	3.0517
	Variância	$\hat{\lambda}$	0.0082	0.0049	0.0024	0.0012	0.0005
		$\hat{\alpha}$	0.2664	0.1282	0.0557	0.0258	0.0099
		$\hat{\beta}$	28.0805	2.6576	0.7277	0.2943	0.1030
	Viés	$\hat{\lambda}$	-0.0001	-0.0001	-0.0002	-0.0001	-0.0002
		$\hat{\alpha}$	0.2126	0.1204	0.0572	0.0261	0.0119
		$\hat{\beta}$	1.3808	0.6131	0.2624	0.1233	0.0517
	EQM	$\hat{\lambda}$	0.0082	0.0049	0.0024	0.0012	0.0005
		$\hat{\alpha}$	0.3116	0.1427	0.0590	0.0264	0.0101
		$\hat{\beta}$	29.9871	3.0335	0.7965	0.3095	0.1057

## 2.6.2 Estimação intervalar

Além da estimação pontual, também foram calculados os intervalos de confiança assintóticos, descritos na Seção 2.4, para os parâmetros da distribuição Kumaraswamy inflacionada em um, bem como as amplitudes destes intervalos e as respectivas probabilidades de cobertura e não cobertura.

Na Tabela 2.2 estão apresentadas as amplitudes dos intervalos de confiança resultantes das simulações com  $\lambda = (0.05, 0.25, 0.50)$  e considerando os níveis de confiança  $1 - \delta = (99\%, 95\%, 90\%)$ . Nota-se que à medida em que o tamanho amostral aumenta, os intervalos

de confiança ficam mais precisos, ou seja as amplitudes dos intervalos diminuem. Também é possível observar, analisando a Tabela 2.2, que quanto menor é o nível  $1 - \delta$ , os intervalos de confiança têm menor amplitude.

**Tabela 2.2:** Amplitudes médias dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em um;  $\alpha = 1.5$  e  $\beta = 3.0$ .

$\lambda$	$1 - \delta$	$\hat{\theta}$	$n$				
			30	50	100	500	1000
0.05	99%	$\hat{\lambda}$	0.2211	0.1589	0.1093	0.0500	0.0354
		$\hat{\alpha}$	1.5747	1.1916	0.8302	0.3673	0.2594
		$\hat{\beta}$	5.8552	4.1378	2.7454	1.1731	0.8248
	95%	$\hat{\lambda}$	0.1682	0.1209	0.0832	0.0380	0.0270
		$\hat{\alpha}$	1.1982	0.9067	0.6317	0.2795	0.1974
		$\hat{\beta}$	4.4553	3.1485	2.0890	0.8926	0.6276
	90%	$\hat{\lambda}$	0.1412	0.1015	0.0698	0.0319	0.0226
		$\hat{\alpha}$	1.0056	0.7609	0.5301	0.2346	0.1657
		$\hat{\beta}$	3.7390	2.6423	1.7531	0.7491	0.5267
0.25	99%	$\hat{\lambda}$	0.3982	0.3111	0.2215	0.0996	0.0705
		$\hat{\alpha}$	1.7956	1.3524	0.9385	0.4138	0.2920
		$\hat{\beta}$	7.0632	4.8015	3.1352	1.3234	0.9285
	95%	$\hat{\lambda}$	0.3030	0.2367	0.1686	0.0758	0.0536
		$\hat{\alpha}$	1.3663	1.0291	0.7141	0.3149	0.2222
		$\hat{\beta}$	5.3744	3.6535	2.3856	1.0070	0.7065
	90%	$\hat{\lambda}$	0.2543	0.1987	0.1415	0.0636	0.0450
		$\hat{\alpha}$	1.1466	0.8636	0.5993	0.2642	0.1865
		$\hat{\beta}$	4.5103	3.0661	2.0020	0.8451	0.5929
0.50	99%	$\hat{\lambda}$	0.4624	0.3607	0.2563	0.1151	0.0814
		$\hat{\alpha}$	2.3032	1.7026	1.1641	0.5079	0.3582
		$\hat{\beta}$	11.6019	6.5680	4.0093	1.6334	1.1428
	95%	$\hat{\lambda}$	0.3519	0.2744	0.1950	0.0876	0.0619
		$\hat{\alpha}$	1.7525	1.2955	0.8857	0.3864	0.2725
		$\hat{\beta}$	8.8279	4.9976	3.0507	1.2428	0.8696
	90%	$\hat{\lambda}$	0.2953	0.2303	0.1637	0.0735	0.0520
		$\hat{\alpha}$	1.4708	1.0872	0.7433	0.3243	0.2287
		$\hat{\beta}$	7.4086	4.1941	2.5602	1.0430	0.7298

Para avaliar se os intervalos de confiança contêm os verdadeiros valores de cada parâmetro impostos nas simulações, foram calculadas as probabilidades de cobertura e as probabilidades de não cobertura à esquerda e à direita, estando dispostos estes resultados na Tabela 2.3. Note que à medida em que  $n$  aumenta, as probabilidades de cobertura convergem para os níveis de confiança adotados em cada simulação. Adicionalmente, as probabilidades de não cobertura apresentam comportamento mais simétrico à medida em que o tamanho amostral aumenta, o que é esperado uma vez que os intervalos de confiança assintóticos são baseados na distribuição normal.



**Tabela 2.3:** Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em um;  $\alpha = 1.5$  e  $\beta = 3.0$ .

$\lambda$	$1 - \delta$	$\hat{\theta}$	$n$			
			30	50	100	500
0.05	99%	$\hat{\lambda}$	100.00	100.00	96.54	98.04
			(0.00; 0.00)	(0.00; 0.00)	(0.04; 3.42)	(0.17; 1.79)
		$\hat{\alpha}$	98.83	99.01	99.03	98.88
			(0.51; 0.66)	(0.47; 0.52)	(0.46; 0.51)	(0.59; 0.53)
		$\hat{\beta}$	98.67	98.86	99.00	98.98
			(0.00; 1.33)	(0.00; 1.14)	(0.04; 0.96)	(0.25; 0.77)
	95%	$\hat{\lambda}$	99.68	99.63	88.01	93.32
			(0.32; 0.00)	(0.37; 0.00)	(0.52; 11.47)	(1.25; 5.43)
		$\hat{\alpha}$	94.86	94.91	95.15	95.02
			(2.87; 2.27)	(2.84; 2.25)	(2.57; 2.28)	(2.77; 2.21)
		$\hat{\beta}$	96.36	96.07	95.85	95.23
			(0.00; 3.64)	(0.49; 3.44)	(1.10; 3.05)	(2.02; 2.75)
	90%	$\hat{\lambda}$	98.09	98.69	85.58	88.61
			(1.91; 0.00)	(1.31; 0.00)	(2.95; 11.47)	(2.83; 8.56)
		$\hat{\alpha}$	89.30	89.64	89.76	90.24
			(6.35; 4.35)	(6.05; 4.31)	(5.78; 4.46)	(5.30; 4.46)
		$\hat{\beta}$	93.23	91.75	91.28	90.11
			(0.94; 5.83)	(2.50; 5.75)	(3.39; 5.33)	(4.68; 5.21)
0.50	99%	$\hat{\lambda}$	98.43	98.60	98.87	98.98
			(0.84; 0.73)	(0.73; 0.67)	(0.60; 0.53)	(0.51; 0.51)
		$\hat{\alpha}$	98.86	98.85	98.72	98.88
			(0.49; 0.65)	(0.44; 0.71)	(0.54; 0.74)	(0.55; 0.57)
		$\hat{\beta}$	98.69	98.63	98.98	98.98
			(0.00; 1.31)	(0.00; 1.37)	(0.00; 1.02)	(0.20; 0.82)
	95%	$\hat{\lambda}$	96.08	93.72	94.62	94.54
			(2.05; 1.87)	(3.29; 2.99)	(2.71; 2.67)	(2.83; 2.63)
		$\hat{\alpha}$	94.60	94.52	94.53	94.80
			(3.41; 1.99)	(3.19; 2.29)	(3.05; 2.42)	(2.72; 2.48)
		$\hat{\beta}$	96.74	96.53	96.32	95.33
			(0.00; 3.26)	(0.00; 3.47)	(0.49; 3.19)	(1.67; 3.00)
	90%	$\hat{\lambda}$	90.65	88.42	91.66	90.28
			(4.76; 4.59)	(6.00; 5.58)	(4.11; 4.23)	(4.79; 4.93)
		$\hat{\alpha}$	89.35	89.51	89.46	89.86
			(6.94; 3.71)	(6.70; 3.79)	(6.16; 4.38)	(5.40; 4.74)
		$\hat{\beta}$	94.93	94.08	92.23	90.38
			(0.00; 5.07)	(0.56; 5.36)	(2.50; 5.27)	(4.43; 5.19)

### 2.6.3 Testes de hipóteses

Foram realizadas também simulações de tamanho dos testes razão de verossimilhanças ( $RV$ ), Wald ( $W$ ) e escore ( $S$ ), considerando as hipóteses  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$ , com  $\lambda_0 = (0.05, 0.10, 0.20, 0.50)$  para a distribuição Kumaraswamy inflacionada em zero. Os resultados dessas simulações se encontram na Tabela 2.4.

Nota-se que as taxas de rejeição dos testes, calculadas sob  $H_0$ , convergem para os respectivos níveis nominais, a saber: 1%, 5% e 10%, à medida em que  $n$  cresce. Destaca-se ainda que o teste da razão de verossimilhanças apresenta desempenho superior aos outros

**Tabela 2.4:** Taxas (%) de rejeição nulas dos testes, considerando  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em zero.

$\lambda_0$	Nível nominal	Estatística de teste	$n$					
			20	30	50	100	200	500
0.05	1%	$RV$	0.41	0.32	0.37	0.52	1.50	1.07
		$W$	0.01	0.00	0.00	3.46	2.48	1.96
		$S$	2.52	1.91	1.31	1.19	1.45	1.09
	5%	$RV$	2.52	1.91	1.31	6.37	5.03	5.21
		$W$	0.41	0.32	0.37	11.99	7.40	6.68
		$S$	11.60	7.90	4.08	2.95	3.44	4.62
	10%	$RV$	11.60	7.90	4.08	9.78	10.72	9.68
		$W$	2.52	1.91	1.31	14.42	15.03	11.39
		$S$	11.60	7.90	4.08	9.78	7.07	7.44
0.10	1%	$RV$	0.32	0.87	0.35	1.37	1.26	0.79
		$W$	0.07	0.03	2.94	2.48	1.85	1.37
		$S$	1.37	0.87	0.89	1.40	1.23	0.80
	5%	$RV$	1.37	2.86	5.25	4.44	5.87	5.21
		$W$	0.32	15.56	12.20	7.05	6.88	5.75
		$S$	4.99	2.86	2.39	6.35	4.39	4.37
	10%	$RV$	4.99	2.86	8.64	9.87	9.68	10.31
		$W$	1.37	17.55	13.70	13.90	11.88	11.22
		$S$	4.99	8.10	8.64	13.29	12.24	8.59
0.20	1%	$RV$	0.24	1.42	1.13	1.10	1.08	0.89
		$W$	6.12	4.69	2.04	1.24	1.37	0.95
		$S$	0.94	0.97	0.64	0.72	1.10	0.72
	5%	$RV$	3.29	6.81	5.02	4.48	5.29	5.15
		$W$	6.82	5.27	6.30	6.43	5.96	5.36
		$S$	3.29	3.54	5.02	5.98	4.31	5.15
	10%	$RV$	9.17	10.90	10.87	10.11	9.23	11.02
		$W$	9.17	14.55	13.60	11.41	9.81	11.04
		$S$	14.88	10.90	10.87	10.11	9.23	11.02
0.50	1%	$RV$	1.09	1.57	0.48	1.13	0.93	1.02
		$W$	4.15	1.57	1.40	1.13	1.42	1.02
		$S$	1.09	0.42	0.48	1.13	0.93	1.02
	5%	$RV$	4.15	3.92	6.28	5.38	5.57	5.46
		$W$	4.15	3.92	6.28	5.38	5.57	5.46
		$S$	4.15	3.92	6.28	5.38	5.57	5.46
	10%	$RV$	11.50	9.35	11.58	8.34	10.01	9.72
		$W$	11.50	9.35	11.58	8.34	10.01	9.72
		$S$	11.50	9.35	11.58	8.34	10.01	9.72

dois testes, uma vez que suas taxas de rejeição se aproximam mais rapidamente dos níveis nominais adotados quando o tamanho amostral aumenta. Por fim, é notável que quando aumenta o valor de  $\lambda_0$ , os tamanhos dos testes se aproximam de forma mais rápida dos níveis nominais e os desempenhos dos testes se tornam similares em quase todos os cenários.

Com base nas simulações de tamanho apresentadas na Tabela 2.4 foi possível calcular e armazenar os quantis 0.99, 0.95 e 0.90 referentes a cada estatística de teste, a fim de que fossem utilizados como valores críticos nas simulações de poder que estão dispostas na Tabela 2.5. Isso é feito para comparar poderes de testes que têm mesmo tamanho. Nessas simulações de poder, para os valores nulos de  $\lambda$ , a saber (0.05, 0.10, 0.20, 0.50),

**Tabela 2.5:** Taxas (%) de rejeição não nulas dos testes, considerando  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em zero.

$\lambda_0$	Nível nominal	Estatística de teste	$n$					
			20	30	50	100	200	500
0.05	1%	<i>RV</i>	22.63	36.69	59.09	88.52	98.29	100.00
		<i>W</i>	22.41	37.90	58.87	4.00	90.33	100.00
		<i>S</i>	17.88	29.11	48.30	83.60	99.05	100.00
	5%	<i>RV</i>	42.85	58.44	79.44	91.20	99.74	100.00
		<i>W</i>	43.70	59.16	78.13	83.60	98.29	100.00
		<i>S</i>	36.45	48.02	78.13	94.29	99.75	100.00
	10%	<i>RV</i>	56.87	70.81	86.78	97.11	99.87	100.00
		<i>W</i>	57.70	70.75	78.15	83.60	99.75	100.00
		<i>S</i>	36.45	68.61	78.13	97.11	99.87	100.00
0.10	1%	<i>RV</i>	9.03	25.72	42.16	63.26	93.64	100.00
		<i>W</i>	9.00	24.13	0.16	25.76	78.98	99.97
		<i>S</i>	9.00	24.13	42.16	63.26	93.64	100.00
	5%	<i>RV</i>	37.48	46.93	53.73	82.92	97.39	100.00
		<i>W</i>	37.45	24.38	28.66	59.67	95.67	100.00
		<i>S</i>	37.41	39.30	55.60	87.19	98.12	100.00
	10%	<i>RV</i>	47.83	57.49	68.92	87.32	98.94	100.00
		<i>W</i>	47.75	24.77	28.74	80.58	98.12	100.00
		<i>S</i>	37.41	57.37	68.91	87.19	98.88	100.00
0.20	1%	<i>RV</i>	11.68	8.13	19.41	43.04	75.95	99.52
		<i>W</i>	1.90	0.03	4.64	22.08	65.45	99.44
		<i>S</i>	11.28	15.72	21.52	44.75	75.95	99.52
	5%	<i>RV</i>	22.53	27.12	42.69	65.39	90.89	99.95
		<i>W</i>	2.32	10.97	21.30	53.54	88.17	99.86
		<i>S</i>	22.33	27.08	31.18	62.19	90.89	99.92
	10%	<i>RV</i>	25.05	38.93	48.21	77.64	95.21	99.98
		<i>W</i>	22.90	27.20	43.11	70.69	93.66	99.98
		<i>S</i>	22.33	27.08	43.03	70.66	95.21	99.98
0.50	1%	<i>RV</i>	4.77	7.29	12.09	28.69	60.08	97.52
		<i>W</i>	4.63	6.90	12.92	28.75	60.11	97.53
		<i>S</i>	1.59	4.38	9.24	24.17	58.91	96.87
	5%	<i>RV</i>	13.21	19.99	28.75	53.02	80.26	99.42
		<i>W</i>	13.15	19.76	30.77	52.72	80.84	99.45
		<i>S</i>	13.04	17.94	23.83	46.03	78.35	99.38
	10%	<i>RV</i>	25.43	29.84	41.40	64.51	88.96	99.82
		<i>W</i>	25.54	30.19	42.53	65.35	88.91	99.81
		<i>S</i>	13.04	29.25	33.70	62.04	86.07	99.81

foram considerados respectivamente, os seguintes valores de  $\lambda$  na geração da amostra: (0.15, 0.20, 0.30, 0.40).

Examinando os resultados das taxas de rejeição não nulas (Tabela 2.5) é notável que estas convergem para a probabilidade máxima à medida que  $n$  aumenta, como desejável. O teste escore apresenta desempenho inferior aos testes *RV* e Wald quando o tamanho amostral é pequeno, a saber  $n < 30$ . Contudo, esta diferença de desempenho é menor quando o valor do parâmetro de mistura imposto na simulação aumenta, isto é, quando há mais observações iguais a zero na amostra os testes apresentam desempenhos similares, contudo suas taxas de rejeição demoram mais para atingir 100%.

Na maioria dos cenários o teste da razão de verossimilhanças apresenta desempenho

superior aos testes Wald e escore, todavia quando  $\lambda_0 = 0.50$ , o teste Wald apresenta desempenho melhor do que os outros dois testes.

**Tabela 2.6:** Taxas (%) de rejeição nulas dos testes, considerando  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em um.

$\lambda_0$	Nível nominal	Estatística de teste	$n$					
			20	30	50	100	200	500
0.05	1%	$RV$	0.41	0.32	0.37	0.52	1.50	1.07
		$W$	0.01	0.00	0.00	3.46	2.48	1.96
		$S$	2.52	1.91	1.31	1.19	1.45	1.09
	5%	$RV$	2.52	1.91	1.31	6.37	5.03	5.21
		$W$	0.41	0.32	0.37	11.99	7.40	6.68
		$S$	11.60	7.90	4.08	2.95	3.44	4.62
	10%	$RV$	11.60	7.90	4.08	9.78	10.72	9.68
		$W$	2.52	1.91	1.31	14.42	15.03	11.39
		$S$	11.60	7.90	4.08	9.78	7.07	7.44
0.10	1%	$RV$	0.32	0.87	0.35	1.37	1.26	0.79
		$W$	0.07	0.03	2.94	2.48	1.85	1.37
		$S$	1.37	0.87	0.89	1.40	1.23	0.80
	5%	$RV$	1.37	2.86	5.25	4.44	5.87	5.21
		$W$	0.32	15.56	12.20	7.05	6.88	5.75
		$S$	4.99	2.86	2.39	6.35	4.39	4.37
	10%	$RV$	4.99	2.86	8.64	9.87	9.68	10.31
		$W$	1.37	17.55	13.70	13.90	11.88	11.22
		$S$	4.99	8.10	8.64	13.29	12.24	8.59
0.20	1%	$RV$	0.24	1.42	1.13	1.10	1.08	0.89
		$W$	6.12	4.69	2.04	1.24	1.37	0.95
		$S$	0.94	0.97	0.64	0.72	1.10	0.72
	5%	$RV$	3.29	6.81	5.02	4.48	5.29	5.15
		$W$	6.82	5.27	6.30	6.43	5.96	5.36
		$S$	3.29	3.54	5.02	5.98	4.31	5.15
	10%	$RV$	9.17	10.90	10.87	10.11	9.23	11.02
		$W$	9.17	14.55	13.60	11.41	9.81	11.04
		$S$	14.88	10.90	10.87	10.11	9.23	11.02
0.50	1%	$RV$	1.09	1.57	0.48	1.13	0.93	1.02
		$W$	4.15	1.57	1.40	1.13	1.42	1.02
		$S$	1.09	0.42	0.48	1.13	0.93	1.02
	5%	$RV$	4.15	3.92	6.28	5.38	5.57	5.46
		$W$	4.15	3.92	6.28	5.38	5.57	5.46
		$S$	4.15	3.92	6.28	5.38	5.57	5.46
	10%	$RV$	11.50	9.35	11.58	8.34	10.01	9.72
		$W$	11.50	9.35	11.58	8.34	10.01	9.72
		$S$	11.50	9.35	11.58	8.34	10.01	9.72

Também foram feitas simulações para verificar os desempenhos dos testes razão de verossimilhanças, Wald e escore considerando a distribuição Kumaraswamy inflacionada em um, ou seja, agora  $c = 1$  e os resultados para estas simulações estão apresentados nas Tabelas 2.6 e 2.7. As taxas de rejeição calculadas sob  $H_0$  com  $c = 1$  estão apresentadas na Tabela 2.6. Os resultados são similares aos obtidos com  $c = 0$ ; o teste da  $RV$  novamente se destaca, apresentando taxas de rejeição nulas mais próximas aos níveis nominais na maioria dos cenários. Quando  $\lambda_0$  aumenta, os desempenhos dos testes se tornam semelhantes.

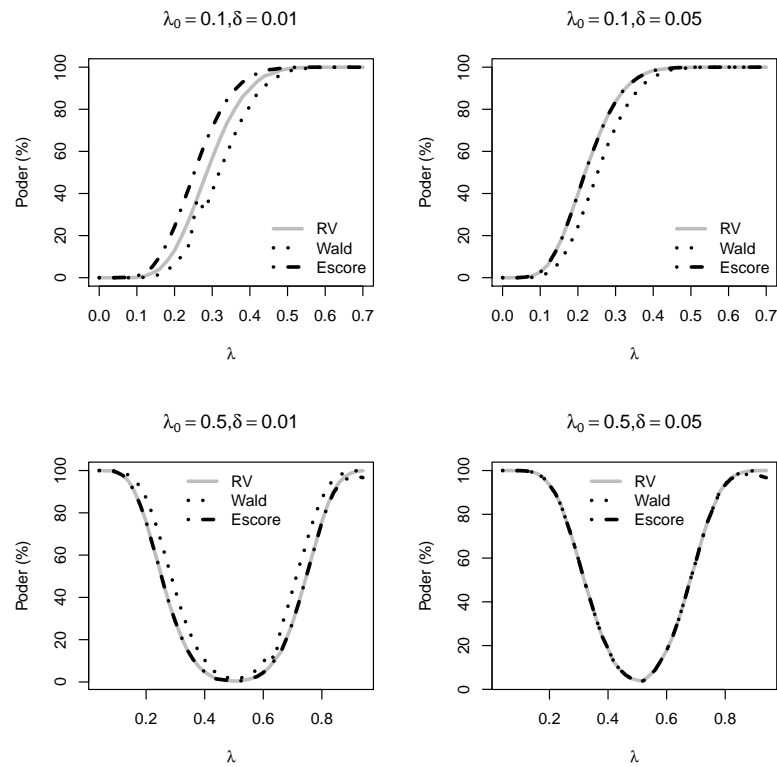
Da mesma maneira que anteriormente, os valores críticos utilizados nas simulações

**Tabela 2.7:** Taxas (%) de rejeição não nulas dos testes, considerando  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em um.

$\lambda_0$	Nível nominal	Estatística de teste	$n$					
			20	30	50	100	200	500
0.05	1%	$RV$	22.63	36.69	59.09	88.52	98.29	100.00
		$W$	22.41	37.90	58.87	4.00	90.33	100.00
		$S$	17.88	29.11	48.30	83.60	99.05	100.00
	5%	$RV$	42.85	58.44	79.44	91.20	99.74	100.00
		$W$	43.70	59.16	78.13	83.60	98.29	100.00
		$S$	36.45	48.02	78.13	94.29	99.75	100.00
	10%	$RV$	56.87	70.81	86.78	97.11	99.87	100.00
		$W$	57.70	70.75	78.15	83.60	99.75	100.00
		$S$	36.45	68.610	78.13	97.11	99.87	100.00
0.10	1%	$RV$	9.03	25.72	42.16	63.26	93.64	100.00
		$W$	9.00	24.13	0.16	25.76	78.98	99.97
		$S$	9.00	24.13	42.16	63.26	93.64	100.00
	5%	$RV$	37.48	46.93	53.73	82.92	97.39	100.00
		$W$	37.45	24.38	28.66	59.67	95.67	100.00
		$S$	37.41	39.30	55.60	87.19	98.12	100.00
	10%	$RV$	47.83	57.49	68.92	87.32	98.94	100.00
		$W$	47.75	24.77	28.74	80.58	98.12	100.00
		$S$	37.41	57.37	68.91	87.19	98.88	100.00
0.20	1%	$RV$	11.68	8.13	19.41	43.04	75.95	99.52
		$W$	1.90	0.03	4.64	22.08	65.45	99.44
		$S$	11.28	15.72	21.52	44.75	75.95	99.52
	5%	$RV$	22.53	27.12	42.69	65.39	90.89	99.95
		$W$	2.32	10.97	21.30	53.54	88.17	99.86
		$S$	22.33	27.08	31.18	62.19	90.89	99.92
	10%	$RV$	25.05	38.93	48.21	77.64	95.21	99.98
		$W$	22.90	27.20	43.11	70.69	93.66	99.98
		$S$	22.33	27.08	43.03	70.66	95.21	99.98
0.50	1%	$RV$	4.77	7.29	12.09	28.69	60.08	97.52
		$W$	4.63	6.90	12.92	28.75	60.11	97.53
		$S$	1.59	4.38	9.24	24.17	58.91	96.87
	5%	$RV$	13.21	19.99	28.75	53.02	80.26	99.42
		$W$	13.15	19.76	30.77	52.72	80.84	99.45
		$S$	13.04	17.94	23.83	46.03	78.35	99.38
	10%	$RV$	25.43	29.84	41.40	64.51	88.96	99.82
		$W$	25.54	30.19	42.53	65.35	88.91	99.81
		$S$	13.04	29.25	33.70	62.04	86.07	99.81

de poder para  $c = 1$  foram obtidos a partir das simulações de tamanho (Tabela 2.6). Para essas simulações de poder, os valores de  $\lambda$  utilizados na geração da amostra foram (0.15, 0.20, 0.30, 0.40), respectivamente. Os desempenhos dos testes continuam similares ao que já foi apresentado, o teste da razão de verossimilhanças apresentando melhores resultados na maioria dos cenários, quando comparado aos testes Wald e escore, como é notado na Tabela 2.7. Adicionalmente, à medida em que o valor hipotetizado aumenta, os poderes dos testes demoram mais a atingir 100%. Mesmo assim, contudo, os três testes apresentam altos poderes à medida que o tamanho amostral aumenta.

Para as simulações de poder com  $c = 1$  foram feitos gráficos que apresentam as taxas de rejeição calculadas sob a hipótese alternativa, ou seja, os poderes dos testes (Figura



**Figura 2.5:** Poderes dos testes, distribuição Kumaraswamy inflacionada em um.

2.5). Para a construção das curvas de poder foram utilizadas 5000 réplicas de Monte Carlo, tamanho amostral igual a 30,  $\alpha = 1.5$  e  $\beta = 3.0$ .

Analisando a Figura 2.5 é possível avaliar os poderes dos testes razão de verossimilhanças, escore e Wald e notar que estes de fato apresentam bom desempenho, uma vez que seus poderes convergem para 100% à medida em que os valores de  $\lambda$  se afastam do valor hipotetizado. Vale salientar que no caso em que  $\lambda_0 = 0.10$ , os gráficos não apresentam uma forma comum para poder de teste bilateral, visto que  $\lambda \in (0, 1)$  e estes testes não são tão sensíveis a erros pequenos, ou seja, para que as taxas de poder alcancem valores mais altos fazem-se necessários valores de  $\lambda$  mais distantes do valor hipotetizado, o que não acontece neste caso pois 0.10 já é bem próximo do limite inferior que  $\lambda$  pode assumir. Quando  $\delta = 1\%$  o teste escore funciona melhor para valores menores de  $\lambda_0$ , nesse caso  $\lambda_0 = 0.10$  e o teste Wald funciona melhor para valores maiores,  $\lambda_0 = 0.50$ . Note ainda que para  $\lambda_0 = 0.50$  são necessários valores mais distantes de  $\lambda_0$  para obter poderes mais altos, por exemplo, no caso em que  $\lambda_0 = 0.10$  e  $\delta = 0.01$ , considerando uma diferença do valor nulo de 0.20, ou seja  $\lambda = 0.30$ , se obtém um valor para o poder em torno de 60%, já no caso em que  $\lambda_0 = 0.50$ , considerando também uma diferença de 0.20 com relação ao valor nulo, isto é  $\lambda = 0.70$ , o poder é aproximadamente 40%. Adicionalmente, os testes *RV* e *escore* apresentam desempenhos bastante similares, a não ser quando  $\lambda_0 = 0.10$  e o nível

de significância dos testes é de 1%, em que o teste escore apresenta poderes mais altos. É válido ressaltar que à medida que os valores de  $\lambda$  aproximam-se dos limites do suporte de  $\lambda$  há falhas de convergência, isso ocorre com mais frequência quando os valores de  $\lambda$  se aproximam de um.

## 2.7 Aplicação

Para ilustrar a utilidade das distribuições Kumaraswamy inflacionada em zero (KIZ) e Kumaraswamy inflacionada em um (KIU) foram realizadas duas aplicações com dados reais restritos aos intervalos  $[0, 1)$  e  $(0, 1]$ , respectivamente. A primeira variável utilizada é a proporção de pessoas em domicílios nos municípios brasileiros com abastecimento de água e esgoto sanitário inadequados em 2010 enquanto a segunda variável é a proporção da população brasileira em domicílios com banheiro e água encanada no ano de 2010, ambas retiradas do Atlas do Desenvolvimento Humano no Brasil 2013 (Brasil, 2014).

Foi empregado o método de máxima verossimilhança para estimar os parâmetros das distribuições através da função `optim` do software R, utilizando o método L-BFGS-B com primeiras derivadas analíticas. As distribuições Kumaraswamy inflacionadas foram implementadas no R com o auxílio de funções referentes à distribuição Kumaraswamy, disponíveis no pacote `VGAM` (Yee, 2008).

Visando comparar o ajuste da distribuição Kumaraswamy inflacionada em zero ou em um com outra distribuição que serve para modelar dados nos intervalos  $[0, 1)$  ou  $(0, 1]$ , também foram ajustadas as distribuições beta inflacionadas em zero (BIZ) e beta inflacionada em um (BIU) propostas por Ospina e Ferrari (2010). Uma variável aleatória que tem distribuição beta inflacionada em  $c$ , sendo  $c = 0$  ou  $c = 1$ , com parâmetros  $\lambda, \mu$  e  $\phi$ , possui densidade dada por

$$bi_c(y; \lambda, \mu, \phi) = \begin{cases} \lambda, & \text{se } y = c, \\ (1 - \lambda)f(y; \mu, \phi), & \text{se } y \in (0, 1), \end{cases}$$

em que  $f(y; \mu, \phi)$  é a função densidade de probabilidade da distribuição beta com parâmetros  $\mu$  (média) e  $\phi$  (parâmetro de precisão). Para estimar os parâmetros da distribuição beta inflacionada em  $c$ , foram utilizadas as funções `BEZI` e `BE0I` implementadas no pacote `gamlss.dist` (Stasinopoulos et. al, 2012).

### 2.7.1 Aplicação da distribuição KIZ

O primeiro conjunto de dados possui 5565 observações sobre a proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados em municípios brasileiros no ano de 2010. Essa proporção é dada pela razão entre o total de pessoas que vivem em domicílios cujo abastecimento de água não provém de rede geral e cujo esgoto sanitário não é realizado por rede coletora de esgoto ou fossa séptica e a população total residente em domicílios particulares permanentes multiplicado. São considerados apenas os domicílios particulares permanentes e para essa aplicação os dados serão divididos por 100.

Essas informações foram retiradas do Atlas do Desenvolvimento Humano no Brasil 2013 que se encontra disponível em <http://www.atlasbrasil.org.br/2013/pt/download/>. Os dados contêm 434 municípios com proporção igual a zero e por isso será ajustada a distribuição Kumaraswamy inflacionada em zero.

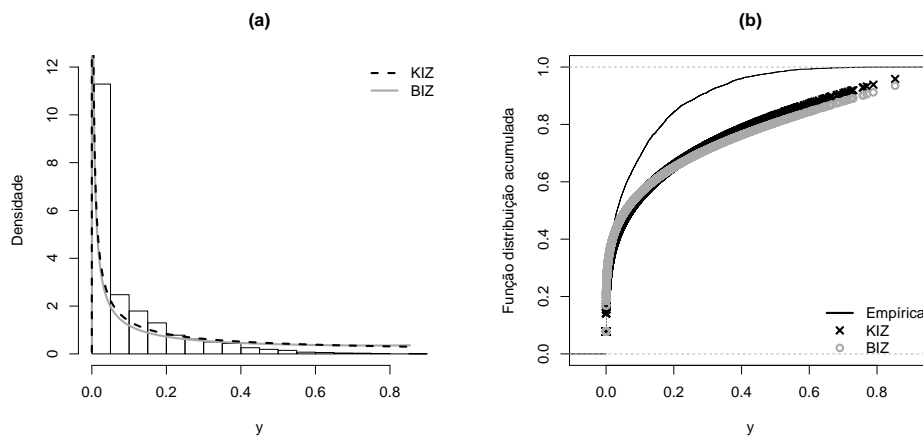
Na Tabela 2.8 são apresentadas algumas medidas resumo da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados. Nota-se que o valor mínimo da variável de interesse é zero e que 75% das observações estão abaixo de 0.1302.

**Tabela 2.8:** Medidas resumo da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
0.0000	0.0053	0.0326	0.0920	0.1302	0.8536

As estimativas de máxima verosimilhança dos parâmetros da distribuição KIZ, com os respectivos erros-padrão entre parênteses, são  $\hat{\lambda} = 0.0780$  (0.0036),  $\hat{\alpha} = 0.2906$  (0.0051) e  $\hat{\beta} = 0.9990$  (0.0175). Para a distribuição BIZ as estimativas são  $\hat{\lambda} = 0.0780$  (0.0036),  $\hat{\mu} = 0.2413$  (0.0039) e  $\hat{\phi} = 0.9990$  (0.0193).

Na Figura 2.6 estão apresentados (a) o histograma da distribuição dos dados juntamente com as densidades KIZ e BIZ estimadas, além (b) das funções de distribuição acumuladas empírica e das respectivas distribuições inflacionadas ajustadas. O painel (a) sugere que a densidade possui forma de “J” invertido.



**Figura 2.6:** (a) Histograma e densidades estimadas e (b) funções distribuição acumuladas para a proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados.

Foram calculadas as estatísticas dos testes Kolmogorov-Smirnov (KS), Anderson-Darling (AD) e Cramer-von Mises (CVM), as quais são utilizadas para verificar se os dados são provenientes de alguma distribuição, nesse caso se os dados vêm da distribuição



Kumaraswamy inflacionada em zero ou da distribuição beta inflacionada em zero. Quanto maior forem essas estatísticas há mais indícios de que os dados não são provenientes das respectivas distribuições. Para a distribuição KIZ essas estatísticas são, respectivamente, iguais a 0.2832, 412.6745 e 72.4510, enquanto que para distribuição BIZ essas estatísticas são iguais a 0.3132, 448.6282 e 77.2219, indicando que há mais indícios dos dados serem provenientes da KIZ do que da distribuição BIZ.

### 2.7.2 Aplicação da distribuição KIU

A segunda aplicação foi realizada com base em um conjunto de dados contendo informações sobre a proporção da população que reside em domicílios com banheiro e água encanada nos 5566 municípios brasileiros no ano de 2010. As informações aqui utilizadas foram retiradas do Atlas do Desenvolvimento Humano no Brasil 2013 e podem ser encontradas no sítio <http://www.atlasbrasil.org.br/2013/pt/consulta/>. Esse conjunto de dados contém 73 valores iguais a um, e por isso será ajustada a distribuição Kumaraswamy inflacionada em um.

As medidas resumo desse conjunto de dados encontram-se na Tabela 2.9, na qual é possível observar que o maior valor que os dados assumem é o limite superior do intervalo  $(0, 1]$  e 75% das observações são maiores que 0.6778. A distribuição dos dados parece ser assimétrica à esquerda, exibindo forma de ‘J’, o que pode ser verificado na Figura 2.7 (a).

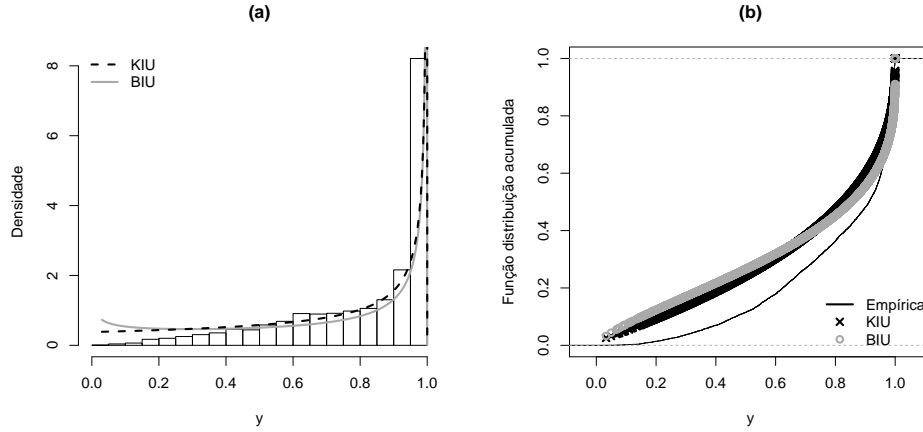
**Tabela 2.9:** Medidas resumo da proporção da população em domicílios com banheiro e água encanada.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
0.0326	0.6778	0.9124	0.8087	0.9800	1.0000

Além de ajustar a distribuição Kumaraswamy inflacionada em um, também foi ajustada a distribuição beta inflacionada em um. Maximizando as funções de log-verossimilhança dessas duas distribuições foram obtidas as estimativas de máxima verossimilhança, com seus respectivos erros-padrão (entre parênteses). Para a distribuição KIU essas estimativas são:  $\hat{\lambda} = 0.0131$  (0.0015),  $\hat{\alpha} = 0.9990$  (0.0255) e  $\hat{\beta} = 0.3846$  (0.0175) enquanto para a distribuição BIU são  $\hat{\lambda} = 0.0131$  (0.0015),  $\hat{\mu} = 0.7029$  (0.0036) e  $\hat{\phi} = 0.9990$  (0.0151).

Na Figura 2.7 estão apresentados o histograma da proporção da população que reside em domicílios com banheiro e água encanada, simultaneamente com as densidades KIU e BIU estimadas, além das funções distribuição acumuladas empírica, KIU e BIU.

Para verificar se os dados são provenientes da distribuição Kumaraswamy inflacionada em um ou se são oriundos da distribuição beta inflacionada em um, foram calculadas as estatísticas Kolmogorov-Smirnov e Cramer-von Mises utilizando a parte contínua das duas distribuições. Para a distribuição KIU essas estatísticas são, respectivamente, iguais a 0.1565 e 47.5689, enquanto que para distribuição BIU essas estatísticas são iguais a 0.2149 e 49.9574, indicando que há mais indícios dos dados serem provenientes da Kumaraswamy inflacionada em um do que da distribuição beta inflacionada em um.



**Figura 2.7:** (a) Histograma e densidades estimadas e (b) funções distribuição acumuladas para a proporção da população em domicílios com banheiro e água encanada.

## 2.8 Conclusões

Dados de taxas e proporções podem incluir zeros ou uns. Nesse caso distribuições como a beta e a Kumaraswamy não são adequadas, pois possuem suporte no intervalo  $(0, 1)$ . Neste capítulo foi proposta uma nova distribuição que permite a modelagem de zeros ou uns na amostra de dados: a distribuição Kumaraswamy inflacionada em  $c$ , em que  $c = 0$  ou  $c = 1$ . Essa nova distribuição permite modelar dados nos intervalos  $[0, 1)$  e  $(0, 1]$ .

Adicionalmente foram discutidas algumas propriedades dessa nova distribuição, bem como a estimação dos parâmetros através do método de máxima verossimilhança. A matriz de informação de Fisher foi derivada e, a partir dela, foram definidos os intervalos de confiança assintóticos; ademais foram apresentadas estatísticas de teste usadas para testar hipóteses sobre os parâmetros que indexam a distribuição Kumaraswamy inflacionada em  $c$ .

Para avaliar os desempenhos dos estimadores de máxima verossimilhança e dos testes da razão de verossimilhanças, Wald e escore foram realizadas simulações de Monte Carlo, nas quais se avaliaram propriedades desses estimadores, bem como os tamanhos e os poderes dos testes. Com base nos resultados encontrados é possível concluir que os estimadores de máxima verossimilhança dos parâmetros da distribuição Kumaraswamy inflacionada em  $c$  e também os testes de hipóteses usados apresentam bom desempenho.

Por fim foram realizadas duas aplicações a dados reais, ajustando as distribuições Kumaraswamy inflacionada em zero e Kumaraswamy inflacionada em um, sendo notável a aplicabilidade dessas distribuições para modelarem dados nos intervalos  $[0, 1)$  e  $(0, 1]$ , respectivamente.

## 3 DISTRIBUIÇÃO KUMARASWAMY INFLACIONADA EM ZERO E UM

### 3.1 Introdução

Existem situações em que os dados são observados no intervalo  $[0, 1]$ , por exemplo os dados na forma de taxas e proporções podem incluir tanto o valor zero quanto o valor um, e observações desta natureza não podem ser modeladas por uma distribuição puramente contínua, como a distribuição Kumaraswamy ou a distribuição beta, uma vez que o suporte de variáveis aleatórias com estas distribuições é  $(0, 1)$ , não permitindo a presença de valores discretos nos limites do suporte. Contudo, é possível criar uma nova distribuição que viabiliza a modelagem tanto da parte contínua quanto da parte discreta de dados que pertencem ao intervalo  $[0, 1]$ . No capítulo anterior foi proposta uma nova distribuição que atribui uma massa de probabilidade à constante  $c$  que pode assumir os valores zero ou um. Para o caso em que as proporções são observadas no intervalo  $[0, 1]$  a proposta é, analogamente ao caso de inflacionamento em apenas um dos limites do suporte, construir uma nova distribuição que resulta da mistura entre duas distribuições, uma contínua e uma discreta, permitindo assim modelar os valores zeros e uns em uma amostra de dados no intervalo  $[0, 1]$ .

Ospina e Ferrari (2010) propuseram uma distribuição inflacionada em zero e um considerando como base a distribuição beta que também serve para modelar observações no intervalo  $(0, 1)$ . Todavia, tendo em vista alguns benefícios da distribuição Kumaraswamy, desenvolvida por Kumaraswamy (1976), quando comparada à distribuição beta, a proposta deste capítulo é desenvolver uma nova distribuição a partir da mistura de uma distribuição discreta com a distribuição Kumaraswamy, posto que esta última apresenta algumas vantagens interessantes, como a simplicidade da função distribuição acumulada, que consequentemente fornece uma função quantílica simples, independente de qualquer função especial, como é o caso da distribuição beta, e isso permite facilmente utilizar a função inversa para gerar ocorrências de uma variável aleatória Kumaraswamy.

Adicionalmente, no que pertence à modelagem dos quantis a distribuição Kumaraswamy apresenta vantagem pelo mesmo motivo citado acima; além desta distribuição apresentar também bastante flexibilidade em seu formato, podendo sua densidade assumir as formas unimodal, ‘U’, crescente, decrescente e constante, decorrentes dos valores que os parâmetros desta distribuição assumem, podendo modelar desta maneira diversos conjuntos de dados que pertencem ao intervalo  $(0, 1)$  e ser particularmente útil em fornecer bom ajuste para dados que retratam fenômenos naturais cujos resultados têm limites inferior e superior, ou resultados também limitados em pesquisa biomédica e epidemiológica, segundo Wang et al. (2017).

Isto posto, ter-se-á uma distribuição que propicia o ajuste de dados em  $[0, 1]$  e que possui as mesmas utilidades encontradas na distribuição Kumaraswamy.

A seguir será apresentada a nova distribuição na Seção 3.2, bem como suas propriedades; estimação dos parâmetros que a indexam juntamente com sua matriz de informação de Fisher estão na Seção 3.3; resultados de simulações de Monte Carlo encontram-se na Seção 3.5.

## 3.2 Distribuição Kumaraswamy inflacionada em zero e um

Na prática os dados que representam proporções podem apresentar observações zeros e uns, sendo assim apenas a distribuição Kumaraswamy não é adequada para modelar esses dados. Então, tal como ajustado para o caso em que amostras de dados apresentam valores zeros ou uns, nesse caso também é possível ser feito um ajuste para que uma nova distribuição, proveniente da distribuição Kumaraswamy, comporte as observações no intervalo  $[0, 1]$ . Para esse tipo de observações propõe-se uma distribuição resultante de uma mistura entre as distribuições Kumaraswamy e Bernoulli, sendo esta última responsável por modelar os inteiros zero e um, atribuindo probabilidades não nulas a estes dois valores. A esta distribuição resultante desta mistura chama-se distribuição Kumaraswamy inflacionada zero e um, denotada por KIZU, em que a distribuição Kumaraswamy ajusta o componente contínuo dos dados e a distribuição Bernoulli modela o componente discreto.

Considere agora a variável  $Y$  com distribuição mistura Kumaraswamy inflacionada em zero e um cuja função distribuição acumulada é dada por

$$\text{KIZU}(y; \lambda, p, \alpha, \beta) = \lambda \text{Ber}(y; p) + (1 - \lambda)G(y; \alpha, \beta),$$

com  $y \in [0, 1]$ , em que  $\lambda \in (0, 1)$  é o parâmetro de mistura,  $\text{Ber}(y; p)$  denota a função distribuição acumulada de uma variável aleatória Bernoulli com parâmetro  $p$ , sendo  $p$  a probabilidade de  $Y$  ser igual a um e  $1 - p$  a probabilidade de  $Y$  ser igual a zero considerando a subamostra de valores discretos e por fim,  $G(y; \alpha, \beta)$  é a função distribuição acumulada da Kumaraswamy expressa no Capítulo 2 em (2.1), com  $\alpha > 0$  e  $\beta > 0$  sendo os parâmetros de forma da distribuição Kumaraswamy.

Dessa maneira,  $Y \in [0, 1]$  é uma variável aleatória Kumaraswamy inflacionada em zero e um, denotada por  $Y \sim \text{KIZU}(y; \lambda, p, \alpha, \beta)$ , se sua função densidade de probabilidade é dada por

$$\text{kizu}(y; \lambda, p, \alpha, \beta) = \begin{cases} \lambda p, & \text{se } y = 1, \\ \lambda(1 - p), & \text{se } y = 0, \\ (1 - \lambda)g(y; \alpha, \beta), & \text{se } y \in (0, 1), \end{cases} \quad (3.1)$$

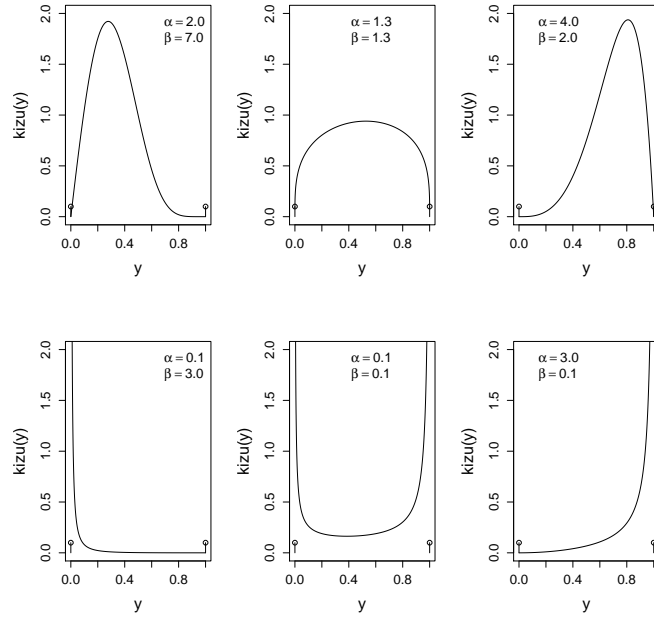
em que  $0 < \lambda < 1$  é o parâmetro de mistura,  $0 \leq p \leq 1$  é o parâmetro da distribuição Bernoulli e  $g(y; \alpha, \beta)$  é a densidade da distribuição Kumaraswamy dada em (2.2) com parâmetros  $\alpha > 0$  e  $\beta > 0$ . Note que  $\lambda p$  representa a probabilidade de  $Y$  ser igual a um e

$\lambda(1 - p)$  representa a probabilidade de  $Y$  ser igual a zero considerando toda a amostra de dados no intervalo  $[0, 1]$ , ou seja,  $\Pr(Y = 1) = \lambda p$  e  $\Pr(Y = 0) = \lambda(1 - p)$ . Já no caso em que  $y \in (0, 1)$ , é possível calcular

$$\Pr[Y \in (a, b)] = (1 - \lambda) \int_a^b g(y; \alpha, \beta) dy,$$

com  $0 < a < b < 1$ .

Nota-se na Figura 3.1 as diversas formas que a densidade Kumaraswamy inflacionada em zero e um pode assumir, dependendo dos valores dos parâmetros que a indexam. De maneira similar às distribuições Kumaraswamy e Kumaraswamy inflacionada em zero ou em um, a distribuição KIZU pode assumir as formas de ‘J’, ‘U’, ‘U’ invertido, bem como apresentar assimetria e simetria, além da forma constante. As linhas verticais na Figura 3.1 com um ponto em cima representam as probabilidades de  $Y$  ser igual a zero ou  $Y$  ser igual a um, neste caso  $\Pr(Y = 0) = \Pr(Y = 1) = 0.1$ .



**Figura 3.1:** Densidades Kumaraswamy inflacionadas em zero e um com  $\lambda = 0.2$ ,  $p = 0.5$  e diferentes valores de  $(\alpha, \beta)$ .

O  $r$ -ésimo momento de uma variável aleatória Kumaraswamy inflacionada em zero e um é dado por

$$\mathbb{E}(Y^r) = \lambda p + (1 - \lambda)\mu_r, \quad r = (1, 2, \dots), \quad (3.2)$$

em que  $\mu_r$  é o  $r$ -ésimo momento da distribuição Kumaraswamy apresentado em (2.5). A partir da expressão (3.2) tem-se que a média e a variância da variável aleatória com distribuição KIZU são

$$\begin{aligned}\mathbb{E}(Y) &= \lambda p + (1 - \lambda)\mu_1, \\ &= \lambda p + \beta(1 - \lambda)B\left(1 + \frac{1}{\alpha}, \beta\right) \quad \text{e}\end{aligned}$$

$$\begin{aligned}\text{Var}(Y) &= \lambda p + (1 - \lambda)\mu_2 - [\lambda p + (1 - \lambda)\mu_1]^2, \\ &= \lambda p + \beta(1 - \lambda)B\left(1 + \frac{2}{\alpha}, \beta\right) - \left[\lambda p + \beta(1 - \lambda)B\left(1 + \frac{1}{\alpha}, \beta\right)\right]^2 \\ &= \lambda p(1 - \lambda p) + (1 - \lambda)\beta \left\{ B\left(1 + \frac{2}{\alpha}, \beta\right) - B\left(1 + \frac{1}{\alpha}, \beta\right) [2\lambda p + \right. \\ &\quad \left. + \beta(1 - \lambda)B\left(1 + \frac{1}{\alpha}, \beta\right)] \right\},\end{aligned}$$

respectivamente, com  $\mu_1$  denotando a média e  $\mu_2$  o segundo momento da distribuição Kumaraswamy que envolvem a função beta simbolizada por  $B(\cdot, \cdot)$ .

### 3.3 Estimação por máxima verossimilhança

É possível reescrever a função densidade de probabilidade da distribuição Kumaraswamy inflacionada em zero e um expressa em (3.1) como

$$\begin{aligned}\text{kizu}(y; \lambda, p, \alpha, \beta) &= [\lambda p^y (1 - p)^{1-y}]^{I_{\{0,1\}}(y)} \times [(1 - \lambda)g(y; \alpha, \beta)]^{1-I_{\{0,1\}}(y)} \\ &= [\lambda^{I_{\{0,1\}}(y)} (1 - \lambda)^{1-I_{\{0,1\}}(y)}] \times [p^y (1 - p)^{1-y}]^{I_{\{0,1\}}(y)} \\ &\quad \times [g(y; \alpha, \beta)^{1-I_{\{0,1\}}(y)}],\end{aligned}$$

com  $y \in [0, 1]$ , em que  $I_{\{0,1\}}(y)$  é a função indicadora que é igual a um se  $y \in \{0, 1\}$ , ou seja, se  $y$  assume zero ou um, e é igual a zero quando  $y \notin \{0, 1\}$ . Note que a densidade fatora em três termos: o primeiro termo dependendo apenas de  $\lambda$ , o segundo termo dependendo somente de  $p$  e o terceiro termo envolvendo  $\alpha$  e  $\beta$ .

Seja  $\boldsymbol{\theta} = (\lambda, p, \alpha, \beta)^\top$  o vetor de parâmetros da distribuição Kumaraswamy inflacionada em zero e um. A função de verossimilhança para  $\boldsymbol{\theta} = (\lambda, p, \alpha, \beta)^\top$  com base na amostra aleatória  $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$  é

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^n \text{kizu}(y_i; \lambda, p, \alpha, \beta) = L_1(\lambda; \mathbf{y}) L_2(p; \mathbf{y}) L_3(\alpha, \beta; \mathbf{y}),$$

em que

$$\begin{aligned}
L_1(\lambda; \mathbf{y}) &= \prod_{i=1}^n \lambda^{\mathbb{I}_{\{0,1\}}(y_i)} (1-\lambda)^{1-\mathbb{I}_{\{0,1\}}(y_i)} = \lambda^{\sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i)} (1-\lambda)^{n-\sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i)}, \\
L_2(p; \mathbf{y}) &= \prod_{i=1}^n [p^{y_i} (1-p)^{1-y_i}]^{\mathbb{I}_{\{0,1\}}(y_i)} = p^{\sum_{i=1}^n y_i \mathbb{I}_{\{0,1\}}(y_i)} (1-p)^{\sum_{i=1}^n (1-y_i) \mathbb{I}_{\{0,1\}}(y_i)} \\
&= p^{\sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i)} (1-p)^{[\sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) - \sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i)]} \quad e \\
L_3(\alpha, \beta; \mathbf{y}) &= \prod_{i=1}^n g(y_i; \alpha, \beta)^{1-\mathbb{I}_{\{0,1\}}(y_i)} \\
&= (\alpha\beta)^{[n-\sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i)]} \prod_{i=1}^n y_i^{(\alpha-1)[1-\mathbb{I}_{\{0,1\}}(y_i)]} (1-y_i)^{(\beta-1)[1-\mathbb{I}_{\{0,1\}}(y_i)]}.
\end{aligned}$$

Note que além da função indicadora já definida,  $\mathbb{I}_{\{0,1\}}(y_i)$ , tem-se ainda  $\mathbb{I}_{\{1\}}(y_i)$  que representa a função de indicadora para  $y = 1$ , ou seja,  $\mathbb{I}_{\{1\}}(y_i) = 1$ , se  $y = 1$  e  $\mathbb{I}_{\{1\}}(y_i) = 0$  caso contrário.

Observe ainda que a função de verossimilhança  $L(\boldsymbol{\theta}; \mathbf{y})$  também pode ser fatorada em três termos, em que o primeiro termo envolve apenas  $\lambda$ , o segundo termo depende apenas de  $p$  e o terceiro e último termo é função de  $\alpha$  e  $\beta$ .

Assim, a função de log-verossimilhança da distribuição Kumaraswamy inflacionada em zero e um pode ser escrita como

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \ell_1(\lambda; \mathbf{y}) + \ell_2(p; \mathbf{y}) + \ell_3(\alpha, \beta; \mathbf{y}),$$

em que

$$\begin{aligned}
\ell_1(\lambda; \mathbf{y}) &= \ln(\lambda) \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) + \ln(1-\lambda) \left[ n - \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) \right] \\
\ell_2(p; \mathbf{y}) &= \ln(p) \sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i) + \ln(1-p) \left[ \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) - \sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i) \right] \quad e \\
\ell_3(\alpha, \beta; \mathbf{y}) &= \ln(\alpha\beta) \left[ n - \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) \right] + (\alpha-1) \sum_{i=1}^n \ln(y_i) [1 - \mathbb{I}_{\{0,1\}}(y_i)] + \\
&\quad (\beta-1) \sum_{i=1}^n \ln(1-y_i) [1 - \mathbb{I}_{\{0,1\}}(y_i)].
\end{aligned}$$

Note que para o cálculo da log-verossimilhança é preciso adotar a seguinte convenção:  $\ln(0) \times 0 = 0$ . Esta notação será utilizada convenientemente nas inferências referentes à distribuição proposta neste capítulo.

Para obter a função escore, a função de log-verossimilhança,  $\ell(\boldsymbol{\theta}; \mathbf{y})$ , foi diferenciada com respeito a cada um dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um. A função escore é dada por  $U(\boldsymbol{\theta}) = [U_\lambda(\lambda), U_p(p), U_\alpha(\alpha, \beta), U_\beta(\alpha, \beta)]$ , em que

$$\begin{aligned}
U_\lambda(\lambda) &= \frac{\partial \ell_1(\lambda; \mathbf{y})}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) - \frac{1}{1-\lambda} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) \right], \\
U_p(p) &= \frac{\partial \ell_2(p; \mathbf{y})}{\partial p} = \frac{1}{p} \sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i) - \frac{1}{1-p} \left[ \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) - \sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i) \right], \\
U_\alpha(\alpha, \beta) &= \frac{\partial \ell_3(\alpha, \beta; \mathbf{y})}{\partial \alpha} = \frac{1}{\alpha} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) \right] + \sum_{i=1}^n \ln(y_i) [1 - \mathbb{I}_{\{0,1\}}(y_i)] + \\
&\quad (\beta - 1) \sum_{i=1}^n \frac{y_i^\alpha}{y_i^\alpha - 1} \ln(y_i) [1 - \mathbb{I}_{\{0,1\}}(y_i)] \quad \text{e} \\
U_\beta(\alpha, \beta) &= \frac{\partial \ell_3(\alpha, \beta; \mathbf{y})}{\partial \beta} = \frac{1}{\beta} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) \right] + \sum_{i=1}^n \ln(1 - y_i^\alpha) [1 - \mathbb{I}_{\{0,1\}}(y_i)].
\end{aligned}$$

É possível encontrar os estimadores de máxima verossimilhança de  $\lambda$  e  $p$ , resolvendo as equações  $U_\lambda(\hat{\lambda}) = 0$  e  $U_p(\hat{p}) = 0$ , respectivamente. A partir daí encontram-se os estimadores de máxima verossimilhança de  $\lambda$  e  $p$  que são, respectivamente,  $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i)$ , que representa a proporção de valores na amostra que são iguais a zero ou a um, e  $\hat{p} = \sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i) / \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i)$ , que representa a proporção de valores iguais a um, considerando agora apenas a parte discreta da amostra. No caso dos estimadores de  $\alpha$  e  $\beta$  é preciso utilizar um método de otimização não-linear para obtê-los.

Calculando as segundas derivadas da função de log-verossimilhança é possível obter a matriz de informação observada e posteriormente calcular a matriz de informação esperada. As segundas derivadas são dadas por

$$\begin{aligned}
\frac{\partial^2 \ell_1(\lambda; \mathbf{y})}{\partial \lambda \partial \lambda} &= -\frac{1}{\lambda^2} \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) - \frac{1}{(1-\lambda)^2} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) \right], \\
\frac{\partial^2 \ell_2(p; \mathbf{y})}{\partial p \partial p} &= -\frac{1}{p^2} \sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i) - \frac{1}{(1-p)^2} \left[ \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) - \sum_{i=1}^n \mathbb{I}_{\{1\}}(y_i) \right], \\
\frac{\partial^2 \ell_3(\alpha, \beta; \mathbf{y})}{\partial \alpha \partial \alpha} &= -\frac{1}{\alpha^2} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) \right] - (\beta - 1) \sum_{i=1}^n \frac{y_i^\alpha}{(y_i^\alpha - 1)^2} \ln^2(y_i) [1 - \mathbb{I}_{\{0,1\}}(y_i)], \\
\frac{\partial^2 \ell_3(\alpha, \beta; \mathbf{y})}{\partial \beta \partial \beta} &= -\frac{1}{\beta^2} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{0,1\}}(y_i) \right] \quad \text{e} \\
\frac{\partial^2 \ell_3(\alpha, \beta; \mathbf{y})}{\partial \alpha \partial \beta} &= \frac{\partial^2 \ell_3(\alpha, \beta; \mathbf{y})}{\partial \beta \partial \alpha} = \sum_{i=1}^n \frac{y_i^\alpha}{y_i^\alpha - 1} \ln(y_i) [1 - \mathbb{I}_{\{0,1\}}(y_i)].
\end{aligned}$$



Note que

$$\begin{aligned} \frac{\partial^2 \ell_1(\lambda; \mathbf{y})}{\partial \lambda \partial p} &= \frac{\partial^2 \ell_1(\lambda; \mathbf{y})}{\partial \lambda \partial \alpha} = \frac{\partial^2 \ell_1(\lambda; \mathbf{y})}{\partial \lambda \partial \beta} = \frac{\partial^2 \ell_2(p; \mathbf{y})}{\partial p \partial \lambda} = \frac{\partial^2 \ell_2(p; \mathbf{y})}{\partial p \partial \alpha} = \frac{\partial^2 \ell_2(p; \mathbf{y})}{\partial p \partial \beta} = \\ &= \frac{\partial^2 \ell_3(\alpha, \beta; \mathbf{y})}{\partial \alpha \partial \lambda} = \frac{\partial^2 \ell_3(\alpha, \beta; \mathbf{y})}{\partial \alpha \partial p} = \frac{\partial^2 \ell_3(\alpha, \beta; \mathbf{y})}{\partial \beta \partial \lambda} = \frac{\partial^2 \ell_3(\alpha, \beta; \mathbf{y})}{\partial \beta \partial p} = 0. \end{aligned}$$

Para obter a matriz de informação esperada de Fisher calcula-se a esperança das segundas derivadas multiplicadas por  $-1$ , resultando em

$$K(\boldsymbol{\theta}) = \begin{pmatrix} k_{\lambda\lambda} & 0 & 0 & 0 \\ 0 & k_{pp} & 0 & 0 \\ 0 & 0 & k_{\alpha\alpha} & k_{\alpha\beta} \\ 0 & 0 & k_{\beta\alpha} & k_{\beta\beta} \end{pmatrix}, \quad (3.3)$$

em que

$$\begin{aligned} k_{\lambda\lambda} &= \frac{n}{\lambda(1-\lambda)}, \\ k_{pp} &= \frac{n\lambda}{p(1-p)}, \\ k_{\alpha\alpha} &= \frac{n(1-\lambda)}{\alpha^2} + \frac{n\beta(1-\lambda)}{\alpha^2(\beta-2)} \{[\psi(\beta) - \psi(2)]^2 - [\psi'(\beta) - \psi'(2)]\}, \\ k_{\alpha\beta} &= k_{\beta\alpha} = -\frac{n(1-\lambda)}{\alpha(\beta-1)} \{[\psi(\beta+1) - \psi(2)]\}, \\ k_{\beta\beta} &= \frac{n(1-\lambda)}{\beta^2}, \end{aligned}$$

sendo  $\psi(\cdot)$  a função digama, isto é,  $\psi(\cdot) = \partial \ln \Gamma(\cdot) / \partial z$ , em que  $\Gamma(\cdot)$  é a função gama. Então,  $\psi'(\cdot)$  é a derivada de primeira ordem da função  $\psi(\cdot)$ , sendo conhecida como função trigama.

Sabe-se que os estimadores de máxima verossimilhança têm distribuição assintótica normal, com média  $\boldsymbol{\theta}$  e variância dada pela matriz inversa da informação de Fisher, como denotado a seguir

$$\hat{\boldsymbol{\theta}} \stackrel{a}{\sim} \mathcal{N}_4(\boldsymbol{\theta}, K(\boldsymbol{\theta})^{-1}),$$

em que  $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{p}, \hat{\alpha}, \hat{\beta})$  é o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ ,  $\stackrel{a}{\sim}$  denota assintoticamente distribuído,  $\mathcal{N}_4$  denota a distribuição normal tetravariada e  $K(\boldsymbol{\theta})^{-1}$  é a inversa da matriz de informação de Fisher. Com base neste resultado, é possível construir intervalos de confiança assintóticos para  $\lambda, p, \alpha$  e  $\beta$ , dados por

$$\hat{\lambda} \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\lambda}), \quad \hat{p} \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{p}), \quad \hat{\alpha} \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\alpha}) \quad \text{e} \quad \hat{\beta} \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\beta}),$$

respectivamente, em que  $\text{ep}(\cdot)$  é o erro-padrão associado a cada estimador, que é dado pela raiz quadrada da variância estimada do estimador,  $z_{(1-\frac{\delta}{2})}$  é o quantil  $1 - \frac{\delta}{2}$  da distribuição normal padrão e  $1 - \delta$  é o nível de confiança adotado. As variâncias dos estimadores são obtidas pelos elementos da diagonal de  $K(\boldsymbol{\theta})^{-1}$  avaliados em  $\hat{\boldsymbol{\theta}}$ .

### 3.4 Testes de Hipóteses

Além de realizar estimação pontual e intervalar é possível testar hipóteses sobre os parâmetros que indexam a distribuição Kumaraswamy inflacionada em zero e um. Para verificar a veracidade das hipóteses estabelecidas são utilizadas estatísticas de teste, calculadas a partir de uma amostra, que são usadas para decidir pela rejeição ou não da hipótese nula ( $H_0$ ), isto é, da hipótese assumida como verdadeira.

Neste trabalho, serão usadas as estatísticas da razão de verossimilhanças, Wald e escore para testar hipóteses sobre os parâmetros da distribuição Kumaraswamy inflacionada em zero e um. O teste da razão de verossimilhanças baseia-se na comparação de dois modelos: um restrito que é obtido impondo-se o valor atribuído na hipótese nula ao parâmetro testado e outro irrestrito que não leva em consideração a restrição da hipótese nula. Já a estatística do teste Wald se baseia na estimativa de máxima verossimilhança irrestrita e na matriz de informação de Fisher avaliada nas estimativas irrestritas. Por fim, a estatística escore utiliza a função escore, que é dada pelas primeiras derivadas da função de log-verossimilhança, e também a matriz de informação de Fisher, ambas avaliadas nas estimativas restritas.

Estas três estatísticas, sob  $H_0$  e considerando determinadas condições de regularidade (Lehmann, 2006), têm distribuição assintótica  $\chi^2_\nu$ , em que  $\nu$  indica o número de restrições impostas na hipótese nula. Sendo assim, é possível definir a região crítica com base na distribuição assintótica. Logo,  $H_0$  é rejeitada se a estatística de teste for maior do que  $\chi^2_{\nu, 1-\delta}$ , em que  $\chi^2_{\nu, 1-\delta}$  é o quantil  $1 - \delta$  da distribuição  $\chi^2$  com  $\nu$  graus de liberdade.

Para maiores detalhes sobre essas três estatísticas de teste ver as descrições das mesmas na Seção 2.5.

### 3.5 Resultados numéricos

De forma análoga à distribuição Kumaraswamy inflacionada em  $c$ , também foram realizadas simulações de Monte Carlo para verificar o desempenho dos estimadores de máxima verossimilhança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um. Foram calculadas algumas propriedades destes estimadores, a saber, média, variância, viés e erro quadrático médio (EQM).

Adicionalmente, foram testadas hipóteses a fim de verificar o desempenho dos testes da razão de verossimilhanças, Wald e escore, baseados nas estatísticas descritas na Seção 2.5. As hipóteses foram  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$ , sendo  $\lambda_0 = (0.05, 0.10, 0.20, 0.50)$ , e  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$ , sendo  $p_0 = (0.05, 0.10, 0.50, 0.75)$ . Os valores de  $\alpha$  e  $\beta$  foram

fixados em 1.5 e 3.0, respectivamente. Note que, para os respectivos valores de  $\lambda_0$  e  $p_0$ , temos as seguintes probabilidades para  $Y$  igual a zero: (0.0475, 0.0450, 0.0250, 0.0125) e para  $Y$  igual a um: (0.0025, 0.0050, 0.0250, 0.0375).

O número de réplicas de Monte Carlo foi 10000 e os tamanhos amostrais foram  $n = (20, 30, 50, 100, 200, 500, 1000)$ .

As simulações foram realizadas utilizando a linguagem matricial de programação Ox. Maximizações numéricas da função de log-verossimilhança foram feitas usando o método quasi-Newton BFGS (Nocedal, 1980) com primeiras derivadas analíticas. Os chutes iniciais utilizados nas simulações foram escolhidos arbitrariamente.

### 3.5.1 Estimação pontual

Foram avaliados os desempenhos dos estimadores de máxima verossimilhança dos parâmetros que indexam a distribuição Kumaraswamy inflacionada em zero e um. Nas Tabelas 3.1 e 3.2 estão apresentados a média, a variância, o viés e o erro quadrático médio desses estimadores.

Na Tabela 3.1 encontram-se os resultados das simulações considerando  $p = 0.5, \alpha = 1.5, \beta = 3.0$  e  $\lambda = (0.05, 0.25, 0.50)$ . Já na Tabela 3.2 os valores dos parâmetros fixados são  $\lambda = 0.5, \alpha = 1.5, \beta = 3.0$  e variou-se o valor de  $p$ ,  $p = (0.25, 0.50, 0.75)$ .

Observa-se na Tabela 3.1 que, à medida em que o tamanho amostral aumenta, as estimativas médias convergem para o valor verdadeiro, bem como, a variância, o viés e o erro quadrático médio diminuem com o aumento de  $n$ , como esperado. Uma vez que esses estimadores são assintoticamente não viesados, nota-se que o viés tende a zero quando  $n$  aumenta. Assim sendo, os estimadores de máxima verossimilhança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um apresentam bom desempenho.

Analogamente aos resultados apresentados na Tabela 3.1, ao analisar a Tabela 3.2 percebe-se que as médias das estimativas de Monte Carlo convergem para o verdadeiro valor considerado para cada parâmetro à medida em que o tamanho da amostra aumenta. O viés de cada estimador tende para zero à medida em que  $n$  tende para infinito. Adicionalmente, a variância e, conseqüentemente, o erro quadrático médio diminuem ao se aumentar o tamanho da amostra. Assim, os resultados exibidos na Tabela 3.2 reforçam o bom desempenho dos estimadores de máxima verossimilhança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um.

**Tabela 3.1:** Médias, variâncias, vieses e EQMs dos estimadores dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de  $\lambda$ ;  $p = 0.5, \alpha = 1.5, \beta = 3.0$ .

$\lambda$	Medida	Estimador	$n$				
			30	50	100	200	500
0.05	Média	$\hat{\lambda}$	0.0934	0.0696	0.0544	0.0504	0.0501
		$\hat{p}$	0.5001	0.4986	0.4987	0.4993	0.4997
		$\hat{\alpha}$	1.6058	1.5585	1.5277	1.5126	1.5057
		$\hat{\beta}$	3.5465	3.2891	3.1317	3.0626	3.0268
	Variância	$\hat{\lambda}$	0.0010	0.0007	0.0004	0.0002	0.0001
		$\hat{p}$	0.0167	0.0243	0.0310	0.0255	0.0104
		$\hat{\alpha}$	0.1103	0.0593	0.0267	0.0130	0.0051
		$\hat{\beta}$	2.0089	0.8137	0.3104	0.1381	0.0521
	Viés	$\hat{\lambda}$	0.0434	0.0196	0.0044	0.0004	0.0001
		$\hat{p}$	0.0001	-0.0014	-0.0013	-0.0007	-0.0003
		$\hat{\alpha}$	0.1058	0.0585	0.0277	0.0126	0.0057
		$\hat{\beta}$	0.5465	0.2891	0.1317	0.0626	0.0268
	EQM	$\hat{\lambda}$	0.0028	0.0011	0.0004	0.0002	0.0001
		$\hat{p}$	0.0167	0.0243	0.0310	0.0255	0.0104
		$\hat{\alpha}$	0.1215	0.0627	0.0275	0.0132	0.0052
		$\hat{\beta}$	2.3075	0.8973	0.3278	0.1420	0.0528
0.25	Média	$\hat{\lambda}$	0.2532	0.2497	0.2490	0.2495	0.2495
		$\hat{p}$	0.5022	0.5031	0.5009	0.5008	0.5001
		$\hat{\alpha}$	1.6339	1.5746	1.5359	1.5166	1.5068
		$\hat{\beta}$	3.7044	3.3645	3.1702	3.0794	3.0318
	Variância	$\hat{\lambda}$	0.0059	0.0037	0.0019	0.0009	0.0004
		$\hat{p}$	0.0296	0.0211	0.0104	0.0050	0.0020
		$\hat{\alpha}$	0.1468	0.0752	0.0347	0.0166	0.0063
		$\hat{\beta}$	3.2115	1.1414	0.4199	0.1811	0.0655
	Viés	$\hat{\lambda}$	0.0032	-0.0003	-0.0010	-0.0005	-0.0005
		$\hat{p}$	0.0022	0.0031	0.0009	0.0008	0.0001
		$\hat{\alpha}$	0.1339	0.0746	0.0359	0.0166	0.0068
		$\hat{\beta}$	0.7044	0.3645	0.1702	0.0794	0.0318
	EQM	$\hat{\lambda}$	0.0059	0.0037	0.0019	0.0009	0.0004
		$\hat{p}$	0.0296	0.0211	0.0104	0.0050	0.0020
		$\hat{\alpha}$	0.1648	0.0808	0.0360	0.0169	0.0064
		$\hat{\beta}$	3.7077	1.2743	0.4488	0.1874	0.0665
0.50	Média	$\hat{\lambda}$	0.4995	0.4992	0.4992	0.4997	0.4998
		$\hat{p}$	0.4997	0.5012	0.5006	0.5004	0.5001
		$\hat{\alpha}$	1.7135	1.6169	1.5539	1.5245	1.5107
		$\hat{\beta}$	4.3928	3.6099	3.2605	3.1185	3.0497
	Variância	$\hat{\lambda}$	0.0080	0.0049	0.0025	0.0012	0.0005
		$\hat{p}$	0.0170	0.0101	0.0051	0.0025	0.0010
		$\hat{\alpha}$	0.2654	0.1283	0.0552	0.0253	0.0095
		$\hat{\beta}$	86.4677	2.6944	0.7409	0.2873	0.0996
	Viés	$\hat{\lambda}$	-0.0005	-0.0008	-0.0008	-0.0003	-0.0002
		$\hat{p}$	-0.0003	0.0012	0.0006	0.0004	0.0001
		$\hat{\alpha}$	0.2135	0.1169	0.0539	0.0245	0.0107
		$\hat{\beta}$	1.3928	0.6099	0.2605	0.1185	0.0497
	EQM	$\hat{\lambda}$	0.0080	0.0049	0.0025	0.0012	0.0005
		$\hat{p}$	0.0170	0.0102	0.0051	0.0025	0.0010
		$\hat{\alpha}$	0.3109	0.1419	0.0581	0.0259	0.0096
		$\hat{\beta}$	88.4077	3.0663	0.8087	0.3013	0.1020

**Tabela 3.2:** Médias, variâncias, vieses e EQMs dos estimadores dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de  $p$ ;  $\lambda = 0.5, \alpha = 1.5, \beta = 3.0$ .

$p$	Medida	Estimador	$n$				
			30	50	100	200	500
0.25	Média	$\hat{\lambda}$	0.5009	0.4993	0.4992	0.4997	0.4998
		$\hat{p}$	0.2555	0.2509	0.2504	0.2502	0.2499
		$\hat{\alpha}$	1.7144	1.6169	1.5539	1.5245	1.5107
		$\hat{\beta}$	4.3970	3.6098	3.2605	3.1185	3.0497
	Variância	$\hat{\lambda}$	0.0080	0.0049	0.0025	0.0012	0.0005
		$\hat{p}$	0.0118	0.0075	0.0037	0.0019	0.0007
		$\hat{\alpha}$	0.2661	0.1282	0.0552	0.0253	0.0095
		$\hat{\beta}$	86.5866	2.6926	0.7409	0.2873	0.0996
	Viés	$\hat{\lambda}$	0.0009	-0.0007	-0.0008	-0.0003	-0.0002
		$\hat{p}$	0.0055	0.0009	0.0004	0.0002	-0.0001
		$\hat{\alpha}$	0.2144	0.1169	0.0539	0.0245	0.0107
		$\hat{\beta}$	1.3970	0.6098	0.2605	0.1185	0.0497
	EQM	$\hat{\lambda}$	0.0080	0.0049	0.0025	0.0012	0.0005
		$\hat{p}$	0.0118	0.0075	0.0037	0.0019	0.0007
		$\hat{\alpha}$	0.3121	0.1419	0.0581	0.0259	0.0096
		$\hat{\beta}$	88.5383	3.0645	0.8087	0.3013	0.1020
0.50	Média	$\hat{\lambda}$	0.4995	0.4992	0.4992	0.4997	0.4998
		$\hat{p}$	0.4997	0.5012	0.5006	0.5004	0.5001
		$\hat{\alpha}$	1.7135	1.6169	1.5539	1.5245	1.5107
		$\hat{\beta}$	4.3928	3.6099	3.2605	3.1185	3.0497
	Variância	$\hat{\lambda}$	0.0080	0.0049	0.0025	0.0012	0.0005
		$\hat{p}$	0.0170	0.0101	0.0051	0.0025	0.0010
		$\hat{\alpha}$	0.2654	0.1283	0.0552	0.0253	0.0095
		$\hat{\beta}$	86.4677	2.6944	0.7409	0.2873	0.0996
	Viés	$\hat{\lambda}$	-0.0005	-0.0008	-0.0008	-0.0003	-0.0002
		$\hat{p}$	-0.0003	0.0012	0.0006	0.0004	0.0001
		$\hat{\alpha}$	0.2135	0.1169	0.0539	0.0245	0.0107
		$\hat{\beta}$	1.3928	0.6099	0.2605	0.1185	0.0497
	EQM	$\hat{\lambda}$	0.0080	0.0049	0.0025	0.0012	0.0005
		$\hat{p}$	0.0170	0.0102	0.0051	0.0025	0.0010
		$\hat{\alpha}$	0.3109	0.1419	0.0581	0.0259	0.0096
		$\hat{\beta}$	88.4077	3.0663	0.8087	0.3013	0.1020
0.75	Média	$\hat{\lambda}$	0.5010	0.4993	0.4992	0.4997	0.4998
		$\hat{p}$	0.7451	0.7507	0.7507	0.7507	0.7503
		$\hat{\alpha}$	1.7142	1.6167	1.5539	1.5245	1.5107
		$\hat{\beta}$	4.3949	3.6096	3.2605	3.1185	3.0497
	Variância	$\hat{\lambda}$	0.0079	0.0049	0.0025	0.0012	0.0005
		$\hat{p}$	0.0121	0.0077	0.0038	0.0019	0.0008
		$\hat{\alpha}$	0.2674	0.1283	0.0552	0.0253	0.0095
		$\hat{\beta}$	86.5739	2.6947	0.7409	0.2873	0.0996
	Viés	$\hat{\lambda}$	0.0010	-0.0007	-0.0008	-0.0003	-0.0002
		$\hat{p}$	-0.0049	0.0007	0.0007	0.0007	0.0003
		$\hat{\alpha}$	0.2142	0.1167	0.0539	0.0245	0.0107
		$\hat{\beta}$	1.3949	0.6096	0.2605	0.1185	0.0497
	EQM	$\hat{\lambda}$	0.0079	0.0049	0.0025	0.0012	0.0005
		$\hat{p}$	0.0122	0.0077	0.0038	0.0019	0.0008
		$\hat{\alpha}$	0.3133	0.1419	0.0581	0.0259	0.0096
		$\hat{\beta}$	88.5195	3.0663	0.8087	0.3013	0.1020

### 3.5.2 Estimação intervalar

Foram calculados os intervalos de confiança assintóticos para os parâmetros da distribuição Kumaraswamy inflacionada em zero e um, descritos na Seção 3.3. Adicionalmente, também foram calculadas a amplitudes média dos intervalos, as proporções de cobertura e não cobertura.

**Tabela 3.3:** Amplitudes médias dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de  $\lambda$ ;  $p = 0.5$ ,  $\alpha = 1.5$  e  $\beta = 3.0$ .

$\lambda$	$1 - \delta$	$\hat{\theta}$	$n$				
			30	50	100	500	1000
0.05	99%	$\hat{\lambda}$	0.2696	0.1819	0.1146	0.0500	0.0354
		$\hat{p}$	1.5449	1.3821	1.0909	0.5109	0.3629
		$\hat{\alpha}$	1.6070	1.2025	0.8315	0.3672	0.2593
		$\hat{\beta}$	6.0639	4.1854	2.7483	1.1717	0.8239
	95%	$\hat{\lambda}$	0.2052	0.1384	0.0872	0.0380	0.0270
		$\hat{p}$	1.1755	1.0517	0.8301	0.3887	0.2761
		$\hat{\alpha}$	1.2227	0.9150	0.6327	0.2794	0.1973
		$\hat{\beta}$	4.6141	3.1847	2.0912	0.8916	0.6269
	90%	$\hat{\lambda}$	0.1722	0.1162	0.0732	0.0319	0.0226
		$\hat{p}$	0.9865	0.8826	0.6966	0.3262	0.2317
		$\hat{\alpha}$	1.0262	0.7679	0.5310	0.2345	0.1656
		$\hat{\beta}$	3.8723	2.6727	1.7550	0.7482	0.5261
0.25	99%	$\hat{\lambda}$	0.4003	0.3111	0.2213	0.0996	0.0705
		$\hat{p}$	0.9077	0.7118	0.5109	0.2302	0.1628
		$\hat{\alpha}$	1.8016	1.3530	0.9377	0.4134	0.2919
		$\hat{\beta}$	7.1437	4.8213	3.1385	1.3215	0.9280
	95%	$\hat{\lambda}$	0.3046	0.2367	0.1684	0.0758	0.0536
		$\hat{p}$	0.6907	0.5416	0.3888	0.1751	0.1239
		$\hat{\alpha}$	1.3708	1.0295	0.7135	0.3146	0.2221
		$\hat{\beta}$	5.4357	3.6686	2.3881	1.0055	0.7061
	90%	$\hat{\lambda}$	0.2556	0.1986	0.1413	0.0636	0.0450
		$\hat{p}$	0.5797	0.4546	0.3263	0.1470	0.1040
		$\hat{\alpha}$	1.1504	0.8640	0.5988	0.2640	0.1864
		$\hat{\beta}$	4.5617	3.0788	2.0042	0.8439	0.5926
0.50	99%	$\hat{\lambda}$	0.4625	0.3606	0.2563	0.1151	0.0814
		$\hat{p}$	0.6492	0.5085	0.3622	0.1627	0.1151
		$\hat{\alpha}$	2.3023	1.6981	1.1611	0.5075	0.3580
		$\hat{\beta}$	11.7802	6.5632	4.0061	1.6320	1.1416
	95%	$\hat{\lambda}$	0.3519	0.2744	0.1950	0.0876	0.0619
		$\hat{p}$	0.4940	0.3869	0.2756	0.1238	0.0876
		$\hat{\alpha}$	1.7519	1.2921	0.8835	0.3862	0.2724
		$\hat{\beta}$	8.9636	4.9940	3.0482	1.2418	0.8687
	90%	$\hat{\lambda}$	0.2954	0.2303	0.1637	0.0735	0.0520
		$\hat{p}$	0.4146	0.3247	0.2313	0.1039	0.0735
		$\hat{\alpha}$	1.4702	1.0844	0.7415	0.3241	0.2286
		$\hat{\beta}$	7.5225	4.1911	2.5582	1.0421	0.7290

Os níveis de confiança adotados nessas simulações foram  $1 - \delta = (99\%, 95\%, 90\%)$ , os parâmetros  $\alpha$  e  $\beta$  foram fixados em 1.5 e 3.0, respectivamente, e variou-se  $\lambda$  e  $p$ , sendo considerados os seguintes valores:  $\lambda = (0.05, 0.25, 0.50)$  e  $p = (0.25, 0.50, 0.75)$ .

A amplitude média dos intervalos de confiança calculados variando o parâmetro de mistura  $\lambda$ , estão reportados na Tabela 3.3 e as amplitudes variando o parâmetro  $p$  estão apresentados na Tabela 3.4.

**Tabela 3.4:** Amplitudes médias dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de  $p$ ;  $\lambda = 0.3$ ,  $\alpha = 1.5$  e  $\beta = 3.0$ .

$p$	$1 - \delta$	$\hat{\theta}$	$n$				
			30	50	100	500	1000
0.25	99%	$\hat{\lambda}$	0.4251	0.3301	0.2344	0.1054	0.0746
		$\hat{p}$	0.7431	0.5658	0.4026	0.1817	0.1286
		$\hat{\alpha}$	1.8772	1.4064	0.9723	0.4280	0.3022
		$\hat{\beta}$	7.6192	5.0663	3.2690	1.3696	0.9612
	95%	$\hat{\lambda}$	0.3235	0.2512	0.1784	0.0802	0.0568
		$\hat{p}$	0.5655	0.4305	0.3063	0.1383	0.0979
		$\hat{\alpha}$	1.4284	1.0701	0.7398	0.3257	0.2299
		$\hat{\beta}$	5.7975	3.8550	2.4874	1.0422	0.7313
	90%	$\hat{\lambda}$	0.2715	0.2108	0.1497	0.0673	0.0476
		$\hat{p}$	0.4745	0.3613	0.2571	0.1160	0.0821
		$\hat{\alpha}$	1.1987	0.8981	0.6209	0.2733	0.1930
		$\hat{\beta}$	4.8654	3.2352	2.0875	0.8746	0.6138
0.50	99%	$\hat{\lambda}$	0.4231	0.3297	0.2344	0.1054	0.0746
		$\hat{p}$	0.8315	0.6518	0.4668	0.2101	0.1486
		$\hat{\alpha}$	1.8718	1.4055	0.9723	0.4280	0.3022
		$\hat{\beta}$	7.5466	5.0609	3.2686	1.3696	0.9612
	95%	$\hat{\lambda}$	0.3219	0.2509	0.1784	0.0802	0.0568
		$\hat{p}$	0.6327	0.4959	0.3552	0.1599	0.1131
		$\hat{\alpha}$	1.4243	1.0695	0.7398	0.3257	0.2299
		$\hat{\beta}$	5.7422	3.8509	2.4871	1.0422	0.7313
	90%	$\hat{\lambda}$	0.2702	0.2105	0.1497	0.0673	0.0476
		$\hat{p}$	0.5310	0.4162	0.2981	0.1342	0.0949
		$\hat{\alpha}$	1.1953	0.8975	0.6209	0.2733	0.1930
		$\hat{\beta}$	4.8190	3.2318	2.0873	0.8746	0.6138
0.75	99%	$\hat{\lambda}$	0.4249	0.3301	0.2344	0.1054	0.0746
		$\hat{p}$	0.7423	0.5627	0.4014	0.1817	0.1286
		$\hat{\alpha}$	1.8776	1.4061	0.9723	0.4280	0.3022
		$\hat{\beta}$	7.6285	5.0628	3.2689	1.3696	0.9612
	95%	$\hat{\lambda}$	0.3233	0.2511	0.1784	0.0802	0.0568
		$\hat{p}$	0.5648	0.4282	0.3055	0.1383	0.0979
		$\hat{\alpha}$	1.4286	1.0699	0.7398	0.3257	0.2299
		$\hat{\beta}$	5.8046	3.8523	2.4873	1.0422	0.7313
	90%	$\hat{\lambda}$	0.2714	0.2108	0.1497	0.0673	0.0476
		$\hat{p}$	0.4740	0.3593	0.2564	0.1160	0.0821
		$\hat{\alpha}$	1.1990	0.8979	0.6209	0.2733	0.1930
		$\hat{\beta}$	4.8713	3.2330	2.0874	0.8746	0.6138

Note nas Tabelas 3.3 e 3.4 que à medida em que  $n$  aumenta as amplitudes dos intervalos diminuem e isso também ocorre quando o nível de confiança  $1 - \delta$  é menor, ou seja, se o nível de confiança diminui a amplitude do intervalo também diminui, uma vez que o quantil da distribuição normal relacionado ao nível de confiança será menor. Isso indica que os intervalos de confiança ficam mais precisos quando o tamanho da amostra aumenta ou quanto menor for o nível de confiança adotado.

**Tabela 3.5:** Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de  $\lambda$ ;  $p = 0.5$ ,  $\alpha = 1.5$  e  $\beta = 3.0$ .

$\lambda$	$1 - \delta$	$\hat{\theta}$	$n$			
			30	50	100	500
0.05	99%	$\hat{\lambda}$	99.96 (0.04; 0.00)	99.96 (0.04; 0.00)	99.98 (0.02; 0.00)	98.13 (0.11; 1.76)
		$\hat{p}$	99.99 (0.00; 0.01)	99.75 (0.11; 0.14)	98.04 (0.92; 1.04)	97.88 (0.98; 1.14)
		$\hat{\alpha}$	98.84 (0.43; 0.73)	98.89 (0.56; 0.55)	99.00 (0.49; 0.51)	99.13 (0.49; 0.38)
		$\hat{\beta}$	98.63 (0.00; 1.37)	98.82 (0.00; 1.18)	99.05 (0.01; 0.94)	98.99 (0.25; 0.76)
		$\hat{\lambda}$	98.81 (1.19; 0.00)	99.39 (0.61; 0.00)	94.79 (0.42; 4.79)	93.47 (1.23; 5.30)
		$\hat{p}$	99.79 (0.08; 0.13)	98.84 (0.59; 0.57)	94.20 (2.88; 2.92)	93.24 (3.29; 3.47)
	95%	$\hat{\alpha}$	94.68 (2.97; 2.35)	94.68 (2.88; 2.44)	94.93 (2.63; 2.44)	95.06 (2.57; 2.37)
		$\hat{\beta}$	96.46 (0.00; 3.54)	96.30 (0.43; 3.27)	95.70 (1.22; 3.08)	95.26 (1.93; 2.81)
		$\hat{\lambda}$	94.58 (5.42; 0.00)	97.48 (2.52; 0.00)	92.07 (3.14; 4.79)	88.64 (2.96; 8.40)
		$\hat{p}$	98.28 (0.86; 0.86)	94.49 (2.62; 2.89)	87.46 (6.13; 6.41)	87.61 (6.03; 6.36)
		$\hat{\alpha}$	89.45 (6.26; 4.29)	89.85 (5.84; 4.31)	90.07 (5.47; 4.46)	89.84 (5.25; 4.91)
		$\hat{\beta}$	93.72 (0.76; 5.52)	92.11 (2.47; 5.42)	91.31 (3.65; 5.04)	90.34 (4.55; 5.11)
	90%	$\hat{\lambda}$	98.72 (0.64; 0.64)	98.48 (0.75; 0.77)	98.73 (0.63; 0.64)	98.95 (0.49; 0.56)
		$\hat{p}$	96.92 (1.52; 1.56)	97.91 (1.09; 1.00)	98.42 (0.73; 0.85)	98.89 (0.50; 0.61)
		$\hat{\alpha}$	98.89 (0.44; 0.67)	99.02 (0.43; 0.55)	98.86 (0.64; 0.50)	99.32 (0.32; 0.36)
		$\hat{\beta}$	98.49 (0.00; 1.51)	98.66 (0.00; 1.34)	98.99 (0.00; 1.01)	99.01 (0.12; 0.87)
		$\hat{\lambda}$	96.22 (1.88; 1.90)	93.87 (3.08; 3.05)	94.27 (2.77; 2.96)	94.45 (2.74; 2.81)
		$\hat{p}$	91.39 (4.35; 4.26)	93.38 (3.32; 3.30)	93.63 (3.23; 3.14)	94.63 (2.57; 2.80)
0.50	99%	$\hat{\alpha}$	94.57 (3.43; 2.00)	94.72 (3.13; 2.15)	94.77 (2.96; 2.27)	95.53 (2.19; 2.28)
		$\hat{\beta}$	96.45 (0.00; 3.55)	96.59 (0.00; 3.41)	96.50 (0.43; 3.07)	95.78 (1.40; 2.82)
	95%	$\hat{\lambda}$	90.46 (4.96; 4.58)	88.48 (5.69; 5.83)	90.98 (4.34; 4.68)	90.15 (4.86; 4.99)
		$\hat{p}$	87.21 (6.36; 6.43)	87.95 (6.22; 5.83)	88.71 (5.64; 5.65)	89.51 (4.97; 5.52)
		$\hat{\alpha}$	88.89 (7.36; 3.75)	89.42 (6.74; 3.84)	89.72 (5.61; 4.67)	90.51 (5.02; 4.47)
		$\hat{\beta}$	94.52 (0.00; 5.48)	94.12 (0.45; 5.43)	91.88 (2.69; 5.43)	90.88 (3.97; 5.15)



**Tabela 3.6:** Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, variando o valor de  $p$ ;  $\lambda = 0.5$ ,  $\alpha = 1.5$  e  $\beta = 3.0$ .

$p$	$1 - \delta$	$\hat{\theta}$	$n$				
			30	50	100	500	
0.25	99%	$\hat{\lambda}$	98.79	98.49	98.73	98.95	
			(0.64; 0.57)	(0.75; 0.76)	(0.63; 0.64)	(0.49; 0.56)	
		$\hat{p}$	96.01	96.38	97.91	98.94	
			(0.33; 3.66)	(0.38; 3.24)	(0.27; 1.82)	(0.33; 0.73)	
		$\hat{\alpha}$	98.94	99.02	98.86	99.32	
			(0.42; 0.64)	(0.43; 0.55)	(0.64; 0.50)	(0.32; 0.36)	
		$\hat{\beta}$	98.53	98.68	98.99	99.01	
			(0.00; 1.47)	(0.00; 1.32)	(0.00; 1.01)	(0.12; 0.87)	
		95%	$\hat{\lambda}$	96.35	93.88	94.27	94.45
				(1.89; 1.76)	(3.08; 3.04)	(2.77; 2.96)	(2.74; 2.81)
			$\hat{p}$	91.64	92.41	94.11	94.97
				(1.83; 6.53)	(1.64; 5.95)	(1.50; 4.39)	(1.87; 3.16)
	$\hat{\alpha}$		94.63	94.73	94.77	95.53	
			(3.45; 1.92)	(3.13; 2.14)	(2.96; 2.27)	(2.19; 2.28)	
	$\hat{\beta}$		96.50	96.61	96.50	95.78	
			(0.00; 3.50)	(0.00; 3.39)	(0.43; 3.07)	(1.40; 2.82)	
	90%	$\hat{\lambda}$	90.59	88.51	90.98	90.15	
			(5.03; 4.38)	(5.69; 5.80)	(4.34; 4.68)	(4.86; 4.99)	
		$\hat{p}$	86.52	87.44	89.40	89.52	
			(4.29; 9.19)	(4.23; 8.33)	(3.58; 7.02)	(4.43; 6.05)	
		$\hat{\alpha}$	88.92	89.44	89.72	90.51	
			(7.42; 3.66)	(6.73; 3.83)	(5.61; 4.67)	(5.02; 4.47)	
		$\hat{\beta}$	94.56	94.14	91.88	90.88	
			(0.00; 5.44)	(0.45; 5.41)	(2.69; 5.43)	(3.97; 5.15)	
0.75	99%	$\hat{\lambda}$	98.78	98.49	98.73	98.95	
			(0.64; 0.58)	(0.75; 0.76)	(0.63; 0.64)	(0.49; 0.56)	
		$\hat{p}$	95.35	96.09	97.83	98.86	
			(4.26; 0.39)	(3.57; 0.34)	(1.88; 0.29)	(0.88; 0.26)	
		$\hat{\alpha}$	98.90	99.02	98.86	99.32	
			(0.44; 0.66)	(0.43; 0.55)	(0.64; 0.50)	(0.32; 0.36)	
		$\hat{\beta}$	98.50	98.66	98.99	99.01	
			(0.00; 1.50)	(0.00; 1.34)	(0.00; 1.01)	(0.12; 0.87)	
		95%	$\hat{\lambda}$	96.40	93.90	94.27	94.45
				(1.90; 1.70)	(3.08; 3.02)	(2.77; 2.96)	(2.74; 2.81)
			$\hat{p}$	91.13	91.54	93.34	95.01
				(6.95; 1.92)	(6.85; 1.61)	(4.87; 1.79)	(3.10; 1.89)
	$\hat{\alpha}$		94.56	94.72	94.77	95.53	
			(3.47; 1.97)	(3.13; 2.15)	(2.96; 2.27)	(2.19; 2.28)	
	$\hat{\beta}$		96.48	96.59	96.50	95.78	
			(0.00; 3.52)	(0.00; 3.41)	(0.43; 3.07)	(1.40; 2.82)	
	90%	$\hat{\lambda}$	90.67	88.51	90.98	90.15	
			(5.04; 4.29)	(5.69; 5.80)	(4.34; 4.68)	(4.86; 4.99)	
		$\hat{p}$	85.66	86.53	88.41	89.69	
			(10.02; 4.32)	(9.49; 3.98)	(7.74; 3.85)	(5.98; 4.33)	
		$\hat{\alpha}$	88.83	89.42	89.72	90.51	
			(7.42; 3.75)	(6.74; 3.84)	(5.61; 4.67)	(5.02; 4.47)	
		$\hat{\beta}$	94.58	94.12	91.88	90.88	
			(0.00; 5.42)	(0.45; 5.43)	(2.69; 5.43)	(3.97; 5.15)	

As taxas de cobertura foram calculadas para verificar se os intervalos de confiança cobrem os verdadeiros valores dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um impostos nas simulações, considerando os níveis de confiança pré-estabelecidos. Já as taxas de não cobertura à esquerda e à direita foram calculadas para verificar se tais não-coberturas são aproximadamente simétricas, como espera-se que sejam. Os resultados estão dispostos nas Tabelas 3.5 e 3.6.

É possível observar, ao analisar as Tabelas 3.5 e 3.6, que as taxas de cobertura convergem para os níveis de confiança adotados e as taxas de não cobertura tendem a apresentar comportamento simétrico à medida em que  $n$  aumenta, o que é esperado uma vez que os intervalos de confiança assintóticos descritos na Seção 3.3 são baseados na distribuição normal.

Diante do exposto nesta seção, é possível perceber o bom desempenho da estimação intervalar considerando a distribuição assintótica dos estimadores de máxima verossimilhança, em que os intervalos apresentam boa taxa de cobertura dos parâmetros, bem como aproximada simetria de não cobertura.

### 3.5.3 Testes de hipóteses

A fim de verificar o desempenho dos testes razão de verossimilhanças ( $RV$ ), Wald ( $W$ ) e escore ( $S$ ), foi testada a hipótese  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$  e foram calculados os tamanhos desses testes. Os valores nulos para  $\lambda$  são  $\lambda_0 = (0.05, 0.10, 0.20, 0.50)$ ; os demais parâmetros são fixos e assumem os seguintes valores:  $p = 0.5$ ,  $\alpha = 1.5$  e  $\beta = 3.0$ . Os resultados das simulações de tamanho para este cenário encontram-se na Tabela 3.7, que apresenta as taxas de rejeição nulas, isto é, calculadas sob  $H_0$ , em que o valor nulo do parâmetro testado é imposto na geração da amostra. Foram utilizados os níveis nominais 1%, 5% e 10% e para definir as regiões de rejeição foram utilizados os quantis 0.99, 0.95 e 0.90 da distribuição qui-quadrado com um grau de liberdade.

É possível notar que as taxas de rejeição nulas convergem para os respectivos níveis nominais à medida em que o tamanho amostral cresce, como esperado. Observe ainda na Tabela 3.7 que ao testar valores maiores para  $\lambda$ , as taxas de rejeição nulas tendem a se aproximar mais rapidamente dos níveis nominais adotados, ou seja, quanto maior for o grau de mistura da distribuição, neste caso  $\lambda_0 = 0.50$ , melhor é o desempenho dos testes. Além disso, na maioria dos cenários, o teste Wald apresenta desempenho inferior aos testes da razão de verossimilhanças e escore, e o teste razão de verossimilhanças apresenta desempenho pouco melhor que o teste escore, considerando boa parte dos cenários. Contudo em diversos casos os dois testes, razão de verossimilhanças e escore, apresentam comportamentos bastante similares. Para valores maiores de  $\lambda_0$ , a saber  $\lambda_0 = 0.50$ , os três testes exibem desempenhos semelhantes, convergindo de maneira similar para os níveis nominais.

Também foram calculadas as taxas de rejeição sob a hipótese nula do teste de  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$ , ou seja, os tamanhos dos testes para este novo cenário. Neste caso,  $\lambda = 0.2$ ,  $\alpha = 1.5$  e  $\beta = 3.0$  são os valores fixados para os demais parâmetros, enquanto

**Tabela 3.7:** Taxas (%) de rejeição nulas dos testes, considerando  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em zero e um.

$\lambda_0$	Nível nominal	Estatística de teste	$n$						
			20	30	50	100	200	500	1000
0.05	1%	$RV$	1.54	1.19	0.61	0.42	1.15	0.86	1.18
		$W$	0.01	0.04	0.04	0.02	2.12	1.87	1.57
		$S$	9.54	5.42	2.52	1.30	1.26	0.96	1.23
	5%	$RV$	9.54	5.42	2.52	3.14	4.39	5.04	4.98
		$W$	1.54	1.19	0.61	5.21	6.51	6.53	5.57
		$S$	39.25	19.53	7.39	3.14	2.82	4.72	4.78
	10%	$RV$	39.25	19.53	7.39	7.63	9.71	9.74	10.85
		$W$	9.54	5.42	2.52	7.93	13.89	11.36	9.28
		$S$	39.25	19.53	7.39	7.63	6.48	7.63	9.25
0.10	1%	$RV$	0.63	1.35	0.33	0.96	1.31	0.86	1.01
		$W$	0.18	0.07	0.10	2.11	1.81	1.53	1.16
		$S$	2.85	1.35	1.06	0.95	1.27	0.78	0.94
	5%	$RV$	2.85	3.98	2.90	3.91	6.07	5.44	4.66
		$W$	0.63	1.35	5.61	6.03	7.31	5.72	4.85
		$S$	9.73	3.98	2.90	5.80	4.33	4.61	5.26
	10%	$RV$	9.73	3.98	7.00	8.96	10.03	10.10	10.19
		$W$	2.85	3.98	7.45	12.59	12.04	10.83	10.51
		$S$	9.73	11.89	7.00	12.36	12.30	8.67	10.19
0.20	1%	$RV$	0.29	0.28	1.07	1.29	1.29	0.96	1.07
		$W$	0.29	2.15	1.77	1.44	1.74	1.09	1.07
		$S$	1.26	1.03	0.77	0.97	1.11	0.79	1.05
	5%	$RV$	4.31	4.78	4.61	4.37	5.58	4.84	4.75
		$W$	1.26	2.90	5.57	6.67	6.58	5.14	5.38
		$S$	4.31	2.91	4.61	5.75	4.46	4.84	5.07
	10%	$RV$	4.31	8.66	10.14	10.04	9.75	10.60	9.61
		$W$	4.31	11.02	12.71	11.38	10.05	10.69	10.51
		$S$	11.17	8.66	10.14	10.04	9.75	10.60	10.41
0.50	1%	$RV$	1.00	1.28	0.58	1.27	0.83	1.05	1.06
		$W$	3.52	1.28	1.52	1.27	1.30	1.05	1.06
		$S$	1.00	0.37	0.58	1.27	0.83	1.05	1.06
	5%	$RV$	3.52	3.78	6.13	5.73	5.60	5.55	5.17
		$W$	3.52	3.78	6.13	5.73	5.60	5.55	5.17
		$S$	3.52	3.78	6.13	5.73	5.60	5.55	5.17
	10%	$RV$	11.06	9.54	11.52	9.02	10.22	9.85	8.95
		$W$	11.06	9.54	11.52	9.02	10.22	9.85	10.12
		$S$	11.06	9.54	11.52	9.02	10.22	9.85	8.95

os valores nulos de  $p$  são  $p_0 = (0.05, 0.10, 0.50, 0.75)$ . Os resultados das simulações de tamanho para este segundo cenário referente aos testes de hipóteses encontram-se na Tabela 3.8.

De acordo com os resultados apresentados na Tabela 3.8 nota-se que as taxas de rejeição nulas dos testes da razão de verossimilhanças, Wald e escore convergem para os respectivos níveis nominais (1%, 5% e 10%) à medida em que  $n$  aumenta. Novamente o teste Wald apresenta comportamento inferior aos demais testes, equiparando-se com os testes da razão de verossimilhanças e escore, apenas quando  $n$  é grande ou, em alguns casos, quando o valor hipotetizado de  $p$  é grande, por exemplo  $p = 0.75$ . Quando o valor hipotetizado  $p_0$  é pequeno, por exemplo  $p_0 = 0.05$ , as taxas de rejeição nulas demoram

**Tabela 3.8:** Taxas (%) de rejeição nulas dos testes, considerando  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$ , para a distribuição Kumaraswamy inflacionada em zero e um.

$p_0$	Nível nominal	Estatística de teste	$n$						
			20	30	50	100	200	500	1000
0.05	1%	$RV$	2.98	2.02	0.94	0.68	0.56	0.51	1.10
		$W$	0.09	0.05	0.01	0.02	0.06	3.56	3.07
		$S$	15.73	11.18	6.25	2.80	1.71	1.03	1.21
	5%	$RV$	8.15	8.00	5.06	3.49	2.46	5.02	5.59
		$W$	1.16	0.74	0.35	0.34	0.44	9.23	7.78
		$S$	31.38	17.68	14.27	7.20	4.48	3.92	4.91
	10%	$RV$	16.01	13.56	11.44	7.12	5.01	9.66	10.44
		$W$	3.00	2.24	1.67	1.23	1.79	13.48	12.40
		$S$	52.95	26.63	19.49	11.33	7.27	9.24	9.84
0.10	1%	$RV$	0.91	1.04	0.66	0.56	0.59	1.26	1.15
		$W$	0.33	0.20	0.08	0.08	4.94	2.82	1.93
		$S$	7.13	4.99	2.73	1.35	1.08	1.09	1.09
	5%	$RV$	7.13	5.00	3.20	2.37	3.51	5.20	5.15
		$W$	2.60	1.36	0.67	0.79	8.35	7.05	6.21
		$S$	11.14	9.15	6.32	4.65	3.48	4.90	4.99
	10%	$RV$	11.14	9.15	6.32	5.11	9.38	10.11	10.55
		$W$	3.55	2.76	2.02	3.01	12.40	11.81	11.08
		$S$	21.97	14.77	10.29	7.49	8.14	9.69	10.23
0.50	1%	$RV$	0.00	0.10	0.88	1.20	0.96	0.97	1.06
		$W$	0.96	2.82	4.31	2.50	1.60	1.27	1.17
		$S$	0.00	0.02	0.38	0.85	0.93	0.93	1.02
	5%	$RV$	1.00	3.19	6.17	5.23	4.98	5.31	5.19
		$W$	3.66	7.63	10.57	6.90	5.77	5.79	5.48
		$S$	0.23	1.18	3.45	5.07	4.64	5.29	5.18
	10%	$RV$	3.66	7.63	10.63	10.12	9.84	10.48	10.10
		$W$	10.25	14.00	14.13	11.94	10.69	10.85	10.32
		$S$	1.16	4.39	9.49	9.76	9.53	10.33	10.10
0.75	1%	$RV$	0.49	0.76	0.74	0.72	1.19	1.19	1.13
		$W$	0.49	0.76	1.49	4.14	2.67	1.62	1.37
		$S$	0.49	0.78	0.98	0.94	0.80	1.10	1.10
	5%	$RV$	3.00	2.80	2.39	4.84	5.43	5.55	5.17
		$W$	3.00	3.03	5.30	8.61	7.08	6.30	5.57
		$S$	3.52	4.29	4.08	4.09	4.80	5.29	5.12
	10%	$RV$	5.83	6.65	6.30	10.06	10.33	10.51	10.56
		$W$	3.54	4.90	10.00	14.00	12.16	11.11	10.72
		$S$	10.35	8.03	5.96	9.20	10.07	10.63	10.40

mais para convergirem aos respectivos níveis nominais. Quando a probabilidade de uns na sub-amostra discreta aumenta, então as taxas se aproximam dos níveis nominais mais rapidamente.

Ressalta-se que, nas simulações de tamanho realizadas foram armazenados os quantis de 90%, 95% e 99% das estatísticas de teste a fim de utilizá-los como valores críticos (exatos) nas simulações de poder, visando comparar testes de mesmo tamanho.

Para as simulações de poder foram calculadas as taxas de rejeição sob a hipótese alternativa, ou seja, realizando a geração dos dados sob  $H_1$ . Os resultados obtidos estão dispostos nas Tabelas 3.9 e 3.10. Na Tabela 3.9 encontram-se as taxas de rejeição não nulas quando a hipótese testada é  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$  e na Tabela 3.10 encontram-se as

taxas de rejeição não nulas para a hipótese  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$ . Os valores de  $\alpha$  e  $\beta$  são fixos e iguais a 1.5 e 3.0, respectivamente, além de  $p = 0.5$  na primeira simulação e  $\lambda = 0.2$  na segunda simulação.

**Tabela 3.9:** Taxas (%) de rejeição não nulas dos testes, considerando  $H_0 : \lambda = \lambda_0$  vs.  $H_1 : \lambda \neq \lambda_0$ , para a distribuição Kumaraswamy inflacionada em zero e um.

$\lambda_0$	Nível nominal	Estatística de teste	$n$						
			20	30	50	100	200	500	1000
0.05	1%	$RV$	7.28	10.91	16.45	38.00	61.96	96.68	99.98
		$W$	7.29	10.69	17.21	29.85	20.21	90.17	99.89
		$S$	2.85	3.98	14.35	29.85	63.63	97.30	99.98
	5%	$RV$	17.88	25.88	35.26	61.40	81.60	98.69	100.00
		$W$	18.67	25.79	35.15	29.94	53.74	98.13	100.00
		$S$	9.73	11.89	26.32	55.82	79.71	99.21	100.00
	10%	$RV$	30.00	35.81	48.95	68.93	86.31	99.51	100.00
		$W$	29.86	36.22	48.53	61.84	79.71	99.21	100.00
		$S$	29.39	27.33	44.08	68.88	86.00	99.50	100.00
0.10	1%	$RV$	5.05	6.82	12.04	15.95	37.18	81.07	98.72
		$W$	3.79	6.81	6.78	1.96	14.39	70.20	97.86
		$S$	3.79	3.65	6.71	23.57	37.18	82.38	99.05
	5%	$RV$	15.02	20.66	28.04	38.30	53.73	91.88	99.70
		$W$	15.54	20.57	8.31	15.25	45.60	88.49	99.62
		$S$	10.79	18.32	21.41	43.37	61.18	92.81	99.70
	10%	$RV$	26.11	31.10	35.08	46.71	68.65	95.97	99.93
		$W$	26.31	30.90	30.35	33.01	61.23	94.56	99.87
		$S$	25.86	18.32	34.69	43.40	68.64	95.97	99.93
0.20	1%	$RV$	0.02	0.00	16.47	57.94	91.06	100.00	100.00
		$W$	0.02	9.63	26.51	63.23	93.42	100.00	100.00
		$S$	0.00	0.00	4.55	44.18	85.71	100.00	100.00
	5%	$RV$	0.22	18.34	37.93	79.74	97.71	100.00	100.00
		$W$	2.64	19.02	55.92	84.61	98.46	100.00	100.00
		$S$	0.18	0.00	34.83	70.15	97.30	100.00	100.00
	10%	$RV$	0.49	24.53	55.93	87.42	99.21	100.00	100.00
		$W$	22.82	42.91	64.56	90.41	99.51	100.00	100.00
		$S$	0.18	18.41	34.83	79.74	99.09	100.00	100.00
0.50	1%	$RV$	4.35	6.94	11.76	27.77	60.73	97.44	100.00
		$W$	4.35	7.99	12.36	28.72	60.96	97.51	100.00
		$S$	4.35	4.10	9.32	24.75	58.46	97.02	100.00
	5%	$RV$	11.38	20.28	29.49	52.55	81.13	99.51	100.00
		$W$	11.36	20.35	30.57	52.99	81.12	99.48	100.00
		$S$	11.17	17.83	23.59	47.26	78.89	99.46	100.00
	10%	$RV$	24.23	30.61	42.63	64.78	88.83	99.82	100.00
		$W$	24.36	30.62	42.48	64.44	88.78	99.82	100.00
		$S$	11.17	29.93	34.39	62.92	86.20	99.82	100.00

Na primeira simulação de poder, Tabela 3.9, a geração dos dados foi feita usando os seguintes valores de  $\lambda$ : (0.10, 0.15, 0.10 e 0.40) contra  $\lambda_0 = (0.05, 0.10, 0.20 \text{ e } 0.50)$ , respectivamente, e como valores críticos foram utilizados os quantis das estatísticas de teste calculadas na simulação de tamanho, cujos resultados se encontram na Tabela 3.7.

Como observado a partir dos resultados apresentados nas Tabelas 3.9 e 3.10, em ambos os cenários nota-se que as taxas de rejeição não nulas convergem para 100% quando o tamanho amostral aumenta, o que era esperado, indicando bom desempenho dos testes de

hipóteses realizados.

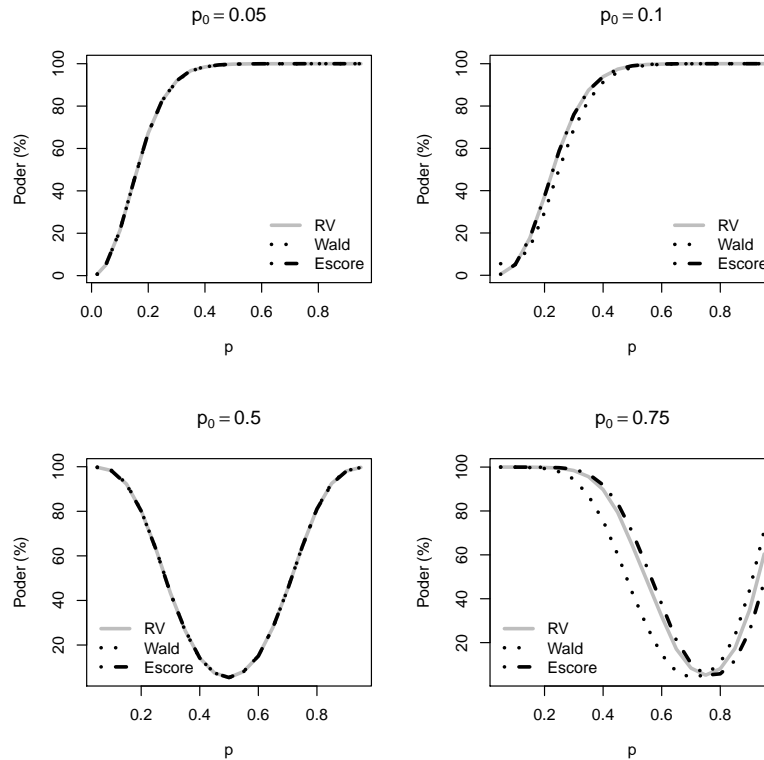
**Tabela 3.10:** Taxas (%) de rejeição não nulas dos testes, considerando  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$ , para a distribuição Kumaraswamy inflacionada em zero e um.

$p_0$	Nível nominal	Estatística de teste	$n$						
			20	30	50	100	200	500	1000
0.05	1%	$RV$	3.03	3.60	4.90	8.72	15.87	34.66	58.13
		$W$	3.22	3.53	5.19	8.69	15.47	0.05	7.90
		$S$	3.19	3.52	5.16	8.61	15.80	39.34	65.76
	5%	$RV$	10.98	12.26	14.65	21.32	35.82	49.81	77.21
		$W$	10.96	11.96	14.63	21.32	30.23	18.10	55.33
		$S$	11.03	11.94	14.68	21.46	35.75	59.62	82.12
	10%	$RV$	18.74	20.96	23.92	32.41	47.92	61.41	85.43
		$W$	18.73	20.86	24.00	32.35	37.99	36.52	75.03
		$S$	17.88	20.49	23.89	32.45	48.12	67.71	87.68
0.10	1%	$RV$	4.35	5.24	9.41	17.35	33.30	60.78	92.01
		$W$	4.36	5.30	9.35	15.13	0.89	14.03	75.59
		$S$	4.26	5.23	9.21	17.87	33.71	69.20	94.22
	5%	$RV$	13.08	16.90	22.95	37.65	49.49	80.38	97.73
		$W$	13.36	16.90	23.20	29.98	11.12	62.59	95.18
		$S$	13.17	16.90	23.11	37.53	57.56	84.60	98.31
	10%	$RV$	22.76	26.43	33.87	49.45	57.35	87.58	98.94
		$W$	22.53	26.71	33.92	38.48	37.41	80.11	98.23
		$S$	22.35	26.75	33.73	50.01	65.00	89.64	99.17
0.50	1%	$RV$	1.57	1.60	2.34	4.51	9.48	29.22	58.80
		$W$	1.57	1.63	2.01	4.41	9.35	29.23	58.80
		$S$	1.57	1.65	2.32	4.63	9.48	29.22	58.72
	5%	$RV$	6.32	6.99	9.20	14.40	24.31	50.58	80.77
		$W$	6.66	7.26	8.96	14.28	24.31	50.52	80.75
		$S$	6.41	6.98	9.20	14.44	24.31	50.56	80.75
	10%	$RV$	11.97	13.29	15.91	22.59	35.12	63.02	88.38
		$W$	11.96	12.69	15.95	22.42	35.26	63.01	88.39
		$S$	11.80	13.19	15.93	22.46	35.24	63.02	88.32
0.75	1%	$RV$	3.12	4.07	4.77	7.16	11.19	34.86	68.93
		$W$	3.06	3.69	3.12	0.53	2.18	20.91	59.36
		$S$	2.20	4.08	4.63	8.89	16.43	39.72	71.64
	5%	$RV$	11.32	13.28	15.36	17.23	27.42	57.87	86.90
		$W$	11.32	12.09	8.84	6.83	17.73	50.45	84.39
		$S$	11.26	13.31	16.87	21.15	31.99	60.98	87.83
	10%	$RV$	19.28	21.11	21.97	24.95	39.06	69.00	91.84
		$W$	19.06	20.27	16.34	16.74	31.57	64.86	90.63
		$S$	18.88	21.67	24.96	28.79	42.58	71.03	92.48

Quando o parâmetro de mistura  $\lambda$  é testado, Tabela 3.9, constata-se que para valores mais baixos de  $\lambda_0$ , as taxas de rejeição não nulas dos testes aumentam mais lentamente, quando comparadas com os poderes registrados ao se utilizarem valores mais altos de  $\lambda_0$ . Ainda nesse cenário, considerando um tamanho amostral pequeno, por exemplo  $n = 20$ , o teste Wald apresenta desempenho superior aos testes da razão de verossimilhanças e escore, apresentando taxas de rejeição não nulas mais altas que os outros dois testes, o que se repete para outros tamanhos amostrais. Contudo, de maneira geral, o teste da razão de verossimilhanças apresenta desempenho melhor do que os dos demais testes utilizados quando tem-se valores mais baixos de  $\lambda_0$ , independente do tamanho amostral. O teste

escore mostra desempenho inferior aos outros dois testes na maioria dos cenários.

Ao se testar  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$ , também foram calculadas as taxas de rejeição sob  $H_1$  e os resultados das simulações para este novo cenário estão dispostos na Tabela 3.10. Nessa simulação foram considerados como valores nulos  $p_0 = (0.05, 0.10, 0.50$  e  $0.75)$  mas a geração dos dados foi feita utilizando os seguintes valores para  $p$ :  $(0.10, 0.20, 0.40$  e  $0.65)$ , respectivamente. Os valores críticos usados na definição da região de rejeição foram extraídos das simulações de tamanho apresentadas na Tabela 3.8, calculando-se os quantis desejados das estatísticas de testes utilizadas. Em primeira instância é notório que as taxas de rejeição não nulas se aproximam mais lentamente de 100% quando a hipótese a ser testada refere-se ao parâmetro  $p$ , isto quando comparada à hipótese que impõe restrições sob o parâmetro  $\lambda$  (Tabela 3.9). Ainda destaca-se para  $p_0 = 0.50$  que os testes apresentam desempenhos bastante similares, todavia, considerando tamanhos amostrais altos, o teste escore se destaca entre os demais, apresentando taxas de rejeição não nulas mais altas. Já o teste Wald exibe desempenho inferior aos outros dois testes em boa parte dos cenários. Em alguns casos, para pequenos tamanhos de amostra, o teste razão de verossimilhanças é o teste que apresenta o melhor desempenho.



**Figura 3.2:** Poderes dos testes, distribuição KIZU.

Ademais, curvas de poder foram construídas considerando  $n = 100$ , 5000 réplicas de Monte Carlo,  $\alpha = 1.5$ ,  $\beta = 3.0$  e nível de significância de 5%. A fim de comparar os poderes dos três testes, quando a hipótese testada é  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$ , as curvas de poder estão expostas na Figura 3.2. Nota-se que para um valor de  $p$  mais alto, por exemplo  $p_0 = 0.75$ , os comportamentos dos testes são mais diferentes, tendo o teste Wald desempenho inferior aos demais. Quando a probabilidade de uns é igual à probabilidade de zeros na subamostra de valores discretos, ou seja  $p_0 = 0.50$ , os testes têm desempenhos bastante próximos. Adicionalmente, comparando os quatro cenários exibidos na Figura 3.2, percebe-se que para valores menores de  $p$ , as taxas de rejeição dos testes convergem mais rapidamente para 100% e são mais sensíveis a pequenas distâncias do valor hipotetizado. Não obstante, em qualquer cenário são evidenciados os bons desempenhos dos três testes utilizados, uma vez que estes têm taxas de rejeição que tendem para 100% à medida em que o valor alternativo se distancia do valor nulo, tal como tende ao nível nominal adotado quando os valores sob  $H_1$  se aproximam do valor nulo.

### 3.6 Aplicação

Nesta seção será apresentada uma aplicação a dados reais restritos ao intervalo  $[0, 1]$  utilizando a distribuição Kumaraswamy inflacionada em zero e um, KIZU. O conjunto de dados usado nessa aplicação possui 109 observações e a variável de interesse é a proporção de acumulação de água em reservatórios do Estado de Pernambuco. As observações foram feitas entre 21/10/2009 e 17/11/2017 e foram retiradas do banco de dados de monitoramento hidrológico da Agência Pernambucana de Águas e Clima (APAC), que se encontra disponível no sítio <http://www.apac.pe.gov.br/monitoramento/>.

A nível de comparação também foi ajustada a distribuição beta inflacionada em zero e um (BIZU) proposta por Ospina e Ferrari (2010). Uma variável aleatória segue distribuição BIZU com parâmetros  $\lambda, p, \mu$  e  $\phi$  se sua função de densidade é dada por

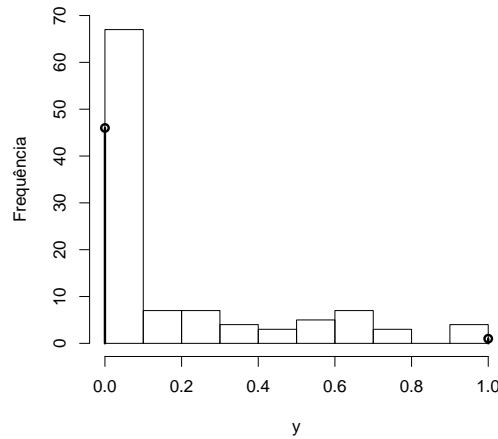
$$\text{bizu}(y; \lambda, p, \mu, \phi) = \begin{cases} \lambda p, & \text{se } y = 1, \\ \lambda(1 - p), & \text{se } y = 0, \\ (1 - \lambda)f(y; \mu, \phi), & \text{se } y \in (0, 1), \end{cases}$$

em que  $f(y; \mu, \phi)$  é a função densidade de probabilidade da distribuição beta com parâmetros  $\mu$  (média) e  $\phi$  (precisão).

Os parâmetros das duas distribuições foram estimados pelo método de máxima verossimilhança utilizando o ambiente computacional estatístico R. Foi usado o método de otimização L-BFGS-B com primeiras derivadas analíticas. Para auxiliar na implementação da função de distribuição KIZU foi utilizada a função `pkumar` do pacote `VGAM` (Yee, 2008); ademais, foi utilizado o pacote `gamlss.dist`, desenvolvido por Mikis Stasinopoulos, Bob Rigby com contribuições de Calliope Akantziliotou e Raydonal Ospina (Stasinopoulos et al., 2012), para a implementação da distribuição BIZU.

O histograma da distribuição da proporção de acumulação de água em reservatórios no Estado de Pernambuco está apresentado na Figura 3.3. É possível verificar que a





**Figura 3.3:** Histograma da proporção de acumulação de água em reservatórios de Pernambuco.

distribuição dos dados apresenta assimetria à direita. Esse comportamento pode ser confirmado ao se analisar a Tabela 3.11 uma vez que a mediana é menor que a média. As duas linhas verticais com um círculo em cima no histograma da Figura 3.3 representam as frequências de zeros e uns na amostra. Observe que a frequência de zeros é maior que a frequência de uns; mais precisamente a amostra tem 34 zeros e 13 valores um.

Na Tabela 3.11 estão apresentadas algumas medidas resumo da proporção de acumulação de água em reservatórios de Pernambuco. Nota-se que o valor mínimo da variável de interesse é zero, o valor máximo é um e 75% das observações não excede 0.8820.

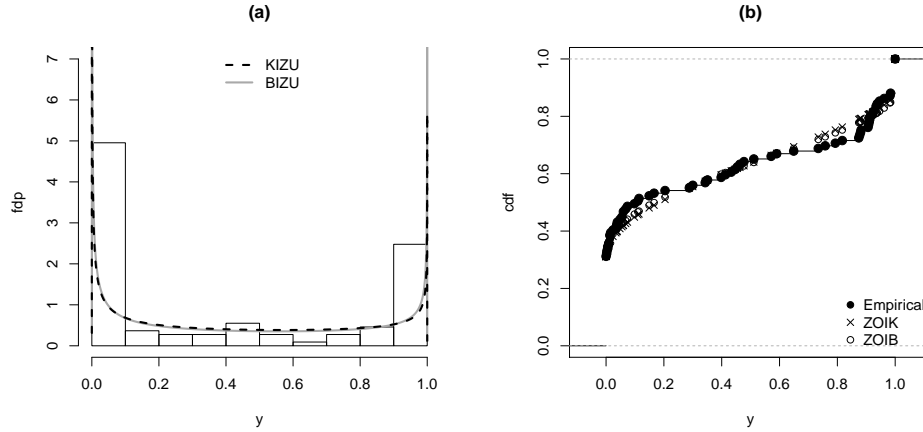
**Tabela 3.11:** Medidas resumo da proporção de acumulação de água em reservatórios de Pernambuco.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
0.0000	0.0000	0.1100	0.3743	0.8820	1.0000

As estimativas de máxima verossimilhança dos parâmetros que indexam a distribuição KIZU com seus respectivos erros-padrão (entre parênteses) são  $\hat{\lambda} = 0.4312$  (0.0474),  $\hat{p} = 0.2766$  (0.0652),  $\hat{\alpha} = 0.4570$  (0.0856) e  $\hat{\beta} = 0.6481$  (0.0998). Para a distribuição BIZU as estimativas são  $\hat{\lambda} = 0.4312$  (0.0474),  $\hat{p} = 0.2766$  (0.0652),  $\hat{\mu} = 0.4409$  (0.0460) e  $\hat{\phi} = 0.9990$  (0.1617).

Na Figura 3.4 estão apresentados (a) o histograma dos dados junto com as densidades KIZU e BIZU estimadas, bem como (b) as funções distribuição acumuladas estimadas KIZU e BIZU e a função distribuição acumulada empírica. Note que a função distribuição acumulada KIZU está mais próxima da função empírica do que a função de distribuição

BIZU.



**Figura 3.4:** (a) Histograma e densidades estimadas e (b) funções distribuição acumuladas para a proporção de acumulação de água em reservatórios de Pernambuco.

Adicionalmente foi calculada a estatística Kolmogorov-Smirnov (KS) considerando apenas a parte contínua dos dados. Essa estatística compara a função distribuição empírica dos dados com a função distribuição acumulada hipotetizada e quanto maior for o seu valor mais indícios existem de que os dados não são provenientes da distribuição em avaliação. A estatística KS é 0.1391 ( $p$ -valor = 0.1787) para a distribuição KIZU e 0.2422 ( $p$ -valor = 0.0014) para a distribuição BIZU, indicando que há mais evidências de que os dados são provenientes da distribuição KIZU.

### 3.7 Conclusões

Em muitas aplicações empíricas os dados assumem valores no intervalo  $[0, 1]$  e, assim, não podem ser modelados por distribuições que possuem suporte em  $(0, 1)$ , como por exemplo as distribuições beta e Kumaraswamy. Ospina e Ferrari (2010) propuseram uma distribuição inflacionada em zero e um tendo a distribuição beta como base. Dadas as vantagens da distribuição Kumaraswamy, neste capítulo foi proposta uma distribuição inflacionada em zero e um a partir da mistura da distribuição Bernoulli com a distribuição Kumaraswamy, que tem uma função densidade de probabilidade bastante flexível, apresenta uma função quantílica simples, que não depende de qualquer função especial, o que possibilita gerar facilmente ocorrências de uma variável aleatória Kumaraswamy inflacionada em zero e um. Essa nova distribuição permite modelar dados no intervalo  $[0, 1]$ .

Estimação pontual, através do método de máxima verossimilhança, foi discutida para a nova distribuição, bem como foram apresentados intervalos de confiança assintóticos e

também testes de hipóteses. Os desempenhos dos estimadores de máxima verossimilhança dos parâmetros que indexam a distribuição Kumaraswamy inflacionada em zero e um foram avaliados através de simulações de Monte Carlo. Também foram realizadas simulações para verificar os desempenhos dos testes da razão de verossimilhanças, Wald e escore, comparando seus tamanhos e poderes. A partir dos resultados encontrados, pode-se concluir que os estimadores de máxima verossimilhança dos parâmetros da distribuição Kumaraswamy inflacionada em zero e um, bem como os testes considerados, apresentam bons desempenhos em amostras finitas.

Também foi realizada uma aplicação a dados reais, através da qual pode-se observar que a distribuição Kumaraswamy inflacionada em zero e um ajusta bem dados no intervalo  $[0, 1]$ . Portanto, fica evidenciada a utilidade da distribuição Kumaraswamy inflacionada em zero e um para modelar dados que assumem valores em  $[0, 1]$ .

## 4 MODELOS DE REGRESSÃO KUMARASWAMY INFLACIONADOS

### 4.1 Introdução

Modelos de regressão são frequentemente usados para analisar dados que são relacionados com outras variáveis. Quando se tem uma relação linear entre a variável resposta (dependente) e a variável regressora (independente), o modelo de regressão linear clássico é comumente utilizado. Contudo, esse modelo além de supor linearidade impõe outros pressupostos, por exemplo, considera indiretamente que a variável resposta é normalmente distribuída. Por conseguinte, o modelo de regressão linear não é apropriado em situações em que a variável resposta assume valores no intervalo  $(0, 1)$ , podendo até fornecer valores ajustados para a variável dependente fora dos limites desse intervalo. Uma solução é usar uma transformação na variável resposta para que ela assuma valores na reta real. Contudo, essa abordagem possui desvantagens, como por exemplo a dificuldade na interpretação dos parâmetros do modelo com relação à variável resposta original.

Ferrari e Cribari-Neto (2004) propuseram o modelo de regressão beta, que é adequado para casos em que a variável resposta assume valores no intervalo  $(0, 1)$ . Esse modelo supõe que tal variável segue distribuição beta.

O modelo de regressão beta não é adequado para modelar variáveis que assumem valores em  $[0, 1)$ ,  $(0, 1]$  ou  $[0, 1]$ . Ospina e Ferrari (2012) propuseram modelos de regressão que são apropriados para modelar variáveis nos intervalos  $[0, 1)$  e  $(0, 1]$ , e Ospina (2008) também propõe, em sua tese de doutorado, o modelo de regressão para variáveis em  $[0, 1]$ .

Como observado nos Capítulos 2 e 3, a distribuição Kumaraswamy proposta por Kumaraswamy (1976), que assim como a distribuição beta, pode ser usada para modelar variáveis no intervalo  $(0, 1)$ , tem características atraentes e por isso será utilizada neste trabalho para criação de novas distribuições que atribuem probabilidades aos limites do intervalo  $(0, 1)$ .

A distribuição Kumaraswamy não possui formas fechadas para seus momentos, mas possui uma função quantílica simples, o que facilita a modelagem de sua mediana, ou de qualquer outro quantil. A simplicidade da função quantílica da distribuição Kumaraswamy se dá pois esta não depende de funções especiais, como é o caso da distribuição beta. Esse fato também facilita a geração de ocorrências de uma variável aleatória que segue distribuição Kumaraswamy, uma vez que é possível utilizar o método da inversão. Adicionalmente, a distribuição Kumaraswamy também é bastante flexível, sua densidade pode assumir diversos formatos a depender dos valores dos parâmetros que a indexam. Isso faz com que a distribuição Kumaraswamy seja bastante versátil, podendo ser ajustada a diversos conjuntos de dados que se encontram no intervalo  $(0, 1)$ .

A parametrização padrão da distribuição Kumaraswamy em termos dos parâmetros de forma fornece expressões para média e variância que dependem de integrais que não possuem forma fechada, o que dificulta seu uso no contexto de modelagem. Entretanto, Mitnik e Baek (2013) discutem duas reparametrizações da distribuição Kumaraswamy que facilitam seu uso em modelos de regressão permitindo modelar tanto o parâmetro de locação (mediana) quanto a dispersão. Neste capítulo será introduzido o modelo de regressão Kumaraswamy para o caso em que a variável resposta assume valor zero e/ou um, com base no modelo apresentado por Mitnik e Baek (2013).

O presente capítulo está dividido em: descrição do modelo de regressão Kumaraswamy, na Seção 4.2; proposta do modelo Kumaraswamy inflacionado em zero ou um na Seção 4.3, bem como apresentação das respectivas inferências consideradas; na Seção 4.4 são apresentados resultados de simulação de Monte Carlo para estimação dos parâmetros do modelo Kumaraswamy inflacionado em zero ou em um; o modelo Kumaraswamy inflacionado em zero e um é proposto e discutido na Seção 4.5, bem como apresentamos a estimação por máxima verossimilhança; na Seção 4.6 apresentamos os resultados numéricos para o modelo Kumaraswamy inflacionado em zero e um; na Seção 4.7 discutimos critérios para seleção de modelo, bem como resíduos e técnicas para verificar a bondade dos ajustes; aplicações a dados reais, dos modelos inflacionados propostos, são apresentadas na Seção 4.8; e, por fim, as conclusões deste capítulo são apresentadas na Seção 4.9.

## 4.2 Modelo de regressão Kumaraswamy

A distribuição Kumaraswamy é uma distribuição biparamétrica de probabilidade contínua definida para variáveis aleatórias com suporte no intervalo  $(0, 1)$ . É uma distribuição similar à distribuição beta e também é bastante flexível, pois sua função densidade de probabilidade pode assumir diversas formas, a depender dos valores dos parâmetros que a indexam. As funções densidade e de distribuição acumulada considerando a parametrização padrão da distribuição Kumaraswamy podem ser vistas na Seção 2.2. Seja  $Y \sim \text{Kum}(\alpha, \beta)$ . A função quantílica Kumaraswamy é dada por

$$y = G^{-1}(u) = [1 - (1 - u)^{1/\beta}]^{1/\alpha}, \quad (4.1)$$

em que  $u \in (0, 1)$ , sendo  $\alpha$  e  $\beta$  os parâmetros de forma da distribuição Kumaraswamy.

De (4.1) é possível obter uma expressão para mediana da distribuição, que é dada por

$$\omega = \text{md}(Y) = (1 - 0.5^{1/\beta})^{1/\alpha}. \quad (4.2)$$

A expressão (4.2) fornece a base para duas reparametrizações introduzidas por Mitnik e Baek (2013).

Segundo Mitnik e Baek (2013), a partir de (4.2) é possível expressar os parâmetros de forma  $\alpha$  e  $\beta$  como

$$\alpha = \frac{\ln(1 - 0.5^{d_\beta})}{\ln(\omega)} \quad (4.3)$$

e

$$\beta = \frac{\ln(0.5)}{\ln(1 - \omega^{d_\alpha^{-1}})}, \quad (4.4)$$

em que  $d_\alpha = \alpha^{-1}$  e  $d_\beta = \beta^{-1}$ . Substituindo (4.3) e (4.4) nas funções densidade de probabilidade, distribuição acumulada e quantílica da Kumaraswamy, os autores obtiveram duas possíveis reparametrizações da distribuição Kumaraswamy. Nesse capítulo será considerada apenas a primeira reparametrização, uma vez que esta será usada na construção do modelo de regressão. Para mais detalhes sobre essas reparametrizações, ver Mitnik e Baek (2013).

A primeira reparametrização é obtida substituindo  $\beta$  pela expressão (4.4). As novas funções densidade de probabilidade, distribuição acumulada e quantílica reparametrizadas são dadas, respectivamente, por

$$\begin{aligned} g(y; \omega, d_\alpha) &= \frac{\ln(0.5)}{d_\alpha \ln(1 - \omega^{d_\alpha^{-1}})} y^{d_\alpha^{-1}-1} \left(1 - y^{d_\alpha^{-1}}\right)^{\frac{\ln(0.5)}{\ln(1 - \omega^{d_\alpha^{-1}})} - 1}, \\ G(y; \omega, d_\alpha) &= 1 - \left(1 - y^{d_\alpha^{-1}}\right)^{\frac{\ln(0.5)}{\ln(1 - \omega^{d_\alpha^{-1}})}} e \\ G^{-1}(y; \omega, d_\alpha) &= \left[1 - (1 - u)^{\frac{\ln(1 - \omega^{d_\alpha^{-1}})}{\ln(0.5)}}\right]^{d_\alpha}. \end{aligned} \quad (4.5)$$

Mitnik e Baek (2013) mostram que  $d_\alpha$  e  $d_\beta$  são parâmetros de dispersão.

As reparametrizações introduzidas por Mitnik e Baek (2013) possibilitam o uso da distribuição Kumaraswamy para desenvolver um modelo de regressão incluindo dois submodelos, um para a mediana e outro para a dispersão. Mitnik e Baek (2013) sugerem o uso da primeira reparametrização por ser mais vantajosa em termos de interpretabilidade.

Considerando  $Y \sim \text{Kum}(\alpha, \beta)$  como variável dependente e usando a primeira reparametrização expressa em (4.5), Mitnik e Baek (2013) apresentam um modelo análogo aos modelos discutidos por McCullagh e Nelder (1989), em que

$$\omega_i = \left(1 + e^{-x_{i1}\gamma_1}\right)^{-1}, \quad (4.6)$$

$$\tau_i = e^{q_{i1}\varsigma_1}, \quad (4.7)$$

$i$  indexando as observações,  $0 < \omega < 1$  sendo a mediana e  $0 < \tau \equiv d_\alpha < \infty$  sendo o parâmetro de dispersão. Aqui,  $x_1$  é um vetor  $n \times r$  e  $q_1$  é um vetor  $n \times u$  de variáveis regressoras (independentes),  $\gamma_1$  e  $\varsigma_1$  são, respectivamente, os vetores  $r \times 1$  e  $u \times 1$  dos parâmetros do modelo a serem estimados e foram usadas as funções de ligação *logit* e logarítmica, respectivamente, para os submodelos de locação ( $\omega_i$ ) e dispersão ( $\tau_i$ ).

É possível mostrar que

$$Y_i = \left[1 - (1 - u_i)^{\frac{\ln\left(1 - \omega_i^{-1}\right)}{\ln(0.5)}}\right]^{\tau_i},$$

em que  $u_i \sim \mathcal{U}(0, 1)$ .

A função de log-verossimilhança para o modelo dado em (4.6) e (4.7), considerando uma amostra independente  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  de tamanho  $n$  da variável aleatória  $Y$ , é dada por

$$\begin{aligned} \ell(\gamma_1, \varsigma_1; \mathbf{y}) &= \sum_{i=1}^n \ln \left( \frac{1}{\tau_i} \right) + \sum_{i=1}^n \ln \left[ \frac{\ln(0.5)}{\ln \left( 1 - \omega_i^{1/\tau_i} \right)} \right] + \sum_{i=1}^n \left( \frac{1}{\tau_i} - 1 \right) \ln(y_i) \\ &+ \sum_{i=1}^n \left[ \frac{\ln(0.5)}{\ln \left( 1 - \omega_i^{1/\tau_i} \right)} - 1 \right] \ln \left( 1 - y_i^{1/\tau_i} \right). \end{aligned}$$

Esse modelo também é discutido por Oliveira e Bayer (2017) e, segundo os autores, a estimação dos parâmetros pode ser feita através da maximização da função de log-verossimilhança  $\ell(\gamma_1, \varsigma_1; \mathbf{y})$ . Contudo, os estimadores de máxima verossimilhança dos parâmetros do modelo não possuem forma fechada. Então, as correspondentes estimativas podem ser obtidas maximizando numericamente  $\ell(\gamma_1, \varsigma_1; \mathbf{y})$ .

### 4.3 Modelo de regressão Kumaraswamy inflacionado em zero ou um

Seja  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  uma amostra independente de tamanho  $n$  da distribuição Kumaraswamy inflacionada em  $c$ , com densidade reparametrizada dada por

$$\text{ki}_c(y; \lambda, \omega, \tau) = \begin{cases} \lambda, & \text{se } y = c, \\ (1 - \lambda)g(y; \omega, \tau), & \text{se } y \in (0, 1), \end{cases} \quad (4.8)$$

em que  $g(y; \omega, \tau)$  é expressa em (4.5) com  $\tau \equiv d_\alpha$ .

#### 4.3.1 Dispersão fixa

O modelo de regressão Kumaraswamy inflacionado em  $c$  considerando dispersão fixa, com  $c = 0$  ou  $c = 1$ , é definido por

$$\begin{cases} h(\lambda_i) = \sum_{j=1}^m z_{ij}\pi_j = \zeta_i \\ f(\omega_i) = \sum_{t=1}^r x_{it}\gamma_t = \eta_i, \end{cases} \quad i = 1, \dots, n, \quad (4.9)$$

em que  $h : (0, 1) \rightarrow \mathbb{R}$  e  $f : (0, 1) \rightarrow \mathbb{R}$  são funções de ligação estritamente monótonas e duas vezes diferenciáveis,  $z_i = (z_{i1}, \dots, z_{im})$  e  $x_i = (x_{i1}, \dots, x_{ir})$ ,  $m + r < n$ , são as variáveis regressoras conhecidas, que podem coincidir total ou parcialmente, e  $\pi = (\pi_1, \dots, \pi_m)^\top$  e  $\gamma = (\gamma_1, \dots, \gamma_r)^\top$  são os vetores de parâmetros do modelo a serem estimados. Diferentes funções de ligação  $h$  e  $f$  podem ser usadas, por exemplo, a função *logit*  $f(\omega) = \ln[\omega/(1-\omega)]$ , a função log-log  $f(\omega) = \ln[-\ln(\omega)]$ , a função log-log complementar  $f(\omega) = \ln[-\ln(1-\omega)]$ , entre outras, tanto para  $f(\omega)$  quanto para  $h(\lambda)$ .

É possível reescrever a densidade em (4.8) de forma que ela seja composta por dois termos, um que depende apenas de  $\lambda$  e outro que depende apenas de  $\omega$  e  $\tau$ :

$$\text{ki}_c(y; \lambda, \omega, \tau) = \left[ \lambda^{\mathbb{I}_{\{c\}}(y)} (1 - \lambda)^{1 - \mathbb{I}_{\{c\}}(y)} \right] \times \left[ g(y; \omega, \tau)^{1 - \mathbb{I}_{\{c\}}(y)} \right], \quad (4.10)$$

em que  $\mathbb{I}_{\{c\}}(y) = 1$  se  $y = 0$  ou  $y = 1$ ,  $\mathbb{I}_{\{c\}}(y) = 0$  se  $y \in (0, 1)$  e  $g(y; \omega, \tau)$  é a função densidade reparametrizada da distribuição Kumaraswamy expressa em (4.5), com  $\tau \equiv d_\alpha$ .

A função de verossimilhança para  $\boldsymbol{\theta} = (\pi^\top, \gamma^\top, \tau)^\top$ , baseada na densidade expressa em (4.10), é dada por

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^n \text{ki}_c(y_i; \lambda_i, \omega_i, \tau) = L_1(\pi) \times L_2(\gamma, \tau),$$

em que

$$L_1(\pi) = \prod_{i=1}^n \lambda_i^{\mathbb{I}_{\{c\}}(y_i)} (1 - \lambda_i)^{1 - \mathbb{I}_{\{c\}}(y_i)} \quad \text{e} \quad L_2(\gamma, \tau) = \prod_{i=1}^n g(y_i; \omega_i, \tau)^{1 - \mathbb{I}_{\{c\}}(y_i)},$$

$\mathbb{I}_{\{c\}}(y_i)$  sendo a função indicadora que vale 1 quando  $y = c$  e 0 caso contrário,  $\lambda_i = h^{-1}(\zeta_i)$ ,  $\omega_i = f^{-1}(\eta_i)$  e  $g(y_i; \omega_i, \tau)$  sendo a função densidade da distribuição Kumaraswamy reparametrizada expressa em (4.5), com  $\tau \equiv d_\alpha$ .

De forma análoga à função de densidade expressa em (4.10), a função de verossimilhança para  $\boldsymbol{\theta} = (\pi^\top, \gamma^\top, \tau)^\top$  também pode ser fatorada em dois termos, um dependente apenas do parâmetro de mistura  $\lambda$  e outro termo que depende da mediana e do parâmetro de dispersão,  $\omega$  e  $\tau$ . Logo, a função de log-verossimilhança para  $\boldsymbol{\theta} = (\pi^\top, \gamma^\top, \tau)^\top$  é dada por

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \ell_1(\pi) + \ell_2(\gamma, \tau),$$

em que

$$\begin{aligned} \ell_1(\pi) &= \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \ln(\lambda_i) + \sum_{i=1}^n [1 - \mathbb{I}_{\{c\}}(y_i)] \ln(1 - \lambda_i) \quad \text{e} \\ \ell_2(\gamma, \tau) &= \sum_{i=1}^n [1 - \mathbb{I}_{\{c\}}(y_i)] \ln[g(y_i; \omega_i, \tau)] = \ln\left(\frac{1}{\tau}\right) \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right] \\ &+ \sum_{i=1}^n \ln \left[ \frac{\ln(0.5)}{\ln(1 - \omega_i^{1/\tau})} \right] [1 - \mathbb{I}_{\{c\}}(y_i)] + \left( \frac{1}{\tau} - 1 \right) \sum_{i=1}^n \ln(y_i) [1 - \mathbb{I}_{\{c\}}(y_i)] \\ &+ \sum_{i=1}^n \left[ \frac{\ln(0.5)}{\ln(1 - \omega_i^{1/\tau})} - 1 \right] \ln(1 - y_i^{1/\tau}) [1 - \mathbb{I}_{\{c\}}(y_i)], \end{aligned}$$

que similarmente fatora em dois termos conforme descrito para a função de verossimilhança. Note que é preciso utilizar a convenção  $\ln(0) \times 0 = 0$ .

Diferenciando a função de log-verossimilhança uma vez obtém-se a função escore, cujos componentes são

$$U_{\pi_j}(\pi) = \frac{\partial \ell_1(\pi)}{\partial \pi_j} = \frac{\partial \ell_1(\pi)}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial \pi_j} = \sum_{i=1}^n \frac{\mathbb{I}_{\{c\}}(y_i) - \lambda_i}{\lambda_i(1 - \lambda_i)} \frac{\partial \lambda_i}{\partial \pi_j},$$



$$\begin{aligned}
U_{\gamma_t}(\gamma, \tau) &= \frac{\partial \ell_2(\gamma, \tau)}{\partial \gamma_t} = \frac{\partial \ell_2(\gamma, \tau)}{\partial \omega_i} \frac{\partial \omega_i}{\partial \gamma_t} \\
&= \frac{1}{\tau} \sum_{i=1}^n \omega_i^{\tau-1-1} [1 - \mathbb{I}_{\{c\}}(y_i)] \left[ \frac{\ln(1 - \omega_i^{\tau-1}) + \ln(0.5) \ln(1 - y_i^{\tau-1})}{(1 - \omega_i^{\tau-1}) \ln^2(1 - \omega_i^{\tau-1})} \right] \frac{\partial \omega_i}{\partial \gamma_t}, \\
U_{\tau}(\gamma, \tau) &= \frac{\partial \ell_2(\gamma, \tau)}{\partial \tau} = -\frac{1}{\tau} \left[ n - \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \right] - \frac{1}{\tau^2} \sum_{i=1}^n \ln(y_i) [1 - \mathbb{I}_{\{c\}}(y_i)] \\
&\quad - \frac{1}{\tau^2} \sum_{i=1}^n [1 - \mathbb{I}_{\{c\}}(y_i)] \left\{ \omega_i^{\tau-1} \ln(\omega_i) \left[ \frac{\ln(1 - \omega_i^{\tau-1}) + \ln(0.5) \ln(1 - y_i^{\tau-1})}{(1 - \omega_i^{\tau-1}) \ln^2(1 - \omega_i^{\tau-1})} \right] \right. \\
&\quad \left. - \frac{y_i^{\tau-1} \ln(y_i) [\ln(0.5) - \ln(1 - \omega_i^{\tau-1})]}{(1 - y_i^{\tau-1}) \ln(1 - \omega_i^{\tau-1})} \right\}.
\end{aligned}$$

Note que

$$\frac{\partial \lambda_i}{\partial \pi_j} = \frac{\partial \lambda_i}{\partial \zeta_i} \frac{\partial \zeta_i}{\partial \pi_j} = \frac{\partial h^{-1}(\zeta_i)}{\partial \zeta_i} z_{ij} \quad \text{e} \quad \frac{\partial \omega_i}{\partial \gamma_t} = \frac{\partial \omega_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \gamma_t} = \frac{\partial f^{-1}(\eta_i)}{\partial \eta_i} x_{it},$$

que dependem das funções de ligação utilizadas para  $\lambda$  e  $\omega$ .

Após realizar estimação pontual é possível construir intervalos de confiança para os parâmetros que indexam o modelo. Os intervalos de confiança assintóticos para  $\pi_j$ ,  $\gamma_t$  e  $\tau$  são dados, respectivamente, por

$$\hat{\pi}_j \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\pi}_j), \quad \hat{\gamma}_t \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\gamma}_t) \quad \text{e} \quad \hat{\tau} \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\tau}),$$

em que  $\text{ep}(\cdot)$  é o erro-padrão de cada estimador, que é obtido pela raiz quadrada dos elementos da diagonal principal da inversa da matriz de informação, substituindo os parâmetros pelos estimadores de máxima verossimilhança;  $z_{(1-\frac{\delta}{2})}$  é o quantil  $1 - \frac{\delta}{2}$  da distribuição normal padrão e  $1 - \delta$  é o nível de confiança adotado.

### 4.3.2 Dispersão variável

É possível modelar o parâmetro de dispersão, inserindo um submodelo para  $\tau$  no modelo apresentado em (4.9). Considere agora o seguinte modelo Kumaraswamy inflacionado em zero ou um com dispersão variável:

$$\begin{cases} h(\lambda_i) = \sum_{j=1}^m z_{ij} \pi_j = \zeta_i \\ f(\omega_i) = \sum_{t=1}^r x_{it} \gamma_t = \eta_i, \\ d(\tau_i) = \sum_{b=1}^u q_{ib} \varsigma_b = \Upsilon_i, \end{cases} \quad i = 1, \dots, n,$$

em que  $h : (0, 1) \rightarrow \mathbb{R}$ ,  $f : (0, 1) \rightarrow \mathbb{R}$  e  $d : (0, \infty) \rightarrow \mathbb{R}$  são funções de ligação estritamente monótonas e duas vezes diferenciáveis,  $z_i = (z_{i1}, \dots, z_{im})$ ,  $x_i = (x_{i1}, \dots, x_{ir})$  e

$q_i = (q_{i1}, \dots, q_{iu})$ ,  $m + r + u < n$ , são variáveis regressoras conhecidas e  $\pi = (\pi_1, \dots, \pi_m)^\top$ ,  $\gamma = (\gamma_1, \dots, \gamma_r)^\top$  e  $\varsigma = (\varsigma_1, \dots, \varsigma_u)^\top$  são os parâmetros do modelo a serem estimados.

Para esse modelo a função de log-verossimilhança de  $\theta = (\pi^\top, \gamma^\top, \varsigma^\top)^\top$  é dada por

$$\ell(\theta; \mathbf{y}) = \ell_1(\pi) + \ell_2(\gamma, \varsigma),$$

em que

$$\begin{aligned} \ell_1(\pi) &= \sum_{i=1}^n \mathbb{I}_{\{c\}}(y_i) \ln(\lambda_i) + \sum_{i=1}^n [1 - \mathbb{I}_{\{c\}}(y_i)] \ln(1 - \lambda_i) \quad \text{e} \\ \ell_2(\gamma, \varsigma) &= \sum_{i=1}^n [1 - \mathbb{I}_{\{c\}}(y_i)] \ln[g(y_i; \omega_i, \tau_i)] = \sum_{i=1}^n [1 - \mathbb{I}_{\{c\}}(y_i)] \ln\left(\frac{1}{\tau_i}\right) \\ &+ \sum_{i=1}^n \ln\left[\frac{\ln(0.5)}{\ln\left(1 - \omega_i^{1/\tau_i}\right)}\right] [1 - \mathbb{I}_{\{c\}}(y_i)] + \sum_{i=1}^n \left(\frac{1}{\tau_i} - 1\right) \ln(y_i) [1 - \mathbb{I}_{\{c\}}(y_i)] \\ &+ \sum_{i=1}^n \left[\frac{\ln(0.5)}{\ln\left(1 - \omega_i^{1/\tau_i}\right)} - 1\right] \ln\left(1 - y_i^{1/\tau_i}\right) [1 - \mathbb{I}_{\{c\}}(y_i)]. \end{aligned}$$

Novamente, utilizamos a convenção  $\ln(0) \times 0 = 0$ .

Agora tome como funções de ligação  $h(\lambda_i) = \ln[-\ln(1 - \lambda_i)] = \zeta_i$ ,  $f(\omega_i) = \ln[\omega_i/(1 - \omega_i)] = \eta_i$  e  $d(\tau_i) = \ln(\tau_i) = \Upsilon_i$ . É possível obter a função escore, cujos elementos são

$$\begin{aligned} U_{\pi_j}(\pi) &= \frac{\partial \ell_1(\pi)}{\partial \pi_j} = \frac{\partial \ell_1(\pi)}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial \zeta_i} \frac{\partial \zeta_i}{\partial \pi_j} = \sum_{i=1}^n \frac{\mathbb{I}_{\{c\}}(y_i) - \lambda_i}{\lambda_i(1 - \lambda_i)} e^{\zeta_i - e^{\zeta_i}} z_{ij}, \quad j = 1, \dots, m, \\ U_{\gamma_t}(\gamma, \varsigma) &= \frac{\partial \ell_2(\gamma, \varsigma)}{\partial \gamma_t} = \frac{\partial \ell_2(\gamma, \varsigma)}{\partial \omega_i} \frac{\partial \omega_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \gamma_t} \\ &= \sum_{i=1}^n \left[ \frac{\ln\left(1 - \omega_i^{\tau_i^{-1}}\right) + \ln(0.5) \ln\left(1 - y_i^{\tau_i^{-1}}\right)}{\tau_i \left(1 - \omega_i^{\tau_i^{-1}}\right) \ln^2\left(1 - \omega_i^{\tau_i^{-1}}\right)} \right] \omega_i^{\tau_i^{-1} - 1} \frac{x_{it} e^{\eta_i} [1 - \mathbb{I}_{\{c\}}(y_i)]}{(1 + e^{\eta_i})^2}, \\ &\quad t = 1, \dots, r, \\ U_{\varsigma_b}(\gamma, \varsigma) &= \frac{\partial \ell_2(\gamma, \varsigma)}{\partial \varsigma_b} = \frac{\partial \ell_2(\gamma, \varsigma)}{\partial \tau_i} \frac{\partial \tau_i}{\partial \Upsilon_i} \frac{\partial \Upsilon_i}{\partial \varsigma_b} = - \sum_{i=1}^n \left\{ \frac{1}{\tau_i} + \frac{1}{\tau_i^2} \ln(y_i) \right. \\ &+ \frac{1}{\tau_i^2} \frac{\omega_i^{\tau_i^{-1}} \ln(\omega_i) \left[ \ln\left(1 - \omega_i^{\tau_i^{-1}}\right) + \ln(0.5) \ln\left(1 - y_i^{\tau_i^{-1}}\right) \right]}{\left(1 - \omega_i^{\tau_i^{-1}}\right) \ln^2\left(1 - \omega_i^{\tau_i^{-1}}\right)} \\ &+ \left. \frac{1}{\tau_i^2} \frac{y_i^{\tau_i^{-1}} \ln(y_i) \left[ \ln(0.5) - \ln\left(1 - \omega_i^{\tau_i^{-1}}\right) \right]}{\left(1 - y_i^{\tau_i^{-1}}\right) \ln\left(1 - \omega_i^{\tau_i^{-1}}\right)} \right\} [1 - \mathbb{I}_{\{c\}}(y_i)] e^{\Upsilon_i} q_{ib}, \\ &\quad b = 1, \dots, u. \end{aligned}$$

Os intervalos de confiança assintóticos para os parâmetros que indexam esse modelo são dados por

$$\hat{\pi}_j \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\pi}_j), \quad \hat{\gamma}_t \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\gamma}_t) \quad \text{e} \quad \hat{\varsigma}_b \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\varsigma}_b),$$

em que  $\text{ep}(\cdot)$  é o erro-padrão de cada estimador, obtido pela raiz quadrada dos elementos da diagonal principal da inversa da matriz de informação, substituindo os parâmetros pelos estimadores de máxima verossimilhança;  $z_{(1-\frac{\delta}{2})}$  é o quantil  $1 - \frac{\delta}{2}$  da distribuição normal padrão e  $1 - \delta$  é o nível de confiança adotado.

### 4.3.3 Testes de hipóteses

Além de estimação pontual e intervalos de confiança, também será abordado o teste da razão de verossimilhanças. Considere que o interesse é testar restrições para um subconjunto de parâmetros de  $\boldsymbol{\theta}$ , então  $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$  vs.  $H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^{(0)}$ , em que  $\boldsymbol{\theta}_1$  é o vetor  $r \times 1$  de parâmetros de interesse. A estatística do teste da razão de verossimilhanças é dada por

$$RV = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})],$$

em que  $\hat{\boldsymbol{\theta}}$  é o estimador de máxima verossimilhança irrestrito de  $\boldsymbol{\theta}$  e  $\tilde{\boldsymbol{\theta}} = \left(\boldsymbol{\theta}_1^{(0)\top}, \tilde{\boldsymbol{\theta}}_2^\top\right)^\top$  é o estimador de máxima verossimilhança restrito de  $\boldsymbol{\theta}$ , obtido pela imposição de  $H_0$ .

Sob condições usuais de regularidade e sob a hipótese nula, a estatística da razão de verossimilhanças tem distribuição assintótica  $\chi_r^2$ , em que  $r$  é o número de restrições impostas em  $H_0$ . Assim,  $H_0$  é rejeitada se  $RV > \chi_{r,1-\delta}^2$ .

No caso particular em que  $\theta_1$  é um escalar, além do teste da razão de verossimilhanças também é possível utilizar o teste Z, cuja estatística de teste apresenta a seguinte forma:

$$Z = \frac{\hat{\theta}_1 - \theta_1^{(0)}}{\text{ep}(\hat{\theta}_1)},$$

em que  $\hat{\theta}_1$  é o estimador de máxima verossimilhança do parâmetro de interesse,  $\theta_1^{(0)}$  é o valor nulo e  $\text{ep}(\hat{\theta}_1)$  é o erro-padrão de  $\hat{\theta}_1$ .

A distribuição aproximada da estatística  $Z$  é normal padrão. Logo,  $H_0$  é rejeitada se  $Z > z_{(1-\frac{\delta}{2})}$  ou se  $Z < -z_{(1-\frac{\delta}{2})}$ , em que  $z$  é o quantil da distribuição normal padrão.

## 4.4 Resultados numéricos para o modelo $\text{KI}_c$

Foram realizadas simulações de Monte Carlo para avaliar os desempenhos dos estimadores de máxima verossimilhança (EMV's) dos parâmetros que indexam o modelo Kumaraswamy inflacionado em zero. Para atingir tal objetivo, foram considerados diversos cenários: C1) submodelo apenas para o parâmetro de mistura,  $\lambda$ ; C2) submodelos para  $\lambda$  e  $\omega$ , sendo esses dois cenários com dispersão fixa, e por fim, C3) submodelos para  $\lambda$ ,  $\omega$  e  $\tau$ , sendo este último modelo com dispersão variável.

Nas simulações foram utilizadas as funções  $\lambda_i = 1 - e^{-e^{\pi_0 + z_{i1}\pi_1}}$ ,  $\omega_i = 1/[1 + e^{-\gamma_0 - x_{i1}\gamma_1}]$  e  $\tau_i = e^{\varsigma_0 + q_{i1}\varsigma_1}$ ,  $i = 1, 2, \dots, n$ . Os valores de  $z_1$ ,  $x_1$  e  $q_1$  foram gerados aleatoriamente da distribuição  $\mathcal{U}(0, 1)$  e os valores dos parâmetros utilizados nas simulações são  $\pi_0 = -0.5$ ,  $\pi_1 = -1.5$ ,  $\gamma_0 = 1.0$ ,  $\gamma_1 = -2.0$ ,  $\varsigma_0 = 0.5$  e  $\varsigma_1 = 1.5$ . Os tamanhos amostrais utilizados nas simulações foram  $n = (30, 50, 100, 200, 500, 1000)$  e foram realizadas 10000 réplicas de

**Tabela 4.1:** Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 1;  $\pi_0 = -0.5, \pi_1 = -1.5$  e  $\tau = 2.0$ .

$\omega$	Medida	Estimador	$n$				
			40	60	100	500	1000
0.25	Média	$\hat{\pi}_0$	-0.5271	-0.5361	-0.5106	-0.5019	-0.4999
		$\hat{\pi}_1$	-1.6270	-1.5391	-1.5367	-1.5066	-1.5032
		$\hat{\omega}$	0.2568	0.2544	0.2527	0.2503	0.2503
		$\hat{\tau}$	1.9553	1.9719	1.9817	1.9987	1.9972
	Variância	$\hat{\pi}_0$	0.5529	0.3674	0.1719	0.0250	0.0127
		$\hat{\pi}_1$	2.2066	1.3408	0.6638	0.1075	0.0538
		$\hat{\omega}$	0.0047	0.0032	0.0019	0.0004	0.0002
		$\hat{\tau}$	0.2168	0.1509	0.0909	0.0185	0.0091
	Viés	$\hat{\pi}_0$	-0.0271	-0.0361	-0.0106	-0.0019	0.0001
		$\hat{\pi}_1$	-0.1270	-0.0391	-0.0367	-0.0066	-0.0032
		$\hat{\omega}$	0.0068	0.0044	0.0027	-0.0003	0.0003
		$\hat{\tau}$	-0.0447	-0.0281	-0.0183	-0.0013	-0.0028
	EQM	$\hat{\pi}_0$	0.5536	0.3687	0.1720	0.0250	0.0127
		$\hat{\pi}_1$	2.2227	1.3423	0.6651	0.1076	0.0538
		$\hat{\omega}$	0.0048	0.0032	0.0020	0.0004	0.0002
		$\hat{\tau}$	0.2188	0.1517	0.0913	0.0185	0.0091
0.50	Média	$\hat{\pi}_0$	-0.5271	-0.5361	-0.5106	-0.5018	-0.4998
		$\hat{\pi}_1$	-1.6270	-1.5391	-1.5367	-1.5067	-1.5033
		$\hat{\omega}$	0.5027	0.5015	0.5010	0.4999	0.5001
		$\hat{\tau}$	1.9533	1.9706	1.9803	1.9988	1.9968
	Variância	$\hat{\pi}_0$	0.5529	0.3674	0.1719	0.0250	0.0127
		$\hat{\pi}_1$	2.2066	1.3408	0.6638	0.1075	0.0538
		$\hat{\omega}$	0.0093	0.0064	0.0039	0.0008	0.0004
		$\hat{\tau}$	0.3041	0.2123	0.1273	0.0261	0.0129
	Viés	$\hat{\pi}_0$	-0.0271	-0.0361	-0.0106	-0.0018	0.0002
		$\hat{\pi}_1$	-0.1270	-0.0391	-0.0367	-0.0067	-0.0033
		$\hat{\omega}$	0.0027	0.0015	0.0010	-0.0001	0.0001
		$\hat{\tau}$	-0.0467	-0.0294	-0.0197	-0.0012	-0.0032
	EQM	$\hat{\pi}_0$	0.5536	0.3687	0.1720	0.0250	0.0127
		$\hat{\pi}_1$	2.2227	1.3423	0.6651	0.1076	0.0538
		$\hat{\omega}$	0.0093	0.0064	0.0039	0.0008	0.0004
		$\hat{\tau}$	0.3063	0.2132	0.1277	0.0261	0.0129

Monte Carlo. As maximizações das funções de log-verossimilhança foram realizadas usando o método quasi-Newton BFGS (Nocedal, 1980) com gradiente analítico. As simulações foram realizadas usando a linguagem matricial de programação Ox (Doornik e Ooms, 2007). Os chutes iniciais utilizados nas simulações foram escolhidos arbitrariamente.

**C1)** Primeiro foi usado um cenário no qual só o parâmetro de mistura é modelado, a saber:

$$h(\lambda_i) = \pi_0 + z_{i1}\pi_1, \quad i = 1, \dots, n.$$

Neste caso, a mediana é igual a 0.25 e 0.50, e o parâmetro de dispersão é fixo e igual a 2.0.

Na Tabela 4.1 encontram-se as estimativas médias, as variâncias, os vieses e os erros quadráticos médios (EQM) dos estimadores de máxima verossimilhança no

cenário C1). É possível observar que as médias das estimativas convergem para os verdadeiros valores dos parâmetros adotados nas simulações, bem como o viés, a variância e o EQM diminuem à medida que  $n$  aumenta, indicando bom desempenho dos estimadores de máxima verossimilhança.

**Tabela 4.2:** Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 1;  $\pi_0 = -0.5, \pi_1 = -1.5$  e  $\tau = 2.0$ .

$\omega$	$1 - \delta$	$\hat{\theta}$	$n$				
			40	60	100	500	1000
0.25	99%	$\hat{\pi}_0$	3.5813	3.0120	2.0843	0.8148	0.5771
		$\hat{\pi}_1$	7.0129	5.7174	4.0548	1.6739	1.1809
		$\hat{\omega}$	0.3483	0.2864	0.2246	0.1023	0.0724
		$\hat{\tau}$	2.4078	1.9767	1.5444	0.7017	0.4955
	95%	$\hat{\pi}_0$	2.7250	2.2918	1.5859	0.6200	0.4391
		$\hat{\pi}_1$	5.3361	4.3504	3.0853	1.2737	0.8986
		$\hat{\omega}$	0.2650	0.2179	0.1709	0.0778	0.0551
		$\hat{\tau}$	1.8321	1.5041	1.1751	0.5340	0.3770
	90%	$\hat{\pi}_0$	2.2869	1.9234	1.3310	0.5203	0.3685
		$\hat{\pi}_1$	4.4782	3.6509	2.5893	1.0689	0.7541
		$\hat{\omega}$	0.2224	0.1829	0.1434	0.0653	0.0462
		$\hat{\tau}$	1.5376	1.2623	0.9862	0.4481	0.3164
0.50	99%	$\hat{\pi}_0$	3.5813	3.0120	2.0843	0.8148	0.5771
		$\hat{\pi}_1$	7.0129	5.7174	4.0548	1.6740	1.1809
		$\hat{\omega}$	0.4890	0.4050	0.3192	0.1463	0.1063
		$\hat{\tau}$	2.8392	2.3440	1.8312	0.8324	0.5896
	95%	$\hat{\pi}_0$	2.7250	2.2918	1.5859	0.6200	0.4391
		$\hat{\pi}_1$	5.3361	4.3504	3.0853	1.2737	0.8986
		$\hat{\omega}$	0.3721	0.3081	0.2429	0.1113	0.0809
		$\hat{\tau}$	2.1603	1.7836	1.3933	0.6334	0.4487
	90%	$\hat{\pi}_0$	2.2869	1.9234	1.3310	0.5203	0.3685
		$\hat{\pi}_1$	4.4782	3.6509	2.5893	1.0689	0.7541
		$\hat{\omega}$	1.8130	0.2586	0.2038	0.0934	0.0679
		$\hat{\tau}$	0.3122	1.4968	1.1693	0.5315	0.3765

Além de estimação pontual foram construídos intervalos de confiança para cada parâmetro e as amplitudes médias dos intervalos encontram-se na Tabela 4.2. Nota-se que as amplitudes médias diminuem à medida que o tamanho amostral aumenta, sendo assim os intervalos se tornam mais precisos.

Já na Tabela 4.3 estão apresentadas as taxas de cobertura e não cobertura dos intervalos. Note que as taxas de cobertura convergem para os respectivos níveis de confiança utilizados quando  $n$  cresce; já as taxas de não cobertura apresentam comportamento menos assimétrico à medida que o tamanho amostral aumenta.

**C2)** Em seguida foi considerado um modelo para os parâmetros de mistura e de mediana, mas a dispersão é fixa:

$$\begin{cases} h(\lambda_i) = \pi_0 + z_{i1}\pi_1 \\ f(\omega_i) = \gamma_0 + x_{i1}\gamma_1, \quad i = 1, \dots, n. \end{cases}$$

**Tabela 4.3:** Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 1;  $\pi_0 = -0.5, \pi_1 = -1.5$  e  $\tau = 2.0$ .

$\omega$	$1 - \delta$	$\hat{\theta}$	$n$				
			40	60	100	500	1000
0.25	99%	$\hat{\pi}_0$	99.01	98.99	98.91	98.96	98.95
			(0.98; 0.01)	(0.96; 0.05)	(0.97; 0.12)	(0.67; 0.37)	(0.68; 0.37)
		$\hat{\pi}_1$	99.64	99.30	99.11	98.80	98.99
			(0.28; 0.08)	(0.39; 0.31)	(0.46; 0.43)	(0.60; 60)	(0.55; 0.46)
		$\hat{\omega}$	97.30	97.68	98.43	98.81	98.94
			(0.53; 2.17)	(0.58; 1.74)	(0.36; 1.21)	(0.35; 0.84)	(0.42; 0.57)
		$\hat{\tau}$	94.92	95.93	97.05	98.61	99.01
			(0.00; 5.08)	(0.00; 4.07)	(0.00; 2.95)	(0.12; 1.27)	(0.16; 0.90)
		$\hat{\pi}_0$	95.38	95.34	95.06	95.08	94.99
			(3.80; 0.82)	(3.30; 1.36)	(3.59; 1.35)	(2.77; 2.15)	(2.92; 2.09)
		$\hat{\pi}_1$	95.77	95.17	94.91	94.78	94.88
			(2.27; 1.96)	(2.53; 2.30)	(2.33; 2.76)	(2.56; 2.66)	(2.75; 2.37)
		$\hat{\omega}$	92.67	93.15	93.76	94.83	95.17
			(2.55; 4.78)	(2.42; 4.43)	(2.36; 3.88)	(2.11; 3.06)	(2.21; 2.62)
		$\hat{\tau}$	90.55	91.51	92.97	94.70	95.22
			(0.03; 9.42)	(0.15; 8.34)	(0.33; 6.70)	(1.18; 4.12)	(1.42; 3.36)
	90%	$\hat{\pi}_0$	90.20	90.30	89.75	90.31	89.60
			(6.86; 2.94)	(6.12; 3.58)	(6.61; 3.64)	(5.25; 4.44)	(5.77; 4.63)
		$\hat{\pi}_1$	90.55	90.25	89.84	90.06	89.26
			(4.65; 4.80)	(4.80; 4.95)	(4.77; 5.39)	(4.93; 5.01)	(5.66; 5.08)
		$\hat{\omega}$	88.08	88.07	88.85	89.93	89.95
			(4.76; 7.16)	(4.61; 7.32)	(4.61; 6.54)	(4.56; 5.51)	(4.77; 5.28)
		$\hat{\tau}$	86.68	87.08	88.11	90.05	90.13
			(0.29; 13.03)	(0.84; 12.08)	(1.63; 10.26)	(3.17; 6.78)	(3.39; 6.48)
0.50	99%	$\hat{\pi}_0$	99.01	98.99	98.91	98.96	98.95
			(0.98; 0.01)	(0.96; 0.05)	(0.97; 0.12)	(0.67; 0.37)	(0.68; 0.37)
		$\hat{\pi}_1$	99.64	99.30	99.11	98.80	98.99
			(0.28; 0.08)	(0.39; 0.31)	(0.46; 0.43)	(0.60; 0.60)	(0.55; 0.46)
		$\hat{\omega}$	97.27	97.70	98.32	98.97	99.04
			(1.36; 1.37)	(1.26; 1.04)	(0.87; 0.81)	(0.44; 0.59)	(0.50; 0.45)
		$\hat{\tau}$	94.11	95.32	96.58	98.46	98.84
			(0.00; 5.88)	(0.01; 4.67)	(0.00; 3.42)	(0.12; 1.42)	(0.16; 0.99)
		$\hat{\pi}_0$	95.38	95.34	95.06	95.08	94.99
			(3.80; 0.82)	(3.30; 1.36)	(3.59; 1.35)	(2.77; 2.15)	(2.92; 2.09)
		$\hat{\pi}_1$	95.77	95.17	94.91	94.78	94.88
			(2.27; 1.96)	(2.53; 2.30)	(2.33; 2.76)	(2.56; 2.66)	(2.75; 2.37)
		$\hat{\omega}$	92.14	93.01	93.67	94.83	95.14
			(4.10; 3.76)	(3.60; 3.39)	(3.40; 2.93)	(2.57; 2.60)	(2.47; 2.38)
		$\hat{\tau}$	89.50	90.85	92.56	94.51	95.14
			(0.03; 10.46)	(0.07; 9.08)	(0.17; 7.27)	(1.13; 4.36)	(1.35; 3.50)
	90%	$\hat{\pi}_0$	90.20	90.30	89.75	90.31	89.60
			(6.86; 2.94)	(6.12; 3.58)	(6.61; 3.64)	(5.25; 4.44)	(5.77; 4.63)
		$\hat{\pi}_1$	90.55	90.25	89.84	90.06	89.26
			(4.65; 4.80)	(4.80; 4.95)	(4.77; 5.39)	(4.93; 5.01)	(5.66; 5.08)
		$\hat{\omega}$	87.16	87.73	88.67	89.91	89.71
			(6.87; 5.97)	(6.28; 5.99)	(5.77; 5.56)	(5.08; 5.01)	(5.29; 4.99)
		$\hat{\tau}$	85.88	86.62	88.04	89.74	90.33
			(0.20; 13.91)	(0.69; 12.69)	(1.17; 10.79)	(2.98; 7.28)	(3.08; 6.58)

**Tabela 4.4:** Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 2;  $\pi_0 = -0.5$ ,  $\pi_1 = -1.5$ ,  $\gamma_0 = 1.0$  e  $\gamma_1 = -2.0$ .

$\tau$	Medida	Estimador	$n$				
			40	60	100	500	1000
0.90	Média	$\hat{\pi}_0$	-0.5271	-0.5361	-0.5106	-0.5019	-0.4999
		$\hat{\pi}_1$	-1.6270	-1.5391	-1.5367	-1.5066	-1.5032
		$\hat{\gamma}_0$	1.0086	1.0100	1.0087	0.9993	1.0016
		$\hat{\gamma}_1$	-2.0046	-2.0115	-2.0115	-1.9997	-2.0023
		$\hat{\tau}$	0.8568	0.8722	0.8827	0.8974	0.8979
	Variância	$\hat{\pi}_0$	0.5529	0.3674	0.1719	0.0250	0.0127
		$\hat{\pi}_1$	2.2065	1.3408	0.6638	0.1075	0.0538
		$\hat{\gamma}_0$	0.4973	0.3423	0.1869	0.0293	0.0141
		$\hat{\gamma}_1$	1.1269	0.7822	0.4296	0.0666	0.0320
		$\hat{\tau}$	0.0380	0.0265	0.0163	0.0034	0.0017
	Viés	$\hat{\pi}_0$	-0.0271	-0.0361	-0.0106	-0.0019	0.0001
		$\hat{\pi}_1$	-0.1270	-0.0391	-0.0367	-0.0066	-0.0032
		$\hat{\gamma}_0$	0.0086	0.0100	0.0087	-0.0007	0.0016
		$\hat{\gamma}_1$	-0.0046	-0.0115	-0.0115	0.0003	-0.0023
		$\hat{\tau}$	-0.0432	-0.0278	-0.0173	-0.0026	-0.0021
	EQM	$\hat{\pi}_0$	0.5536	0.3687	0.1720	0.0250	0.0127
		$\hat{\pi}_1$	2.2227	1.3423	0.6651	0.1076	0.0538
		$\hat{\gamma}_0$	0.4974	0.3424	0.1870	0.0293	0.0141
		$\hat{\gamma}_1$	1.1269	0.7824	0.4297	0.0666	0.0320
		$\hat{\tau}$	0.0398	0.0273	0.0166	0.0034	0.0017
2.00	Média	$\hat{\pi}_0$	-0.5271	-0.5361	-0.5106	-0.5019	-0.4999
		$\hat{\pi}_1$	-1.6270	-1.5391	-1.5367	-1.5064	-1.5031
		$\hat{\gamma}_0$	1.0016	1.0069	1.0084	0.9980	1.0017
		$\hat{\gamma}_1$	-2.0004	-2.0119	-2.0144	-1.9992	-2.0032
		$\hat{\tau}$	1.8996	1.9353	1.9592	1.9943	1.9947
	Variância	$\hat{\pi}_0$	0.5529	0.3674	0.1719	0.0249	0.0126
		$\hat{\pi}_1$	2.2066	1.3408	0.6638	0.1075	0.0538
		$\hat{\gamma}_0$	1.1632	0.8008	0.4283	0.0661	0.0319
		$\hat{\gamma}_1$	2.7884	1.9317	1.0438	0.1623	0.0785
		$\hat{\tau}$	0.2788	0.1963	0.1200	0.0252	0.0123
	Viés	$\hat{\pi}_0$	-0.0271	-0.0361	-0.0106	-0.0019	0.0001
		$\hat{\pi}_1$	-0.1270	-0.0391	-0.0367	-0.0064	-0.0031
		$\hat{\gamma}_0$	0.0016	0.0069	0.0084	-0.0020	0.0017
		$\hat{\gamma}_1$	-0.0004	-0.0119	-0.0144	0.0008	-0.0032
		$\hat{\tau}$	-0.1004	-0.0647	-0.0408	-0.0057	-0.0053
	EQM	$\hat{\pi}_0$	0.5536	0.3687	0.1720	0.0249	0.0126
		$\hat{\pi}_1$	2.2227	1.3423	0.6651	0.1075	0.0538
		$\hat{\gamma}_0$	1.1632	0.8009	0.4284	0.0661	0.0319
		$\hat{\gamma}_1$	2.7884	1.9319	1.0440	0.1623	0.0785
		$\hat{\tau}$	0.2889	0.2005	0.1217	0.0253	0.0124

Os resultados da simulação de estimação pontual para esse cenário encontram-se na Tabela 4.4. Nota-se que os estimadores de máxima verossimilhança dos parâmetros do modelo no cenário 2 continuam apresentando bom desempenho à medida que o tamanho amostral aumenta, visto que as médias das estimativas convergem para os verdadeiros valores dos parâmetros, bem como os vieses, as variâncias e os EQMs diminuem.

**Tabela 4.5:** Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 2;  $\pi_0 = -0.5$ ,  $\pi_1 = -1.5$ ,  $\gamma_0 = 1.0$  e  $\gamma_1 = -2.0$ .

$\tau$	$1 - \delta$	$\hat{\theta}$	$n$				
			40	60	100	500	1000
0.90	99%	$\hat{\pi}_0$	3.5813	3.0120	2.0843	0.8148	0.5771
		$\hat{\pi}_1$	7.0129	5.7174	4.0548	1.6739	1.1809
		$\hat{\gamma}_0$	3.4265	2.8985	2.1634	0.8767	0.6154
		$\hat{\gamma}_1$	5.1461	4.3652	3.2686	1.3242	0.9221
		$\hat{\tau}$	0.9872	0.8180	0.6468	0.2992	0.2113
	95%	$\hat{\pi}_0$	2.7250	2.2918	1.5859	0.6200	0.4391
		$\hat{\pi}_1$	5.3361	4.3504	3.0853	1.2737	0.8986
		$\hat{\gamma}_0$	2.6073	2.2055	1.6461	0.6671	0.4683
		$\hat{\gamma}_1$	3.9157	3.3215	2.4871	1.0076	0.7016
		$\hat{\tau}$	0.7512	0.6224	0.4922	0.2277	0.1608
	90%	$\hat{\pi}_0$	2.2869	1.9234	1.3310	0.5203	0.3685
		$\hat{\pi}_1$	4.4782	3.6509	2.5893	1.0689	0.7541
		$\hat{\gamma}_0$	2.1881	1.8509	1.3815	0.5598	0.3930
		$\hat{\gamma}_1$	3.2861	2.7875	2.0872	0.8456	0.5888
		$\hat{\tau}$	0.6304	0.5223	0.4130	0.1911	0.1349
2.00	99%	$\hat{\pi}_0$	3.5813	3.0120	2.0843	0.8148	0.5771
		$\hat{\pi}_1$	7.0129	5.7174	4.0548	1.6739	1.1810
		$\hat{\gamma}_0$	5.2510	4.4389	3.2801	1.3176	0.9254
		$\hat{\gamma}_1$	8.1301	6.8799	5.1105	2.0695	1.4430
		$\hat{\tau}$	2.6541	2.2244	1.7586	0.8195	0.5851
	95%	$\hat{\pi}_0$	2.7250	2.2918	1.5859	0.6200	0.4391
		$\hat{\pi}_1$	5.3361	4.3504	3.0853	1.2737	0.8986
		$\hat{\gamma}_0$	3.9955	3.3776	2.4959	1.0025	0.7042
		$\hat{\gamma}_1$	6.1862	5.2349	3.8886	1.5747	1.0980
		$\hat{\tau}$	2.0196	1.6926	1.3381	0.6236	0.4452
	90%	$\hat{\pi}_0$	2.2869	1.9234	1.3310	0.5203	0.3685
		$\hat{\pi}_1$	4.4782	3.6509	2.5893	1.0689	0.7541
		$\hat{\gamma}_0$	3.3531	2.8346	2.0946	0.8414	0.5909
		$\hat{\gamma}_1$	5.1917	4.3933	3.2634	1.3215	0.9214
		$\hat{\tau}$	1.6949	1.4204	1.1230	0.5233	0.3736

Na Tabela 4.5 encontram-se as amplitudes médias dos intervalos de confiança. Note que elas diminuem à medida que o tamanho amostral aumenta, indicando maior precisão dos intervalos. Na Tabela 4.6 encontram-se as taxas de cobertura e não cobertura dos intervalos para os parâmetros do segundo cenário. Note que as taxas de cobertura convergem para os respectivos níveis de confiança utilizados e as taxas de não cobertura se tornam mais simétricas à medida que o tamanho amostral aumenta.

**C3)** Por fim, foi considerado o cenário em que a dispersão varia de acordo com as observações. Nesse contexto, o modelo de regressão é

$$\begin{cases} h(\lambda_i) = \pi_0 + z_{i1}\pi_1 \\ f(\omega_i) = \gamma_0 + x_{i1}\gamma_1 \\ d(\tau_i) = \varsigma_0 + q_{i1}\varsigma_1, \quad i = 1, \dots, n. \end{cases}$$

Note que o parâmetro de dispersão agora é variável.



**Tabela 4.6:** Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 2;  $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0$  e  $\gamma_1 = -2.0$ .

$\tau$	$1 - \delta$	$\hat{\theta}$	$n$				
			40	60	100	500	1000
0.90	99%	$\hat{\pi}_0$	99.01 (0.98; 0.01)	98.99 (0.96; 0.05)	98.91 (0.97; 0.12)	98.96 (0.67; 0.37)	98.95 (0.68; 0.37)
		$\hat{\pi}_1$	99.64 (0.28; 0.08)	99.30 (0.39; 0.31)	99.11 (0.46; 0.43)	98.80 (0.60; 0.60)	98.99 (0.55; 0.46)
		$\hat{\gamma}_0$	97.69 (0.30; 2.01)	98.15 (0.25; 1.60)	98.39 (0.25; 1.36)	98.89 (0.40; 0.71)	99.08 (0.33; 0.59)
		$\hat{\gamma}_1$	97.90 (1.71; 0.39)	98.21 (1.28; 0.51)	98.68 (1.00; 0.32)	98.93 (0.66; 0.41)	98.99 (0.57; 0.44)
		$\hat{\tau}$	93.4136 (0.00; 6.64)	95.16 (0.00; 4.84)	96.63 (0.00; 3.37)	98.58 (0.11; 1.31)	98.82 (0.21; 0.97)
		$\hat{\pi}_0$	95.38 (3.80; 0.82)	95.34 (3.30; 1.36)	95.06 (3.59; 1.35)	95.08 (2.77; 2.15)	94.99 (2.92; 2.09)
		$\hat{\pi}_1$	95.77 (2.27; 1.96)	95.17 (2.53; 2.30)	94.91 (2.33; 2.76)	94.78 (2.56; 2.66)	94.88 (2.75; 2.37)
		$\hat{\gamma}_0$	93.22 (1.74; 5.04)	93.88 (1.80; 4.32)	94.33 (1.80; 3.87)	94.69 (2.23; 3.08)	95.41 (2.15; 2.44)
		$\hat{\gamma}_1$	93.11 (4.57; 2.32)	93.87 (3.90; 2.23)	94.02 (3.69; 2.29)	94.90 (2.78; 2.32)	95.18 (2.40; 2.42)
		$\hat{\tau}$	88.25 (0.03; 11.72)	90.08 (0.14; 9.78)	92.04 (0.28; 7.68)	94.54 (1.19; 4.27)	95.08 (1.38; 3.54)
	90%	$\hat{\pi}_0$	90.20 (6.86; 2.94)	90.30 (6.12; 3.58)	89.75 (6.61; 3.64)	90.31 (5.25; 4.44)	89.60 (5.77; 4.63)
		$\hat{\pi}_1$	90.55 (4.65; 4.80)	90.25 (4.80; 4.95)	89.84 (4.77; 5.39)	90.06 (4.93; 5.01)	89.26 (5.66; 5.08)
		$\hat{\gamma}_0$	87.88 (4.22; 7.90)	88.88 (4.17; 6.95)	89.20 (4.17; 6.63)	90.08 (4.28; 5.64)	90.37 (4.58; 5.05)
		$\hat{\gamma}_1$	87.50 (7.47; 5.03)	88.57 (6.68; 4.75)	88.72 (6.35; 4.93)	90.05 (5.21; 4.74)	90.24 (4.94; 4.82)
		$\hat{\tau}$	84.03 (0.29; 15.68)	85.48 (0.73; 13.79)	87.09 (1.50; 11.41)	89.77 (2.97; 7.26)	90.13 (3.14; 6.73)

**Tabela 4.7:** Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 3;  $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$  e  $\varsigma_1 = 1.5$ .

Medida	Estimador	$n$				
		40	60	100	500	1000
Média	$\hat{\pi}_0$	-0.5275	-0.5365	-0.5104	-0.5020	-0.5000
	$\hat{\pi}_1$	-1.6267	-1.5383	-1.5370	-1.5061	-1.5032
	$\hat{\gamma}_0$	1.0347	1.0288	1.0252	1.0008	1.0038
	$\hat{\gamma}_1$	-2.0583	-2.0479	-2.0422	-2.0041	-2.0074
	$\hat{\varsigma}_0$	0.2304	0.3307	0.3947	0.4844	0.4908
	$\hat{\varsigma}_1$	1.7023	1.6181	1.5764	1.5081	1.5041
Variância	$\hat{\pi}_0$	0.5524	0.3675	0.1719	0.0250	0.0126
	$\hat{\pi}_1$	2.2055	1.3410	0.6639	0.1076	0.0536
	$\hat{\gamma}_0$	1.8132	1.2092	0.5997	0.0892	0.0436
	$\hat{\gamma}_1$	5.6528	3.7361	1.8990	0.3063	0.1506
	$\hat{\varsigma}_0$	1.0221	0.6126	0.2964	0.0414	0.0202
	$\hat{\varsigma}_1$	2.8064	1.6300	0.8126	0.1185	0.0561
Viés	$\hat{\pi}_0$	-0.0275	-0.0365	-0.0104	-0.0020	-0.0000
	$\hat{\pi}_1$	-0.1267	-0.0383	-0.0370	-0.0061	-0.0032
	$\hat{\gamma}_0$	0.0347	0.0288	0.0252	0.0008	0.0038
	$\hat{\gamma}_1$	-0.0583	-0.0479	-0.0422	-0.0041	-0.0074
	$\hat{\varsigma}_0$	-0.2696	-0.1693	-0.1053	-0.0156	-0.0092
	$\hat{\varsigma}_1$	0.2023	0.1181	0.0764	0.0081	0.0041
EQM	$\hat{\pi}_0$	0.5532	0.3689	0.1720	0.0250	0.0126
	$\hat{\pi}_1$	2.2216	1.3425	0.6653	0.1077	0.0536
	$\hat{\gamma}_0$	1.8144	1.2101	0.6003	0.0892	0.0436
	$\hat{\gamma}_1$	5.6562	3.7384	1.9008	0.3063	0.1507
	$\hat{\varsigma}_0$	1.0948	0.6412	0.3074	0.0416	0.0203
	$\hat{\varsigma}_1$	2.8473	1.6439	0.8184	0.1186	0.0561

**Tabela 4.8:** Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 3;  $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$  e  $\varsigma_1 = 1.5$ .

$1 - \delta$	$\hat{\theta}$	$n$				
		40	60	100	500	1000
99%	$\hat{\pi}_0$	2.9390	2.1651	1.6803	0.8019	0.5660
	$\hat{\pi}_1$	6.0276	4.6399	3.6352	1.6524	1.1611
	$\hat{\gamma}_0$	5.4494	4.0866	3.1362	1.4937	1.0540
	$\hat{\gamma}_1$	9.9808	7.7123	5.9990	2.7664	1.9400
	$\hat{\varsigma}_0$	3.9021	2.8714	2.1497	0.9899	0.6919
	$\hat{\varsigma}_1$	6.4471	4.8201	3.6284	1.6319	1.1317
95%	$\hat{\pi}_0$	2.2363	1.6474	1.2785	0.6102	0.4307
	$\hat{\pi}_1$	4.5865	3.5305	2.7660	1.2573	0.8835
	$\hat{\gamma}_0$	4.1465	3.1095	2.3864	1.1365	0.8020
	$\hat{\gamma}_1$	7.5945	5.8684	4.5647	2.1050	1.4761
	$\hat{\varsigma}_0$	2.9691	2.1848	1.6357	0.7533	0.5265
	$\hat{\varsigma}_1$	4.9057	3.6676	2.7609	1.2418	0.8611
90%	$\hat{\pi}_0$	1.8767	1.3825	1.0730	0.5121	0.3614
	$\hat{\pi}_1$	3.8491	2.9629	2.3213	1.0552	0.7414
	$\hat{\gamma}_0$	3.4798	2.6096	2.0027	0.9538	0.6731
	$\hat{\gamma}_1$	6.3735	4.9249	3.8308	1.7666	1.2388
	$\hat{\varsigma}_0$	2.4918	1.8336	1.3727	0.6321	0.4418
	$\hat{\varsigma}_1$	4.1170	3.0780	2.3170	1.0421	0.7227

Observa-se que, mais uma vez, os estimadores de máxima verossimilhança dos parâmetros do modelo com dispersão variável têm bom desempenho, conforme pode ser visto na Tabela 4.7. As medidas de viés, variância e EQM tendem para zero à medida em que  $n$  vai aumentando, assim como as médias das estimativas convergem para os valores verdadeiros dos parâmetros.

**Tabela 4.9:** Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero no cenário 3;  $\pi_0 = -0.5, \pi_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$  e  $\varsigma_1 = 1.5$ .

$1 - \delta$	$\hat{\theta}$	$n$				
		40	60	100	500	1000
99%	$\hat{\pi}_0$	98.85	98.94	98.93	98.90	98.99
		(1.15; 0.00)	(0.97; 0.09)	(0.93; 0.14)	(0.70; 0.40)	(0.58; 0.43)
	$\hat{\pi}_1$	99.47	99.28	99.12	99.02	98.98
		(0.38; 0.15)	(0.55; 0.17)	(0.57; 0.31)	(0.57; 0.41)	(0.59; 0.43)
	$\hat{\gamma}_0$	97.39	98.23	98.75	98.92	99.04
		(0.71; 1.90)	(0.53; 1.24)	(0.30; 0.95)	(0.47; 0.61)	(0.33; 0.57)
	$\hat{\gamma}_1$	98.48	98.60	99.07	98.94	99.01
		(1.13; 0.39)	(1.07; 0.33)	(0.64; 0.29)	(0.53; 0.53)	(0.54; 0.40)
	$\hat{\varsigma}_0$	94.70	96.22	97.26	97.84	97.69
		(0.19; 5.11)	(0.20; 3.58)	(0.36; 2.38)	(0.73; 1.43)	(1.00; 1.31)
	$\hat{\varsigma}_1$	96.73	97.92	98.24	97.65	97.32
		(2.50; 0.77)	(1.36; 0.72)	(0.95; 0.81)	(1.18; 1.17)	(1.22; 1.46)
95%	$\hat{\pi}_0$	95.38	95.12	95.27	95.25	95.15
		(3.77; 0.85)	(3.61; 1.27)	(3.17; 1.56)	(2.82; 1.93)	(2.88; 1.97)
	$\hat{\pi}_1$	95.47	95.59	95.34	95.25	95.12
		(2.52; 2.01)	(2.47; 1.94)	(2.65; 2.01)	(2.38; 2.37)	(2.51; 2.37)
	$\hat{\gamma}_0$	92.57	93.13	94.02	94.63	95.02
		(2.59; 4.84)	(2.50; 4.37)	(2.26; 3.72)	(2.54; 2.83)	(2.27; 2.71)
	$\hat{\gamma}_1$	93.59	94.28	94.41	94.53	94.85
		(4.19; 2.22)	(3.53; 2.19)	(3.38; 2.21)	(2.80; 2.67)	(2.93; 2.22)
	$\hat{\varsigma}_0$	88.84	91.10	92.55	93.12	93.05
		(0.65; 10.51)	(0.79; 8.11)	(1.17; 6.28)	(2.36; 4.52)	(3.04; 3.91)
	$\hat{\varsigma}_1$	91.24	92.82	93.44	93.25	92.57
		(5.83; 2.93)	(4.21; 2.97)	(3.67; 2.89)	(3.64; 3.11)	(3.48; 3.95)
90%	$\hat{\pi}_0$	89.83	90.05	90.43	90.31	90.51
		(7.21; 2.96)	(6.53; 3.42)	(5.72; 3.85)	(5.41; 4.28)	(5.31; 4.18)
	$\hat{\pi}_1$	90.05	90.51	90.35	90.81	90.05
		(5.15; 4.80)	(4.89; 4.60)	(5.13; 4.52)	(4.67; 4.52)	(5.08; 4.87)
	$\hat{\gamma}_0$	86.85	88.06	88.81	89.34	89.89
		(5.57; 7.58)	(5.02; 6.92)	(4.80; 6.39)	(5.07; 5.59)	(4.64; 5.47)
	$\hat{\gamma}_1$	87.85	88.69	89.12	89.36	89.94
		(7.39; 4.76)	(6.43; 4.88)	(6.11; 4.77)	(5.23; 5.41)	(5.40; 4.66)
	$\hat{\varsigma}_0$	83.59	85.63	87.28	87.81	87.98
		(1.89; 14.52)	(2.09; 12.28)	(2.94; 9.78)	(4.51; 7.68)	(5.20; 6.82)
	$\hat{\varsigma}_1$	85.30	87.41	88.25	88.02	87.24
		(9.28; 5.42)	(7.50; 5.09)	(6.32; 5.43)	(6.31; 5.67)	(6.14; 6.62)

Para os parâmetros dos modelos, no terceiro cenário, também foram construídos intervalos considerando como níveis de confiança 90%, 95% e 99%. As amplitudes médias desses intervalos foram calculadas e estão reportadas na Tabela 4.8. Note que à medida em que o tamanho amostral vai aumentando, as amplitudes médias diminuem, o que indica

maior precisão dos intervalos de confiança.

Com base nos intervalos de confiança, construídos nesse cenário, foram calculadas as taxas de cobertura e as taxas de não cobertura à esquerda e à direita, cujos resultados estão expostos na Tabela 4.9. Analisando esses resultados nota-se que as taxas de cobertura convergem para os níveis de confiança adotados, conforme o tamanho da amostra aumenta. No que diz respeito às taxas de não cobertura, é perceptível que estas tornam-se menos assimétricas à medida em que  $n$  aumenta, uma vez que os intervalos assintóticos são baseados na distribuição normal.

Adicionalmente às estimações pontual e intervalar, também foram testadas hipóteses sobre os parâmetros, mais especificamente sobre o coeficiente de regressão do submodelo do parâmetro de dispersão,  $\tau$ . Considerando o modelo com dispersão variável, especialmente o submodelo  $d(\tau_i) = \varsigma_0 + q_{i1}\varsigma_1$ , testou-se  $H_0 : \varsigma_1 = \varsigma_1^{(0)}$  vs.  $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , em que  $\varsigma_1^{(0)} = 0$ . As taxas de rejeição nulas estão apresentadas na Tabela 4.10 e as taxas de rejeição não nulas estão expostas na Tabela 4.11. Para as taxas de rejeição nulas os dados foram gerados sob  $H_0$ , em que  $\varsigma_1 = 0$ , e para as taxas de rejeição não nulas, ou poderes dos testes, os dados foram gerados sob  $H_1$ , em que  $\varsigma_1 = (0.5, 1.5)$ .

Foram avaliados os desempenhos de dois testes de hipóteses: o teste da razão de verossimilhanças e o teste Z. Os valores críticos foram os quantis das distribuições qui-quadrado com um grau de liberdade e normal padrão, respectivamente para cada teste.

Na simulação de tamanho, nota-se que as taxas de rejeição nulas convergem para os respectivos níveis nominais utilizados conforme  $n$  aumenta, como é possível verificar analisando a Tabela 4.10.

**Tabela 4.10:** Taxas (%) de rejeição nulas dos testes, considerando  $H_0 : \varsigma_1 = \varsigma_1^{(0)}$  vs.  $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , para o modelo Kumaraswamy inflacionado em zero com dispersão variável.

Nível nominal	Estatística de teste	$n$				
		40	60	100	500	1000
1%	$RV$	1.76	1.44	1.25	1.02	1.07
	$Z$	3.10	2.40	2.00	1.08	1.20
5%	$RV$	7.35	6.27	5.86	5.38	5.06
	$Z$	9.29	7.86	7.11	5.52	5.25
10%	$RV$	13.25	11.90	11.22	10.69	10.13
	$Z$	15.19	13.40	12.51	10.93	10.38

Para as simulações de poder foi considerado que  $\varsigma_1^{(0)} = 0$  e os dados foram gerados sob dois cenários: o primeiro com  $\varsigma_1 = 0.5$  e o segundo com  $\varsigma_1 = 1.5$ , a fim de verificar como os poderes dos testes se comportam com valores mais próximos do valor hipotetizado e com valores mais distantes. Os valores críticos usados nesse caso foram retirados das respectivas simulações de tamanho, a fim de comparar os poderes dos testes de mesmo tamanho.

Diante disso, note primeiramente que à medida em que o valor de  $\varsigma_1$  se distancia do valor hipotetizado em  $H_0$ , os poderes dos testes atingem 100% mais rapidamente. Quando

$n$  aumenta, as taxas de rejeição não nulas também aumentam, convergindo para 100%, como era esperado é observado na Tabela 4.11.

**Tabela 4.11:** Taxas (%) de rejeição não nulas dos testes, considerando  $H_0 : \varsigma_1 = \varsigma_1^{(0)}$  vs.  $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , para o modelo Kumaraswamy inflacionado em zero com dispersão variável.

$\varsigma_1$	Nível nominal	Estatística de teste	$n$				
			40	60	100	500	1000
0.50	1%	$RV$	1.58	2.02	2.84	17.22	40.15
		$Z$	2.13	3.15	4.01	24.34	48.97
	5%	$RV$	6.79	7.71	9.64	36.34	65.67
		$Z$	2.13	11.18	14.66	49.07	75.92
	10%	$RV$	12.67	13.98	16.83	49.17	76.17
		$Z$	17.71	20.08	25.96	63.13	85.41
1.50	1%	$RV$	5.22	8.90	18.55	96.13	99.98
		$Z$	7.53	13.37	24.21	97.39	99.33
	5%	$RV$	17.60	23.18	39.85	99.07	100.00
		$Z$	25.78	33.24	51.48	99.44	99.52
	10%	$RV$	27.26	35.04	53.40	99.60	100.00
		$Z$	38.39	47.40	66.94	99.79	99.78

Os resultados de simulação apresentados indicam que, de forma geral, os testes apresentam bom desempenho em amostras finitas. Ainda é possível destacar que o teste Z, apesar de ter uma convergência relativamente mais lenta nas simulações de tamanho, tem poderes mais altos, apresentando melhor desempenho do que o teste da razão de verossimilhanças.

## 4.5 Modelo de regressão Kumaraswamy inflacionado em zero e um

Nesta seção será introduzido o modelo de regressão Kumaraswamy inflacionado em zero e um, em que a variável resposta possui distribuição Kumaraswamy inflacionada em zero e em um, descrita no Capítulo 3.

Seja  $\mathbf{y} = (y_1, \dots, y_n)$  uma amostra independente da variável  $Y \sim \text{KIZU}(y; \lambda, p, \omega, \tau)$ , com densidade reparametrizada dada por

$$\text{kizu}(y; \lambda, p, \omega, \tau) = \begin{cases} \lambda p, & \text{se } y = 1, \\ \lambda(1 - p), & \text{se } y = 0, \\ (1 - \lambda)g(y; \omega, \tau), & \text{se } y \in (0, 1), \end{cases} \quad (4.11)$$

em que  $0 < \lambda < 1$  é o parâmetro de mistura,  $0 < p < 1$  é o parâmetro da distribuição Bernoulli que representa a probabilidade de que  $Y = 1$  e  $g(y; \omega, \tau)$  é a densidade reparametrizada da distribuição Kumaraswamy dada em (4.5) com mediana  $0 < \omega < 1$  e parâmetro dispersão  $\tau > 0$ .

A densidade (4.11) pode ser escrita como

$$\begin{aligned} \text{kizu}(y; \lambda, p, \omega, \tau) &= [\lambda^{\mathbb{I}_{\{0,1\}}(y)}(1 - \lambda)^{1 - \mathbb{I}_{\{0,1\}}(y)}] \times [p^y(1 - p)^{1 - y}]^{\mathbb{I}_{\{0,1\}}(y)} \\ &\times [g(y; \omega, \tau)^{1 - \mathbb{I}_{\{0,1\}}(y)}] \end{aligned}$$

em que  $\mathbb{I}_{\{0,1\}}(y)$  é a função indicadora que é igual a um se  $y \in \{0, 1\}$  e é igual a zero se  $y \notin \{0, 1\}$ . Observa-se que a densidade fatora em três termos: o primeiro termo depende apenas de  $\lambda$ , o segundo termo depende somente de  $p$  e o terceiro termo envolve  $\omega$  e  $\tau$ .

Considere agora que  $\xi_1 = \lambda p = \Pr(Y = 1)$  e  $\xi_0 = \lambda(1 - p) = \Pr(Y = 0)$ . Então,  $\lambda = \xi_0 + \xi_1$  e  $p = \xi_1/\lambda$ . Dessa forma, a densidade expressa em (4.11) pode ser reescrita como

$$\text{kizu}(y; \xi_0, \xi_1, \omega, \tau) = \begin{cases} \xi_1, & \text{se } y = 1, \\ \xi_0, & \text{se } y = 0, \\ (1 - \xi_0 - \xi_1)g(y; \omega, \tau), & \text{se } y \in (0, 1), \end{cases}$$

ou simplesmente

$$\text{kizu}(y; \xi_0, \xi_1, \omega, \tau) = \xi_0^{\mathbb{I}_{\{0\}}(y)} \xi_1^{\mathbb{I}_{\{1\}}(y)} [(1 - \xi_0 - \xi_1)g(y; \omega, \tau)]^{1 - \mathbb{I}_{\{0\}}(y) - \mathbb{I}_{\{1\}}(y)}. \quad (4.12)$$

#### 4.5.1 Dispersão fixa

Com base na densidade (4.12), o modelo de regressão Kumaraswamy inflacionado em zero e um, considerando dispersão fixa, é dado por

$$\begin{cases} h^*(\xi_{0i}, \xi_{1i}) &= (h(\xi_{0i}, \xi_{1i}), l(\xi_{0i}, \xi_{1i})) = (\zeta_{0i}, \zeta_{1i}), \\ f(\omega_i) &= \sum_{t=1}^r x_{it}\gamma_t = \eta_i, \quad i = 1, \dots, n, \end{cases}$$

em que  $h^*$  é uma transformação bijetora de  $(\xi_{0i}, \xi_{1i})$  a  $\mathbb{R}^2$ ,  $f : (0, 1) \rightarrow \mathbb{R}$  é uma função de ligação duas vezes diferenciável e estritamente monótona,  $\xi_{0i} = \Pr(Y_i = 0)$ ,  $\xi_{1i} = \Pr(Y_i = 1)$ ,  $\zeta_{0i} = \sum_{j=1}^m z_{0ij}\varphi_j$ ,  $\zeta_{1i} = \sum_{k=1}^s z_{1ik}\nu_k$  e  $\eta_i$  são os preditores lineares,  $z_{0i} = (z_{0i1}, \dots, z_{0im})$ ,  $z_{1i} = (z_{1i1}, \dots, z_{1is})$  e  $x_i = (x_{i1}, \dots, x_{ir})$  são as variáveis regressoras conhecidas,  $\varphi = (\varphi_1, \dots, \varphi_m)^\top$ ,  $\nu = (\nu_1, \dots, \nu_s)^\top$  e  $\gamma = (\gamma_1, \dots, \gamma_r)^\top$  são vetores de parâmetros de regressão a serem estimados.

Considere que as funções de ligação são  $h(\xi_{0i}, \xi_{1i}) = \ln[\xi_{0i}/(1 - \xi_{0i} - \xi_{1i})] = \zeta_{0i}$ ,  $l(\xi_{0i}, \xi_{1i}) = \ln[\xi_{1i}/(1 - \xi_{0i} - \xi_{1i})] = \zeta_{1i}$  e  $f(\omega_i) = \ln[\omega_i/(1 - \omega_i)] = \eta_i$ . Tem-se assim que

$$\begin{cases} \xi_{0i}/(1 - \xi_{0i} - \xi_{1i}) = e^{\zeta_{0i}}, \\ \xi_{1i}/(1 - \xi_{0i} - \xi_{1i}) = e^{\zeta_{1i}}. \end{cases} \quad (4.13)$$

Resolvendo o sistema de equações em (4.13) tem-se que

$$\xi_{0i} = \frac{e^{\zeta_{0i}}}{1 + e^{\zeta_{0i}} + e^{\zeta_{1i}}} \quad \text{e} \quad \xi_{1i} = \frac{e^{\zeta_{1i}}}{1 + e^{\zeta_{0i}} + e^{\zeta_{1i}}}. \quad (4.14)$$

Adicionalmente,  $\omega_i = 1/(1 + e^{-\eta_i})$ .

Seja  $\boldsymbol{\theta} = (\varphi^\top, \nu^\top, \gamma^\top, \tau)^\top$  o vetor de parâmetros do modelo KIZU com dispersão fixa. A função de verossimilhança para  $\boldsymbol{\theta}$  com base na amostra independente  $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$  é

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^n \text{kizu}(y_i; \xi_0, \xi_1, \omega_i, \tau) = L_1(\varphi, \nu; \mathbf{y}) \times L_2(\gamma, \tau; \mathbf{y}),$$

em que

$$\begin{aligned}
L_1(\varphi, v; \mathbf{y}) &= \prod_{i=1}^n \xi_{0i}^{\mathbb{I}_{\{0\}}(y_i)} \xi_{1i}^{\mathbb{I}_{\{1\}}(y_i)} (1 - \xi_{0i} - \xi_{1i})^{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}, \\
L_2(\gamma, \tau; \mathbf{y}) &= \prod_{i=1}^n g(y_i; \omega_i, \tau)^{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)} \\
&= \prod_{i=1}^n \left[ \frac{\ln(0.5)}{\tau \ln(1 - \omega_i^{\tau-1})} y_i^{\tau-1-1} \left(1 - y_i^{\tau-1}\right)^{\frac{\ln(0.5)}{\ln(1 - \omega_i^{\tau-1})} - 1} \right]^{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}.
\end{aligned}$$

Assim, a função de log-verossimilhança Kumaraswamy inflacionada em zero e um pode ser escrita como

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \ell_1(\varphi, v; \mathbf{y}) + \ell_2(\gamma, \tau; \mathbf{y}),$$

em que

$$\begin{aligned}
\ell_1(\varphi, v; \mathbf{y}) &= \sum_{i=1}^n \left\{ \mathbb{I}_{\{0\}}(y_i) \ln(\xi_{0i}) + \mathbb{I}_{\{1\}}(y_i) \ln(\xi_{1i}) \right. \\
&\quad \left. + [1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)] \ln(1 - \xi_{0i} - \xi_{1i}) \right\}, \\
\ell_2(\gamma, \tau; \mathbf{y}) &= \sum_{i=1}^n [1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)] \ln[g(y_i; \omega_i, \tau)] \\
&= \sum_{i: y_i \in (0,1)} \left\{ \ln\left(\frac{1}{\tau}\right) + \left(\frac{1}{\tau} - 1\right) \ln(y_i) + \ln \left[ \frac{\ln(0.5)}{\ln(1 - \omega_i^{1/\tau})} \right] \right. \\
&\quad \left. + \left[ \frac{\ln(0.5)}{\ln(1 - \omega_i^{1/\tau})} - 1 \right] \ln(1 - y_i^{1/\tau}) \right\}.
\end{aligned}$$

Note que a função de log-verossimilhança fatora em duas partes, uma expressa em termos dos parâmetros  $\varphi$  e  $v$  e a outra que envolve  $\gamma$  e  $\tau$ , permitindo assim estimação independente destes parâmetros. Novamente utilizamos a convenção  $\ln(0) \times 0 = 0$ .

A fim de obter o vetor de primeiras derivadas, a função de log-verossimilhança é diferenciada em relação a cada parâmetro de interesse, como segue:

$$\begin{aligned}
U_{\varphi_j}(\varphi, v) &= \frac{\partial \ell_1(\varphi, v)}{\partial \varphi_j} = \frac{\partial \ell_1(\varphi, v)}{\partial \xi_{0i}} \frac{\partial \xi_{0i}}{\partial \varphi_j} + \frac{\partial \ell_1(\varphi, v)}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \varphi_j}, \quad j = 1, \dots, m, \\
U_{v_k}(\varphi, v) &= \frac{\partial \ell_1(\varphi, v)}{\partial v_k} = \frac{\partial \ell_1(\varphi, v)}{\partial \xi_{0i}} \frac{\partial \xi_{0i}}{\partial v_k} + \frac{\partial \ell_1(\varphi, v)}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial v_k}, \quad k = 1, \dots, s, \\
U_{\gamma_t}(\gamma, \tau) &= \frac{\partial \ell_2(\gamma, \tau)}{\partial \gamma_t} = \frac{\partial \ell_2(\gamma, \tau)}{\partial \omega_i} \frac{\partial \omega_i}{\partial \gamma_t}, \quad t = 1, \dots, r, \\
U_{\tau}(\gamma, \tau) &= \frac{\partial \ell_2(\gamma, \tau)}{\partial \tau}.
\end{aligned}$$

Com base nas expressões em (4.14), é possível obter

$$\begin{aligned}
\frac{\partial \xi_{0i}}{\partial \zeta_{0i}} &= \frac{e^{\xi_{0i}}(1 + e^{\xi_{1i}})}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})^2} = \frac{e^{\xi_{0i}}}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})} \cdot \frac{(1 + e^{\xi_{1i}})}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})} = \xi_{0i}(1 - \xi_{0i}), \\
\frac{\partial \xi_{1i}}{\partial \zeta_{1i}} &= \frac{e^{\xi_{1i}}(1 + e^{\xi_{0i}})}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})^2} = \frac{e^{\xi_{1i}}}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})} \cdot \frac{(1 + e^{\xi_{0i}})}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})} = \xi_{1i}(1 - \xi_{1i}), \\
\frac{\partial \xi_{1i}}{\partial \zeta_{0i}} &= -\frac{e^{\xi_{0i} + \xi_{1i}}}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})^2} = -\frac{e^{\xi_{0i}}}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})} \cdot \frac{e^{\xi_{1i}}}{(1 + e^{\xi_{0i}} + e^{\xi_{1i}})} = -\xi_{0i}\xi_{1i}.
\end{aligned}$$

Logo, a função escore é dada por

$$\begin{aligned}
U_{\varphi_j}(\varphi, v) &= \sum_{i=1}^n \left[ \frac{\mathbb{I}_{\{0\}}(y_i)}{\xi_{0i}} - \frac{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}{(1 - \xi_{0i} - \xi_{1i})} \right] \xi_{0i}(1 - \xi_{0i})z_{0ij} \\
&\quad - \sum_{i=1}^n \left[ \frac{\mathbb{I}_{\{1\}}(y_i)}{\xi_{1i}} - \frac{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}{(1 - \xi_{0i} - \xi_{1i})} \right] \xi_{0i}\xi_{1i}z_{0ij}, \\
U_{v_k}(\varphi, v) &= -\sum_{i=1}^n \left[ \frac{\mathbb{I}_{\{0\}}(y_i)}{\xi_{0i}} - \frac{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}{(1 - \xi_{0i} - \xi_{1i})} \right] \xi_{0i}\xi_{1i}z_{1ik} \\
&\quad + \sum_{i=1}^n \left[ \frac{\mathbb{I}_{\{1\}}(y_i)}{\xi_{1i}} - \frac{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}{(1 - \xi_{0i} - \xi_{1i})} \right] \xi_{1i}(1 - \xi_{1i})z_{1ik}, \\
U_{\gamma_t}(\gamma, \tau) &= \sum_{i=1}^n \left\{ \frac{x_{it}e^{\eta_i}\omega_i^{\tau-1} [1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)]}{\tau(1 + e^{\eta_i})^2} \right. \\
&\quad \times \left. \left[ \frac{\ln(1 - \omega_i^{\tau-1}) + \ln(0.5) \ln(1 - y_i^{\tau-1})}{(1 - \omega_i^{\tau-1}) \ln^2(1 - \omega_i^{\tau-1})} \right] \right\}, \\
U_{\tau}(\gamma, \tau) &= -\frac{1}{\tau} \sum_{i=1}^n [1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)] \\
&\quad - \frac{1}{\tau^2} \sum_{i=1}^n \ln(y_i) [1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)] \\
&\quad - \frac{1}{\tau^2} \sum_{i=1}^n \left\{ \omega_i^{\tau-1} \ln(\omega_i) \left[ \frac{\ln(1 - \omega_i^{\tau-1}) + \ln(0.5) \ln(1 - y_i^{\tau-1})}{(1 - \omega_i^{\tau-1}) \ln^2(1 - \omega_i^{\tau-1})} \right] \right. \\
&\quad \left. - \frac{y_i^{\tau-1} \ln(y_i) [\ln(0.5) - \ln(1 - \omega_i^{\tau-1})]}{(1 - y_i^{\tau-1}) \ln(1 - \omega_i^{\tau-1})} \right\} [1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)].
\end{aligned}$$

Note que

- Se  $y = 0$ , então

$$\frac{\partial \ell_1(\varphi, v)}{\partial \varphi_j} = \sum_{i=1}^n (1 - \xi_{0i})z_{0ij} \quad \text{e} \quad \frac{\partial \ell_1(\varphi, v)}{\partial v_k} = -\sum_{i=1}^n \xi_{1i}z_{1ik};$$



- Se  $y = 1$ , então

$$\frac{\partial \ell_1(\varphi, v)}{\partial \varphi_j} = -\sum_{i=1}^n \xi_{0i} z_{0ij} \quad \text{e} \quad \frac{\partial \ell_1(\varphi, v)}{\partial v_k} = \sum_{i=1}^n (1 - \xi_{1i}) z_{1ik};$$

- Se  $y \in (0, 1)$ , então

$$\frac{\partial \ell_1(\varphi, v)}{\partial \varphi_j} = -\sum_{i=1}^n \xi_{0i} z_{0ij} \quad \text{e} \quad \frac{\partial \ell_1(\varphi, v)}{\partial v_k} = -\sum_{i=1}^n \xi_{1i} z_{1ik}.$$

Os intervalos de confiança assintóticos para os parâmetros que indexam esse modelo são dados por

$$\hat{\varphi}_j \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\varphi}_j), \quad \hat{v}_k \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{v}_k) \quad \text{e} \quad \hat{\gamma}_t \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\gamma}_t),$$

em que  $\text{ep}(\cdot)$  é o erro-padrão de cada estimador, obtido pela raiz quadrada dos elementos da diagonal principal da inversa da matriz de informação, substituindo os parâmetros pelos estimadores de máxima verossimilhança;  $z_{(1-\frac{\delta}{2})}$  é o quantil  $1 - \frac{\delta}{2}$  da distribuição normal padrão e  $1 - \delta$  é o nível de confiança adotado.

#### 4.5.2 Dipersão variável

Considere agora o seguinte modelo:

$$\begin{cases} h^*(\xi_{0i}, \xi_{1i}) &= (h(\xi_{0i}, \xi_{1i}), l(\xi_{0i}, \xi_{1i})) = (\zeta_{0i}, \zeta_{1i}), \\ f(\omega_i) &= \sum_{t=1}^r x_{it} \gamma_t = \eta_i, \\ d(\tau_i) &= \sum_{b=1}^u q_{ib} \varsigma_b = \Upsilon_i, \end{cases}$$

em que  $i = 1, \dots, n$ ,  $h^*$  é uma transformação bijetora de  $(\xi_{0i}, \xi_{1i})$  a  $\mathbb{R}^2$ ,  $f : (0, 1) \rightarrow \mathbb{R}$  e  $d : (0, \infty) \rightarrow \mathbb{R}$  são funções de ligação duas vezes diferenciáveis e estritamente monótonas,  $\zeta_{0i} = \sum_{j=1}^m z_{0ij} \varphi_j$ ,  $\zeta_{1i} = \sum_{k=1}^s z_{1ik} v_k$ ,  $\eta_i$  e  $\Upsilon_i$  são os preditores lineares,  $z_{0i} = (z_{0i1}, \dots, z_{0im})$ ,  $z_{1i} = (z_{1i1}, \dots, z_{1is})$ ,  $x_i = (x_{i1}, \dots, x_{ir})$  e  $q_i = (q_{i1}, \dots, q_{iu})$  são variáveis independentes conhecidas,  $\varphi = (\varphi_1, \dots, \varphi_m)^\top$ ,  $v = (v_1, \dots, v_s)^\top$ ,  $\gamma = (\gamma_1, \dots, \gamma_r)^\top$  e  $\varsigma = (\varsigma_1, \dots, \varsigma_u)^\top$  são os vetores de parâmetros a serem estimados.

A função de verossimilhança para  $\boldsymbol{\theta} = (\varphi^\top, v^\top, \gamma^\top, \varsigma^\top)^\top$  com base na amostra independente  $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$  é

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^n \text{kizu}(y_i; \xi_0, \xi_1, \omega_i, \tau_i) = L_1(\varphi, v; \mathbf{y}) \times L_2(\gamma, \varsigma; \mathbf{y}),$$

em que

$$\begin{aligned} L_1(\varphi, v; \mathbf{y}) &= \prod_{i=1}^n \xi_{0i}^{\mathbb{I}_{\{0\}}(y_i)} \xi_{1i}^{\mathbb{I}_{\{1\}}(y_i)} (1 - \xi_{0i} - \xi_{1i})^{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}, \\ L_2(\gamma, \varsigma; \mathbf{y}) &= \prod_{i=1}^n g(y_i; \omega_i, \tau_i)^{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)} \\ &= \prod_{i=1}^n \left[ \frac{\ln(0.5)}{\tau_i \ln(1 - \omega_i^{\tau_i^{-1}})} y_i^{\tau_i^{-1} - 1} \left(1 - y_i^{\tau_i^{-1}}\right)^{\frac{\ln(0.5)}{\ln(1 - \omega_i^{\tau_i^{-1}})} - 1} \right]^{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}. \end{aligned}$$

A função de log-verossimilhança do modelo KIZU com dispersão variável é dada por

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \ell_1(\varphi, v; \mathbf{y}) + \ell_2(\gamma, \varsigma; \mathbf{y}),$$

em que

$$\begin{aligned} \ell_1(\varphi, v; \mathbf{y}) &= \sum_{i=1}^n \left\{ \mathbb{I}_{\{0\}}(y_i) \ln(\xi_{0i}) + \mathbb{I}_{\{1\}}(y_i) \ln(\xi_{1i}) \right. \\ &\quad \left. + [1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)] \ln(1 - \xi_{0i} - \xi_{1i}) \right\}, \end{aligned}$$

$$\begin{aligned} \ell_2(\gamma, \varsigma; \mathbf{y}) &= \sum_{i=1}^n [1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)] \ln[g(y_i; \omega_i, \tau_i)] \\ &= \sum_{i: y_i \in (0,1)} \left\{ \ln\left(\frac{1}{\tau_i}\right) + \left(\frac{1}{\tau_i} - 1\right) \ln(y_i) \ln\left[\frac{\ln(0.5)}{\ln(1 - \omega_i^{1/\tau_i})}\right] \right. \\ &\quad \left. + \left[\frac{\ln(0.5)}{\ln(1 - \omega_i^{1/\tau_i})} - 1\right] \ln(1 - y_i^{1/\tau_i}) \right\}. \end{aligned}$$

Como observado anteriormente, a função de log-verossimilhança fatora em duas partes, uma que depende apenas dos parâmetros  $\varphi$  e  $v$  e a outra que depende de  $\gamma$  e  $\varsigma$ . Aqui também é adotada a convenção  $\ln(0) \times 0 = 0$ .

Considerando as funções de ligação  $h(\xi_{0i}) = h(\xi_{0i}, \xi_{1i}) = \ln[\xi_{0i}/(1 - \xi_{0i} - \xi_{1i})] = \zeta_{0i}$ ,  $l(\xi_{1i}) = l(\xi_{0i}, \xi_{1i}) = \ln[\xi_{1i}/(1 - \xi_{0i} - \xi_{1i})] = \zeta_{1i}$ ,  $f(\omega_i) = \ln[\omega_i/(1 - \omega_i)] = \eta_i$  e  $d(\tau_i) = \Upsilon_i$ , é possível obter o vetor escore, cujas componentes são

$$\begin{aligned} U_{\varphi_j}(\varphi, v) &= \sum_{i=1}^n \left[ \frac{\mathbb{I}_{\{0\}}(y_i)}{\xi_{0i}} - \frac{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}{(1 - \xi_{0i} - \xi_{1i})} \right] \xi_{0i}(1 - \xi_{0i}) z_{0ij} \\ &\quad - \sum_{i=1}^n \left[ \frac{\mathbb{I}_{\{1\}}(y_i)}{\xi_{1i}} - \frac{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}{(1 - \xi_{0i} - \xi_{1i})} \right] \xi_{0i} \xi_{1i} z_{0ij}, \quad j = 1, \dots, m, \\ U_{v_k}(\varphi, v) &= - \sum_{i=1}^n \left[ \frac{\mathbb{I}_{\{0\}}(y_i)}{\xi_{0i}} - \frac{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}{(1 - \xi_{0i} - \xi_{1i})} \right] \xi_{0i} \xi_{1i} z_{1ik} \\ &\quad + \sum_{i=1}^n \left[ \frac{\mathbb{I}_{\{1\}}(y_i)}{\xi_{1i}} - \frac{1 - \mathbb{I}_{\{0\}}(y_i) - \mathbb{I}_{\{1\}}(y_i)}{(1 - \xi_{0i} - \xi_{1i})} \right] \xi_{1i}(1 - \xi_{1i}) z_{1ik}, \quad k = 1, \dots, s, \end{aligned}$$

$$\begin{aligned}
U_{\gamma_t}(\gamma, \varsigma) &= \sum_{i=1}^n \omega_i^{\tau_i^{-1}-1} \left[ \frac{\ln(1 - \omega_i^{\tau_i^{-1}}) + \ln(0.5) \ln(1 - y_i^{\tau_i^{-1}})}{\tau_i (1 - \omega_i^{\tau_i^{-1}}) \ln^2(1 - \omega_i^{\tau_i^{-1}})} \right] \\
&\times \frac{x_{it} e^{\eta_i} [1 - \Pi_{\{0\}}(y_i) - \Pi_{\{1\}}(y_i)]}{(1 + e^{\eta_i})^2}, \quad t = 1, \dots, r, \\
U_{\varsigma_b}(\gamma, \varsigma) &= \frac{\partial \ell_2(\gamma, \varsigma)}{\partial \varsigma_b} = \frac{\partial \ell_2(\gamma, \varsigma)}{\partial \tau_i} \frac{\partial \tau_i}{\partial \Upsilon_i} \frac{\partial \Upsilon_i}{\partial \varsigma_b} = - \sum_{i=1}^n \left\{ \frac{1}{\tau_i} + \frac{1}{\tau_i^2} \ln(y_i) \right. \\
&+ \frac{1}{\tau_i^2} \frac{\omega_i^{\tau_i^{-1}} \ln(\omega_i) \left[ \ln(1 - \omega_i^{\tau_i^{-1}}) + \ln(0.5) \ln(1 - y_i^{\tau_i^{-1}}) \right]}{(1 - \omega_i^{\tau_i^{-1}}) \ln^2(1 - \omega_i^{\tau_i^{-1}})} \\
&+ \left. \frac{1}{\tau_i^2} \frac{y_i^{\tau_i^{-1}} \ln(y_i) \left[ \ln(0.5) - \ln(1 - \omega_i^{\tau_i^{-1}}) \right]}{(1 - y_i^{\tau_i^{-1}}) \ln(1 - \omega_i^{\tau_i^{-1}})} \right\} \\
&\times [1 - \Pi_{\{0\}}(y_i) - \Pi_{\{1\}}(y_i)] e^{\Upsilon_i} q_{ib}, \quad b = 1, \dots, u.
\end{aligned}$$

Os intervalos de confiança assintóticos para os parâmetros que indexam esse modelo são dados por

$$\hat{\varphi}_j \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\varphi}_j), \quad \hat{v}_k \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{v}_k), \quad \hat{\gamma}_t \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\gamma}_t) \quad \text{e} \quad \hat{\varsigma}_b \pm z_{(1-\frac{\delta}{2})} \text{ep}(\hat{\varsigma}_b),$$

em que  $\text{ep}(\cdot)$  é o erro-padrão de cada estimador, obtido pela raiz quadrada dos elementos da diagonal principal da inversa da matriz de informação, substituindo os parâmetros pelos estimadores de máxima verossimilhança;  $z_{(1-\frac{\delta}{2})}$  é o quantil  $1 - \frac{\delta}{2}$  da distribuição normal padrão e  $1 - \delta$  é o nível de confiança adotado.

Além da estimação pontual e intervalar, também é possível testar hipóteses sobre os parâmetros desse modelo de forma análoga aos testes de hipóteses apresentados para o modelo de regressão Kumaraswamy inflacionado em  $c$  na Seção 4.3.3.

## 4.6 Resultados numéricos para o modelo KIZU

A fim de avaliar o desempenho dos EMVs dos parâmetros do modelo Kumaraswamy inflacionado em zero e um, foram realizadas simulações de Monte Carlo, com 10000 réplicas, considerando dois cenários: C1) modelo com dispersão fixa e C2) modelo com dispersão variável.

Para o primeiro cenário foram utilizadas as funções  $\xi_{0i} = e^{\zeta_{0i}} / (1 + e^{\zeta_{0i}} + e^{\zeta_{1i}})$ ,  $\xi_{1i} = e^{\zeta_{1i}} / (1 + e^{\zeta_{0i}} + e^{\zeta_{1i}})$  e  $\omega_i = 1 / (1 + e^{-\eta_i})$ . Já para o segundo cenário foram utilizadas as mesmas funções do cenário C1) adicionando  $\tau_i = e^{\Upsilon_i}$ , em que  $\zeta_0, \zeta_1, \eta$  e  $\Upsilon$  são os preditores lineares dos parâmetros do modelo. Também foram realizadas simulações de Monte Carlo para verificar o desempenho dos testes de hipóteses descritos na Seção 4.3.3.

Os tamanhos amostrais utilizados nas simulações foram  $n = (30, 50, 100, 200, 500)$  e as maximizações das funções de log-verossimilhança foram realizadas utilizando o método quasi-Newton BFGS (Nocedal, 1980) com gradiente analítico, utilizando a linguagem

**Tabela 4.12:** Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 1;  $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0$  e  $\tau = 2.0$ .

Medida	Estimador	$n$				
		40	60	100	500	1000
Média	$\hat{\varphi}_0$	-0.5345	-0.5385	-0.5167	-0.4994	-0.5011
	$\hat{\varphi}_1$	-2.0645	-1.5832	-1.5495	-1.5155	-1.5061
	$\hat{v}_0$	-0.5497	-0.5326	-0.5120	-0.4973	-0.5036
	$\hat{v}_1$	-1.7054	-1.5993	-1.5602	-1.5185	-1.5004
	$\hat{\gamma}_0$	1.0102	1.0012	0.9974	0.9984	1.0000
	$\hat{\gamma}_1$	-2.0139	-2.0054	-1.9996	-2.0008	-2.0011
	$\hat{\tau}$	1.8741	1.9205	1.9537	1.9913	1.9948
Variância	$\hat{\varphi}_0$	7.1711	0.8515	0.3908	0.0546	0.0266
	$\hat{\varphi}_1$	814.4086	2.7312	1.2704	0.1951	0.0938
	$\hat{v}_0$	1.5912	0.8789	0.3935	0.0555	0.0276
	$\hat{v}_1$	5.3077	2.8338	1.2873	0.1950	0.0957
	$\hat{\gamma}_0$	1.4818	0.9717	0.5093	0.0792	0.0379
	$\hat{\gamma}_1$	3.5153	2.2784	1.2166	0.1900	0.0918
	$\hat{\tau}$	0.3364	0.2301	0.1405	0.0285	0.0143
Viés	$\hat{\varphi}_0$	-0.0345	-0.0385	-0.0167	0.0006	-0.0011
	$\hat{\varphi}_1$	-0.5645	-0.0832	-0.0495	-0.0155	-0.0061
	$\hat{v}_0$	-0.0497	-0.0326	-0.0120	0.0027	-0.0036
	$\hat{v}_1$	-0.2054	-0.0993	-0.0602	-0.0185	-0.0004
	$\hat{\gamma}_0$	0.0102	0.0012	-0.0026	-0.0016	-0.0000
	$\hat{\gamma}_1$	-0.0139	-0.0054	0.0004	-0.0008	-0.0011
	$\hat{\tau}$	-0.1259	-0.0795	-0.0463	-0.0087	-0.0052
EQM	$\hat{\varphi}_0$	7.1723	0.8530	0.3911	0.0546	0.0266
	$\hat{\varphi}_1$	814.7273	2.7381	1.2728	0.1953	0.0939
	$\hat{v}_0$	1.5937	0.8800	0.3936	0.0556	0.0276
	$\hat{v}_1$	5.3499	2.8437	1.2909	0.1953	0.0957
	$\hat{\gamma}_0$	1.4819	0.9717	0.5093	0.0792	0.0379
	$\hat{\gamma}_1$	3.5155	2.2785	1.2166	0.1900	0.0918
	$\hat{\tau}$	0.2105	0.1505	0.0941	0.0199	0.0091

matricial de programação Ox. Os chutes iniciais utilizados nas simulações foram escolhidos arbitrariamente.

**C1)** No cenário com dispersão fixa, foi considerado o seguinte modelo:

$$\begin{cases} h(\xi_{0i}) = \varphi_0 + z_{0i1}\varphi_1 = \zeta_{0i}, \\ l(\xi_{1i}) = v_0 + z_{1i1}v_1 = \zeta_{1i}, \\ f(\omega_i) = \gamma_0 + x_{i1}\gamma_1 = \eta_i, \end{cases}$$

em que  $h(\xi_{0i}) = \ln[\xi_{0i}/(1-\xi_{0i}-\xi_{1i})]$ ,  $l(\xi_{1i}) = \ln[\xi_{1i}/(1-\xi_{0i}-\xi_{1i})]$  e  $f(\omega_i) = \ln[\omega_i/(1-\omega_i)]$  são as funções de ligação,  $z_{0i1}, z_{1i1}$  e  $x_{i1}$  são as variáveis regressoras conhecidas, cujos seus respectivos valores foram gerados aleatoriamente da distribuição uniforme padrão,  $\mathcal{U}(0, 1)$ ,  $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0$  e  $\gamma_1 = -2.0$  são os valores dos coeficientes de regressão impostos na simulação, e  $\tau = 2.0$  é o valor do parâmetro de dispersão, que neste caso é fixo.

As estimativas médias, variâncias, vieses e erros quadráticos médios (EQMs) para os estimadores dos parâmetros do modelo nesse cenário estão disponíveis na Tabela

**Tabela 4.13:** Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 1;  $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0$  e  $\tau = 2.0$ .

$1 - \delta$	$\hat{\theta}$	$n$				
		40	60	100	500	1000
99%	$\hat{\varphi}_0$	5.4458	4.4594	3.1053	1.2033	0.8486
	$\hat{\varphi}_1$	9.9520	7.9501	5.6151	2.2664	1.5910
	$\hat{v}_0$	7.8876	4.4591	3.1029	1.2028	0.8489
	$\hat{v}_1$	11.7346	7.9580	5.6152	2.2659	1.5907
	$\hat{\gamma}_0$	5.7917	4.8730	3.5939	1.4419	1.0118
	$\hat{\gamma}_1$	8.8886	7.4942	5.5585	2.2430	1.5629
	$\hat{\tau}$	2.8479	2.3776	1.8904	0.8757	0.6192
95%	$\hat{\varphi}_0$	4.1438	3.3932	2.3628	0.9156	0.6457
	$\hat{\varphi}_1$	7.5725	6.0493	4.2726	1.7245	1.2106
	$\hat{v}_0$	6.0017	3.3930	2.3610	0.9152	0.6459
	$\hat{v}_1$	8.9289	6.0553	4.2726	1.7241	1.2103
	$\hat{\gamma}_0$	4.4070	3.7079	2.7346	1.0971	0.7699
	$\hat{\gamma}_1$	6.7634	5.7024	4.2295	1.7067	1.1892
	$\hat{\tau}$	2.1670	1.8091	1.4384	0.6663	0.4711
90%	$\hat{\varphi}_0$	3.4776	2.8476	1.9829	0.7684	0.5419
	$\hat{\varphi}_1$	6.3551	5.0767	3.5856	1.4473	1.0160
	$\hat{v}_0$	5.0368	2.8475	1.9814	0.7681	0.5421
	$\hat{v}_1$	7.4934	5.0818	3.5857	1.4470	1.0158
	$\hat{\gamma}_0$	3.6984	3.1118	2.2949	0.9207	0.6461
	$\hat{\gamma}_1$	5.6760	4.7856	3.5495	1.4323	0.9980
	$\hat{\tau}$	1.8186	1.5183	1.2072	0.5592	0.3954

4.12. Analisando essa tabela é possível observar que os estimadores de máxima verossimilhança apresentam bom desempenho, uma vez que as médias das estimativas convergem para os valores impostos nas simulações à medida em que o tamanho amostral aumenta; bem como os EQMs dos estimadores diminuem à medida que  $n$  cresce.

Também foram construídos intervalos de confiança para os parâmetros do modelo inflacionado em zero e um com dispersão fixa. As amplitudes médias desses intervalos encontram-se na Tabela 4.13 onde é possível observar que os intervalos ficam mais precisos à medida em que  $n$  aumenta, uma vez que as amplitudes médias diminuem.

Na Tabela 4.14 estão apresentadas as taxas de cobertura e não cobertura à esquerda e à direita. Analisando esses resultados é possível perceber que os intervalos de confiança assintóticos para os parâmetros do modelo Kumaraswamy inflacionado em zero e um apresentam boa cobertura, uma vez que as taxas de cobertura convergem para os níveis de confiança utilizados à medida que  $n$  cresce. Quanto às taxas de não cobertura observa-se que estas apresentam comportamento menos assimétrico quando o tamanho amostral aumenta.

**Tabela 4.14:** Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 1;  $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0$  e  $\tau = 2.0$ .

$1 - \delta$	$\hat{\theta}$	$n$				
		40	60	100	500	1000
99%	$\hat{\varphi}_0$	99.65 (0.31; 0.04)	99.58 (0.39; 0.03)	99.21 (0.68; 0.11)	99.10 (0.61; 0.29)	99.04 (0.51; 0.45)
	$\hat{\varphi}_1$	99.84 (0.09; 0.07)	99.63 (0.21; 0.16)	99.28 (0.37; 0.35)	99.21 (0.45; 0.34)	99.04 (0.48; 0.48)
	$\hat{v}_0$	99.48 (0.47; 0.05)	99.43 (0.52; 0.05)	99.40 (0.52; 0.08)	98.99 (0.63; 0.38)	98.92 (0.55; 0.53)
	$\hat{v}_1$	99.83 (0.10; 0.07)	99.57 (0.25; 0.18)	99.32 (0.35; 0.33)	99.09 (0.42; 0.49)	98.84 (0.75; 0.41)
	$\hat{\gamma}_0$	97.90 (0.21; 1.89)	98.37 (0.16; 1.47)	98.59 (0.16; 1.25)	98.91 (0.29; 0.80)	99.13 (0.35; 0.52)
	$\hat{\gamma}_1$	98.02 (1.57; 0.41)	98.46 (1.23; 0.31)	98.81 (0.93; 0.26)	98.96 (0.71; 0.33)	99.03 (0.56; 0.41)
	$\hat{\tau}$	90.33 (0.04; 9.63)	93.49 (0.03; 6.48)	95.62 (0.04; 4.34)	98.32 (0.11; 1.57)	98.54 (0.26; 1.20)
	$\hat{\varphi}_0$	96.93 (2.64; 0.43)	95.31 (2.43; 1.26)	95.44 (2.82; 1.74)	95.27 (2.74; 1.99)	95.54 (2.22; 2.24)
	$\hat{\varphi}_1$	97.13 (1.63; 1.24)	95.96 (2.12; 1.92)	95.38 (2.30; 2.32)	95.01 (2.39; 2.60)	95.24 (2.48; 2.28)
	$\hat{v}_0$	96.59 (2.90; 0.51)	95.70 (3.04; 1.26)	95.58 (2.82; 1.60)	94.91 (2.95; 2.14)	94.98 (2.57; 2.45)
	$\hat{v}_1$	96.61 (1.88; 1.51)	95.50 (2.06; 2.44)	95.55 (1.99; 2.46)	95.06 (2.34; 2.60)	95.14 (2.58; 2.28)
	$\hat{\gamma}_0$	93.17 (1.59; 5.24)	94.21 (1.45; 4.34)	94.41 (1.53; 4.06)	94.60 (2.02; 3.38)	95.62 (1.94; 2.44)
	$\hat{\gamma}_1$	93.18 (4.62; 2.20)	94.19 (3.98; 1.83)	94.43 (3.54; 2.03)	95.02 (3.02; 1.96)	95.32 (2.42; 2.26)
	$\hat{\tau}$	84.68 (0.05; 15.27)	88.15 (0.07; 11.78)	90.81 (0.19; 9.00)	94.42 (1.08; 4.50)	94.51 (1.41; 4.08)
90%	$\hat{\varphi}_0$	91.55 (5.97; 2.48)	90.86 (5.42; 3.72)	90.43 (5.50; 4.07)	90.37 (5.17; 4.46)	90.47 (4.72; 4.81)
	$\hat{\varphi}_1$	91.62 (4.18; 4.20)	90.12 (5.05; 4.83)	90.29 (4.77; 4.94)	90.08 (4.70; 5.22)	90.44 (4.85; 4.71)
	$\hat{v}_0$	91.63 (5.81; 2.56)	90.62 (6.04; 3.34)	90.59 (5.62; 3.79)	90.11 (5.56; 4.33)	89.88 (5.16; 4.96)
	$\hat{v}_1$	91.03 (4.42; 4.55)	90.05 (4.60; 5.35)	90.05 (4.57; 5.38)	90.34 (4.45; 5.21)	89.83 (5.03; 5.14)
	$\hat{\gamma}_0$	87.84 (3.89; 8.27)	88.87 (3.66; 7.47)	89.46 (3.73; 6.81)	89.93 (4.10; 5.97)	90.24 (4.51; 5.25)
	$\hat{\gamma}_1$	87.47 (7.86; 4.67)	88.99 (6.83; 4.18)	89.72 (6.15; 4.13)	89.96 (5.53; 4.51)	90.13 (4.97; 4.90)
	$\hat{\tau}$	80.56 (0.12; 19.32)	83.64 (0.52; 15.84)	86.00 (0.88; 13.12)	89.62 (2.69; 7.69)	89.88 (3.13; 6.99)

**C2)** Para o modelo com dispersão variável, foi considerado o seguinte modelo:

$$\begin{cases} h(\xi_{0i}) = \varphi_0 + z_{0i1}\varphi_1 = \zeta_{0i}, \\ l(\xi_{1i}) = v_0 + z_{1i1}v_1 = \zeta_{1i}, \\ f(\omega_i) = \gamma_0 + x_{i1}\gamma_1 = \eta_i, \\ d(\tau_i) = \varsigma_0 + q_{i1}\varsigma_1 = \Upsilon_i, \end{cases}$$

**Tabela 4.15:** Médias, variâncias, vieses e EQMs dos estimadores de máxima verossimilhança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 2;  $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$  e  $\varsigma_1 = 1.5$ .

Medida	Estimador	$n$				
		40	60	100	500	1000
Média	$\hat{\varphi}_0$	-0.5311	-0.5383	-0.5166	-0.4993	-0.5010
	$\hat{\varphi}_1$	-2.0684	-1.5835	-1.5498	-1.5160	-1.5064
	$\hat{v}_0$	-0.5487	-0.5322	-0.5119	-0.4971	-0.5035
	$\hat{v}_1$	-1.7073	-1.5998	-1.5603	-1.5188	-1.5008
	$\hat{\gamma}_0$	1.0339	1.0285	1.0140	1.0013	1.0011
	$\hat{\gamma}_1$	-2.0574	-2.0511	-2.0268	-2.0067	-2.0036
	$\hat{\varsigma}_0$	0.2141	0.2875	0.3844	0.4807	0.4896
	$\hat{\varsigma}_1$	1.6683	1.6560	1.5752	1.5099	1.5053
Variância	$\hat{\varphi}_0$	7.1897	0.8516	0.3907	0.0547	0.0267
	$\hat{\varphi}_1$	837.5189	2.7311	1.2699	0.1955	0.0940
	$\hat{v}_0$	1.5794	0.8791	0.3935	0.0556	0.0276
	$\hat{v}_1$	5.3091	2.8348	1.2876	0.1952	0.0958
	$\hat{\gamma}_0$	2.4130	1.5471	0.7308	0.1064	0.0524
	$\hat{\gamma}_1$	7.4376	4.5887	2.2678	0.3517	0.1765
	$\hat{\varsigma}_0$	1.3221	0.8005	0.3801	0.0503	0.0240
	$\hat{\varsigma}_1$	3.5541	2.1163	1.0154	0.1379	0.0649
Viés	$\hat{\varphi}_0$	-0.0311	-0.0383	-0.0166	0.0007	-0.0010
	$\hat{\varphi}_1$	-0.5684	-0.0835	-0.0498	-0.0160	-0.0064
	$\hat{v}_0$	-0.0487	-0.0322	-0.0119	0.0029	-0.0035
	$\hat{v}_1$	-0.2073	-0.0998	-0.0603	-0.0188	-0.0008
	$\hat{\gamma}_0$	0.0339	0.0285	0.0140	0.0013	0.0011
	$\hat{\gamma}_1$	-0.0574	-0.0511	-0.0268	-0.0067	-0.0036
	$\hat{\varsigma}_0$	-0.2859	-0.2125	-0.1156	-0.0193	-0.0104
	$\hat{\varsigma}_1$	0.1683	0.1560	0.0752	0.0099	0.0053
EQM	$\hat{\varphi}_0$	7.1906	0.8530	0.3910	0.0547	0.0267
	$\hat{\varphi}_1$	837.8420	2.7380	1.2724	0.1958	0.0941
	$\hat{v}_0$	1.5817	0.8802	0.3936	0.0556	0.0276
	$\hat{v}_1$	5.3521	2.8447	1.2913	0.1955	0.0958
	$\hat{\gamma}_0$	2.4142	1.5479	0.7310	0.1064	0.0524
	$\hat{\gamma}_1$	7.4409	4.5914	2.2686	0.3517	0.1765
	$\hat{\varsigma}_0$	1.0362	0.2875	0.2645	0.0310	0.0135
	$\hat{\varsigma}_1$	3.7224	2.2723	1.0905	0.1478	0.0702

em que  $h(\xi_{0i}) = \ln[\xi_{0i}/(1 - \xi_{0i} - \xi_{1i})]$ ,  $l(\xi_{1i}) = \ln[\xi_{1i}/(1 - \xi_{0i} - \xi_{1i})]$ ,  $f(\omega_i) = \ln[\omega_i/(1 - \omega_i)]$  e  $d(\tau_i) = \ln(\tau_i)$  são funções de ligação,  $z_{0i1}, z_{1i1}, x_{i1}$  e  $q_{i1}$  são as variáveis regressoras conhecidas, cujos seus valores foram gerados aleatoriamente da distribuição  $\mathcal{U}(0, 1)$ ,  $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$  e  $\varsigma_1 = 1.5$  são os coeficientes de regressão impostos na simulação. Note que nesse caso a dispersão também é variável.

Na Tabela 4.15 encontram-se as estimativas médias, variâncias, vieses e EQM's dos estimadores dos parâmetros do modelo KIZU com dispersão variável. Note que as estimativas médias dos estimadores de máxima verossimilhança convergem para os verdadeiros valores à medida em que  $n$  aumenta e as variâncias, os vieses e os EQMs dos estimadores diminuem à medida que  $n$  cresce, como esperado.

**Tabela 4.16:** Amplitudes médias dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 2;  $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$  e  $\varsigma_1 = 1.5$ .

$1 - \delta$	$\hat{\theta}$	$n$				
		40	60	100	500	1000
99%	$\hat{\varphi}_0$	5.4465	4.4589	3.1052	1.2033	0.8486
	$\hat{\varphi}_1$	9.9529	7.9497	5.6153	2.2665	1.5911
	$\hat{v}_0$	6.8542	4.4596	3.1029	1.2027	0.8489
	$\hat{v}_1$	11.4193	7.9599	5.6150	2.2658	1.5907
	$\hat{\gamma}_0$	7.0087	5.8975	4.2268	1.6709	1.1776
	$\hat{\gamma}_1$	12.4938	10.3667	7.5004	3.0610	2.1482
	$\hat{\varsigma}_0$	4.9372	4.0703	2.9045	1.1108	0.7728
	$\hat{\varsigma}_1$	8.1263	6.6417	4.7559	1.8332	1.2642
95%	$\hat{\varphi}_0$	4.1443	3.3928	2.3627	0.9156	0.6457
	$\hat{\varphi}_1$	7.5732	6.0490	4.2727	1.7246	1.2107
	$\hat{v}_0$	5.2154	3.3934	2.3610	0.9152	0.6459
	$\hat{v}_1$	8.6890	6.0567	4.2725	1.7240	1.2103
	$\hat{\gamma}_0$	5.3330	4.4874	3.2162	1.2714	0.8961
	$\hat{\gamma}_1$	9.5066	7.8881	5.7071	2.3292	1.6346
	$\hat{\varsigma}_0$	3.7567	3.0971	2.2100	0.8452	0.5880
	$\hat{\varsigma}_1$	6.1833	5.0537	3.6188	1.3949	0.9619
90%	$\hat{\varphi}_0$	3.4780	2.8474	1.9829	0.7684	0.5419
	$\hat{\varphi}_1$	6.3556	5.0765	3.5857	1.4473	1.0160
	$\hat{v}_0$	4.3769	2.8478	1.9814	0.7680	0.5421
	$\hat{v}_1$	7.2920	5.0830	3.5856	1.4468	1.0157
	$\hat{\gamma}_0$	4.4756	3.7660	2.6991	1.0670	0.7520
	$\hat{\gamma}_1$	7.9782	6.6199	4.7895	1.9547	1.3718
	$\hat{\varsigma}_0$	3.1527	2.5992	1.8547	0.7093	0.4935
	$\hat{\varsigma}_1$	5.1892	4.2412	3.0370	1.1707	0.8073

As amplitudes médias dos intervalos de confiança para os parâmetros do modelo Kumaraswamy inflacionado em zero e um com dispersão variável estão expostas na Tabela 4.16. Nota-se que a estimação intervalar nesse cenário também apresenta bom desempenho uma vez que os intervalos ficam mais precisos à medida que o tamanho da amostra aumenta.

Confirmando o bom desempenho dos intervalos de confiança para os parâmetros do modelo Kumaraswamy inflacionado em zero e um com dispersão variável, é possível averiguar na Tabela 4.17 as taxas de cobertura e não cobertura desses intervalos, que apresentam bom desempenho, visto que suas taxas de cobertura convergem para os respectivos níveis de confiança adotados e as taxas de não cobertura ficam menos assimétricas quando aumenta o tamanho amostral.

Também foram realizados testes de hipóteses sobre o parâmetro do coeficiente de regressão do submodelo da dispersão. Considerando especificamente o submodelo  $d(\tau_i) = \varsigma_0 + q_{i1}\varsigma_1$ , testou-se  $H_0 : \varsigma_1 = \varsigma_1^{(0)}$  vs.  $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , em que  $\varsigma_1^{(0)} = 0$ . Note que sob a hipótese nula há dispersão fixa.

As taxas de rejeição nulas nesse cenário estão apresentadas na Tabela 4.18 e as taxas de rejeição não nulas estão na Tabela 4.19. Como valores críticos do teste da razão de



**Tabela 4.17:** Taxas (%) de cobertura e de não cobertura (à esquerda; à direita) dos intervalos de confiança dos parâmetros do modelo Kumaraswamy inflacionado em zero e um no cenário 2;  $\varphi_0 = -0.5, \varphi_1 = -1.5, v_0 = -0.5, v_1 = -1.5, \gamma_0 = 1.0, \gamma_1 = -2.0, \varsigma_0 = 0.5$  e  $\varsigma_1 = 1.5$ .

$1 - \delta$	$\hat{\theta}$	$n$				
		40	60	100	500	1000
99%	$\hat{\varphi}_0$	99.66 (0.31; 0.03)	99.57 (0.40; 0.03)	99.21 (0.68; 0.11)	99.10 (0.61; 0.29)	99.03 (0.51; 0.46)
	$\hat{\varphi}_1$	99.85 (0.09; 0.06)	99.63 (0.21; 0.16)	99.28 (0.37; 0.35)	99.20 (0.45; 0.35)	99.03 (0.47; 0.52)
	$\hat{v}_0$	99.49 (0.47; 0.04)	99.43 (0.52; 0.05)	99.41 (0.51; 0.08)	98.98 (0.63; 0.39)	98.90 (0.48; 0.49)
	$\hat{v}_1$	99.84 (0.09; 0.07)	99.57 (0.25; 0.18)	99.32 (0.35; 0.33)	99.08 (0.43; 0.49)	98.81 (0.57; 0.53)
	$\hat{\gamma}_0$	96.83 (0.85; 2.32)	97.78 (0.62; 1.60)	98.22 (0.49; 1.29)	98.80 (0.47; 0.73)	99.09 (0.39; 0.52)
	$\hat{\gamma}_1$	98.09 (1.45; 0.46)	98.71 (0.96; 0.33)	98.79 (0.93; 0.28)	99.09 (0.57; 0.34)	98.90 (0.59; 0.51)
	$\hat{\varsigma}_0$	93.98 (0.25; 5.77)	95.35 (0.46; 4.19)	96.31 (0.41; 3.28)	97.61 (0.64; 1.75)	97.73 (0.79; 1.48)
	$\hat{\varsigma}_1$	96.21 (2.82; 0.97)	96.80 (2.16; 1.04)	97.21 (1.93; 0.86)	97.51 (1.47; 1.02)	97.41 (1.45; 1.14)
	$\hat{\varphi}_0$	96.96 (2.63; 0.41)	96.29 (2.45; 1.26)	95.46 (2.82; 1.72)	95.26 (2.75; 1.99)	95.52 (2.27; 2.21)
	$\hat{\varphi}_1$	97.15 (1.61; 1.24)	95.93 (2.12; 1.95)	95.40 (2.28; 2.32)	95.01 (2.38; 2.61)	95.22 (2.49; 2.29)
	$\hat{v}_0$	96.59 (2.91; 0.50)	95.70 (3.04; 1.26)	95.61 (2.80; 1.50)	94.88 (2.97; 2.15)	94.96 (2.59; 2.45)
	$\hat{v}_1$	96.62 (1.87; 1.51)	95.49 (2.07; 2.44)	95.58 (1.98; 2.44)	95.02 (2.36; 2.62)	95.08 (2.60; 2.32)
95%	$\hat{\gamma}_0$	91.66 (2.95; 5.39)	93.31 (2.55; 4.14)	93.92 (2.33; 3.75)	94.67 (2.20; 3.13)	94.83 (2.53; 2.55)
	$\hat{\gamma}_1$	93.24 (4.47; 2.29)	94.31 (3.47; 2.22)	94.74 (3.10; 2.16)	95.06 (2.71; 2.23)	94.98 (2.51; 2.38)
	$\hat{\varsigma}_0$	88.27 (0.82; 10.91)	90.01 (1.20; 8.79)	91.22 (1.24; 7.54)	93.40 (2.04; 4.56)	93.83 (2.07; 4.10)
	$\hat{\varsigma}_1$	90.27 (6.59; 3.14)	91.75 (5.50; 2.75)	92.24 (5.00; 2.76)	93.17 (3.82; 3.01)	93.62 (2.97; 3.41)
90%	$\hat{\varphi}_0$	91.60 (5.95; 2.45)	90.84 (5.44; 3.72)	90.46 (5.49; 4.05)	90.36 (5.19; 4.45)	90.43 (4.78; 4.79)
	$\hat{\varphi}_1$	91.68 (4.15; 4.17)	90.08 (5.06; 4.86)	90.28 (4.78; 4.94)	90.04 (4.71; 5.25)	90.43 (4.87; 4.70)
	$\hat{v}_0$	91.63 (5.81; 2.56)	90.63 (6.04; 3.33)	90.60 (5.61; 3.79)	90.08 (5.60; 5.32)	89.90 (5.24; 4.94)
	$\hat{v}_1$	91.01 (4.42; 4.57)	90.03 (4.61; 5.36)	90.06 (4.57; 5.37)	90.31 (4.46; 5.23)	89.78 (5.07; 5.15)
	$\hat{\gamma}_0$	86.08 (5.72; 8.20)	87.72 (5.03; 7.25)	88.39 (4.98; 6.63)	89.71 (4.58; 5.71)	89.74 (4.89; 5.37)
	$\hat{\gamma}_1$	87.30 (7.45; 5.25)	88.79 (6.53; 4.68)	89.50 (5.98; 4.52)	90.10 (5.37; 4.53)	89.77 (5.41; 4.82)
	$\hat{\varsigma}_0$	82.86 (2.19; 14.95)	84.56 (2.49; 12.95)	85.92 (2.99; 11.09)	88.04 (4.45; 7.51)	88.94 (4.87; 6.19)
	$\hat{\varsigma}_1$	84.37 (10.03; 5.60)	85.64 (9.34; 5.02)	86.82 (8.05; 5.13)	87.82 (6.69; 5.49)	89.46 (5.60; 4.94)

**Tabela 4.18:** Taxas (%) de rejeição nulas dos testes, considerando  $H_0 : \varsigma_1 = \varsigma_1^{(0)}$  vs.  $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , para o modelo Kumaraswamy inflacionada em zero e um com dispersão variável.

Nível nominal	Estatística de teste	$n$				
		50	60	100	500	1000
1%	$RV$	1.66	1.56	1.42	1.16	1.02
	$Z$	3.26	2.62	2.24	1.26	1.23
5%	$RV$	7.09	6.42	6.27	5.05	4.96
	$Z$	9.38	8.56	7.61	5.42	5.63
10%	$RV$	13.33	11.95	11.72	10.27	10.51
	$Z$	14.78	14.26	13.15	10.73	10.60

**Tabela 4.19:** Taxas (%) de rejeição não nulas dos testes, considerando  $H_0 : \varsigma_1 = \varsigma_1^{(0)}$  vs.  $H_1 : \varsigma_1 \neq \varsigma_1^{(0)}$ , para o modelo Kumaraswamy inflacionada em zero e um com dispersão variável.

$\varsigma_1$	Nível nominal	Estatística de teste	$n$				
			50	60	100	500	1000
0.05	1%	$RV$	1.85	1.65	2.00	13.19	33.44
		$Z$	2.35	2.68	3.42	19.31	39.68
	5%	$RV$	6.72	7.36	9.10	32.61	58.34
		$Z$	10.11	11.28	13.75	43.52	68.45
	10%	$RV$	13.01	13.70	15.52	44.10	69.47
		$Z$	18.49	18.96	23.21	58.16	81.16
1.50	1%	$RV$	6.16	6.77	13.03	91.45	99.97
		$Z$	8.01	10.44	19.17	94.81	99.45
	5%	$RV$	17.34	20.55	33.42	98.05	100.00
		$Z$	24.91	29.72	44.91	98.86	99.64
	10%	$RV$	27.18	31.57	46.01	99.04	100.00
		$Z$	38.08	43.45	59.50	99.53	99.89

verossimilhanças foram utilizados o quantil de referência da distribuição qui-quadrado com um grau de liberdade, já para o teste Z foram utilizados os quantis de referência da distribuição normal padrão.

Observando a Tabela 4.18 nota-se que as taxas de rejeição nulas convergem para os respectivos níveis nominais utilizados quando o tamanho amostral aumenta.

Nas simulações de poder foram utilizados dois valores para  $\varsigma_1$  na geração dos dados, a saber, 0.5 e 1.5, sendo o valor nulo  $\varsigma_1 = 0$ . Os valores críticos foram obtidos das simulações de tamanho, visando comparar os poderes de testes com mesmo tamanho. Note que quando  $n$  aumenta os poderes dos testes convergem para 100%, como observado na Tabela 4.19. Além disso, à medida em que o valor verdadeiro de  $\varsigma_1$  se distancia do valor nulo, os poderes dos testes atingem 100% mais rapidamente.

## 4.7 Seleção do modelo, resíduos e bondade do ajuste

Após estimar os parâmetros do modelo de regressão, uma etapa importante é realizar análise de diagnóstico para certificar que o modelo ajustado é uma aproximação adequada do modelo populacional. Essa etapa consiste em um conjunto de técnicas utilizadas para

averiguar a adequabilidade do modelo, dentre elas está a análise dos resíduos.

Na literatura estatística a análise de resíduos é largamente abordada. Cox & Snell (1968) fornecem uma definição mais geral de resíduos e ilustram algumas de suas propriedades e aplicações. Para o modelo de regressão beta inflacionado, Ospina e Ferrari (2012) discutem sobre os resíduos padronizados e resíduos quantis aleatorizados. No contexto de regressão Kumaraswamy, Oliveira e Bayer (2017) utilizam resíduo ordinário padronizado e resíduo quantílico para investigar a adequacidade do modelo proposto.

Além de análise de resíduo é muito comum calcular critérios para seleção de modelo, bem como o coeficiente de determinação que é útil para mensurar a qualidade do ajuste. Adicionalmente a essas ferramentas, também será apresentado o teste RESET que é utilizado para verificar se o modelo está corretamente especificado.

#### 4.7.1 Resíduos quantis aleatorizados

Oliveira e Bayer (2017) utilizam o resíduo quantílico na análise de diagnóstico do modelo de regressão Kumaraswamy. Segundo os autores, simulações piloto indicam que o resíduo quantílico é o mais adequado no contexto desse modelo. Ospina e Ferrari (2012) fazem uma adequação do resíduo quantílico para os modelos de regressão beta inflacionados, com base no resíduo generalizado proposto por Cox & Snell (1968). Utilizaremos então os resíduos quantis aleatorizados propostos por Ospina e Ferrari (2012).

Seja  $U = G(y, \boldsymbol{\theta})$  uma variável uniformemente distribuída no intervalo  $(0, 1)$ , em que  $Y$  é uma variável aleatória contínua. Considere também que  $\Phi(\cdot)$  é a função distribuição acumulada da distribuição normal padrão. Então,

$$N = \Phi^{-1} [G(y, \boldsymbol{\theta})] = \Phi^{-1}(U)$$

é uma variável aleatória que tem distribuição normal padrão. Como  $\boldsymbol{\theta}$  é uma quantidade desconhecida, devemos substituí-lo por sua estimativa de máxima verossimilhança.

No caso dos modelos Kumaraswamy inflacionados,  $G(y, \boldsymbol{\theta})$  não é contínua, da mesma maneira que a função distribuição acumulada no contexto de regressão beta inflacionada. Diante disto, Ospina e Ferrari (2012) propuseram uma definição mais geral para o resíduo  $N$ , optando por uma versão aleatorizada do resíduo de Cox & Snell (1968), definindo assim o resíduo quantil aleatorizado para o modelo de regressão beta inflacionado.

O resíduo quantil aleatorizado é dado por

$$r_i^q = \Phi^{-1}(u_i), \quad i = 1, \dots, n,$$

em que  $u_i$  é uma variável aleatória uniforme no intervalo  $(a_i, b_i]$ , com

$$a_i = \lim_{y \rightarrow y_i} \text{KI}_c(y; \hat{\lambda}_i, \hat{\omega}_i, \hat{\tau}_i) \quad \text{e} \quad b_i = \text{KI}_c(y_i; \hat{\lambda}_i, \hat{\omega}_i, \hat{\tau}_i).$$

No caso do modelo de regressão Kumaraswamy inflacionado em zero note que

- Se  $y_i = 0$

$$\begin{aligned} a_i &= \lim_{y \rightarrow 0} \text{KI}_0(y; \hat{\lambda}_i, \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i) \lim_{y \rightarrow 0} G(y; \hat{\omega}_i, \hat{\tau}_i) = 0 \quad \text{e} \\ b_i &= \text{KI}_0(0; \hat{\lambda}_i, \hat{\omega}_i, \hat{\tau}_i) = \hat{\lambda}_i, \end{aligned}$$

em que  $G(y; \hat{\omega}_i, \hat{\tau}_i) = \left[ 1 - \left( 1 - y^{1/\hat{\tau}_i} \right)^{\frac{\ln(0.5)}{\ln(1 - \hat{\omega}_i^{1/\hat{\tau}_i})}} \right]$  é a função distribuição acumulada da Kumaraswamy reparametrizada. Então, nesse caso,  $u_i \sim \mathcal{U}(0, \hat{\lambda}_i]$ .

- Se  $y_i \in (0, 1) \Rightarrow u_i = \text{KI}_0(y_i; \hat{\lambda}_i, \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i)G(y_i; \hat{\omega}_i, \hat{\tau}_i)$ .

De forma análoga ao caso anterior, para o modelo de regressão Kumaraswamy inflacionado em um:

- Se  $y_i = 1$

$$\begin{aligned} a_i &= \lim_{y \rightarrow 1} \text{KI}_1(y; \hat{\lambda}_i, \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i) \lim_{y \rightarrow 1} G(y; \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i) \quad \text{e} \\ b_i &= \text{KI}_1(1; \hat{\lambda}_i, \hat{\omega}_i, \hat{\tau}_i) = 1. \end{aligned}$$

Assim,  $u_i \sim \mathcal{U}(1 - \hat{\lambda}_i, 1]$ .

- Se  $y_i \in (0, 1) \Rightarrow u_i = \text{KI}_1(y_i; \hat{\lambda}_i, \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i)G(y_i; \hat{\omega}_i, \hat{\tau}_i)$ .

Já para o modelo de regressão Kumaraswamy inflacionado em zero e um temos que  $r_i^q = \Phi^{-1}(u_i)$ ,  $i = 1, \dots, n$ , em que  $u_i \sim \mathcal{U}(a_i, b_i]$ , com  $a_i = \lim_{y \rightarrow y_i} \text{KIZU}(y; \hat{\lambda}_i, \hat{p}_i, \hat{\omega}_i, \hat{\tau}_i)$  e  $b_i = \text{KIZU}(y_i; \hat{\lambda}_i, \hat{p}_i, \hat{\omega}_i, \hat{\tau}_i)$ . Nesse caso note que

- Se  $y_i = 0$

$$\begin{aligned} a_i &= \lim_{y \rightarrow 0} \text{KIZU}(y; \hat{\lambda}_i, \hat{p}_i, \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i) \lim_{y \rightarrow 0} G(y; \hat{\omega}_i, \hat{\tau}_i) = 0 \quad \text{e} \\ b_i &= \text{KIZU}(0; \hat{\lambda}_i, \hat{p}_i, \hat{\omega}_i, \hat{\tau}_i) = \hat{\lambda}_i(1 - \hat{p}_i). \end{aligned}$$

Então,  $u_i \sim \mathcal{U}(0, \hat{\lambda}_i(1 - \hat{p}_i)]$ .

- Se  $y_i = 1$

$$\begin{aligned} a_i &= \lim_{y \rightarrow 1} \text{KIZU}(y; \hat{\lambda}_i, \hat{p}_i, \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i) \lim_{y \rightarrow 1} G(y; \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i) \quad \text{e} \\ b_i &= \text{KIZU}(1; \hat{\lambda}_i, \hat{p}_i, \hat{\omega}_i, \hat{\tau}_i) = 1. \end{aligned}$$

Então,  $u_i \sim \mathcal{U}(1 - \hat{\lambda}_i, 1]$ .

- Se  $y \in (0, 1) \Rightarrow u_i = \text{KIZU}(y_i; \hat{\lambda}_i, \hat{p}_i, \hat{\omega}_i, \hat{\tau}_i) = (1 - \hat{\lambda}_i)G(y_i; \hat{\omega}_i, \hat{\tau}_i)$

Segundo Ospina e Ferrari (2012), o procedimento de aleatorização é introduzido para produzir resíduos normais contínuos.

### 4.7.2 Critérios de seleção do modelo

É muito comum utilizar critério de seleção de modelo no contexto de modelos de regressão. Os mais utilizados e que são abordados nessa tese são o critério de informação de Akaike (AIC) (Akaike, 1974) e o critério de informação Bayesiano (BIC) dados por

$$\begin{aligned} \text{AIC} &= -2\ell(\hat{\boldsymbol{\theta}}) + 2\kappa, \\ \text{BIC} &= -2\ell(\hat{\boldsymbol{\theta}}) + \kappa \log(n), \end{aligned}$$

em que  $\ell(\boldsymbol{\theta})$  é o logaritmo da função de verossimilhança avaliada no estimador de máxima verossimilhança e  $\kappa$  é o número de parâmetros estimados do modelo. O modelo selecionado é o que possuir menor valor para AIC ou BIC.

### 4.7.3 Coeficiente de determinação generalizado

Visando medir a qualidade do ajuste do modelo selecionado, também é bastante útil calcular o coeficiente de determinação generalizado de Nagelkerke (1991) dado por

$$R_G^2 = 1 - \left( \frac{L_{null}}{L_{fit}} \right)^{\frac{2}{n}} = 1 - \exp \left\{ -\frac{2}{n} \left[ \ell(\hat{\boldsymbol{\theta}}) - \ell(0) \right] \right\},$$

em que  $\ell(0) = \log(L_{null})$  é a log-verossimilhança maximizada referente ao modelo nulo e  $\ell(\hat{\boldsymbol{\theta}}) = \log(L_{fit})$  é a log-verossimilhança do modelo ajustado. O coeficiente de determinação  $R_G^2$  mede a proporção da variabilidade da variável resposta que é explicada pelas covariáveis utilizadas no modelo ajustado. Logo,  $R_G^2 \in (0, 1)$  e quanto mais próximo de 1 melhor é a qualidade do ajuste.

Outros *pseudo*  $R^2$  podem ser definidos como o quadrado do coeficiente de correlação amostral entre os valores da variável resposta,  $y_1, y_2, \dots, y_n$  e seus correspondentes valores preditos dados por  $\widehat{\mathbb{E}(y_1)}, \widehat{\mathbb{E}(y_2)}, \dots, \widehat{\mathbb{E}(y_n)}$  ou  $R_p^2 = 1 - \log L_{fit} / \log L_{null}$  (McFadden, 1974).

### 4.7.4 Teste RESET

Uma prática importante quando se deseja verificar a bondade do ajuste é a determinação da existência ou não de erro de especificação no modelo. Nesse contexto, um teste bastante empregado é o *Ramsey Regression Equation Specification Error Test* (RESET), proposto por Ramsey (1969), que verifica a validade da hipótese de correta especificação do modelo. Nesta tese usaremos uma adaptação do teste RESET, que inicialmente foi introduzido para o modelo de regressão linear. O teste RESET avalia se combinações não-lineares das variáveis independentes são relevantes ou não na explicação da variável resposta. A ideia é que se o modelo está corretamente especificado, essas combinações não-lineares devem ser irrelevantes.

O teste funciona da seguinte forma: sob  $H_0$ , o modelo está corretamente especificado e sob  $H_1$  existe algum erro de especificação no modelo. Para testar tal hipótese, no modelo de regressão Kumaraswamy inflacionado, adicionamos  $\hat{\eta}^2$  como covariável no submodelo da mediana e então estimamos o modelo considerando a inclusão desse novo regressor. Utilizamos o teste da razão de verossimilhanças para testar a hipótese de correta especificação do modelo. Se  $H_0$  não for rejeitada o modelo está corretamente especificado.

## 4.8 Aplicações dos modelos Kumaraswamy inflacionados

Nessas aplicações foi utilizado um conjunto de dados do Atlas do Desenvolvimento Humano no Brasil para o ano de 2010, contendo informações sobre 5565 municípios brasileiros. O Atlas do Desenvolvimento Humano 2013 contém indicadores de renda, educação, trabalho, demografia, habitação e vulnerabilidade, baseados em informações retiradas do Censo Demográfico 2010.

Nessa Seção, objetivamos fazer um ajuste inicial dos modelos Kumaraswamy inflacionados propostos e detectar quais variáveis regressoras são significativas para explicação da variável reposta em cada caso. Também foi realizado o teste RESET para verificar se os modelos ajustados estão corretamente especificados ou não.

Na Seção 4.8.1 é apresentada a aplicação do modelo Kumaraswamy inflacionado em zero e, na Seção 4.8.2, é apresentada a aplicação do modelo Kumaraswamy inflacionado em zero e um.

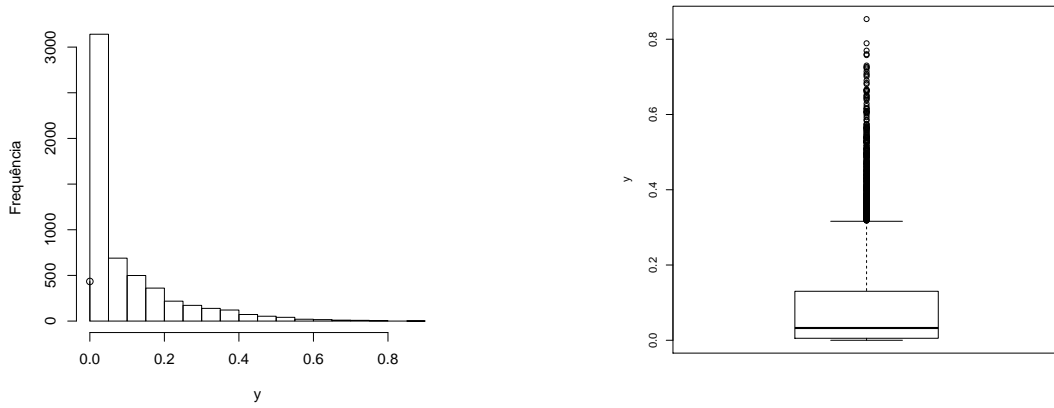
### 4.8.1 Modelo de regressão Kumaraswamy inflacionado em zero

O objetivo dessa aplicação é ajustar o modelo de regressão Kumaraswamy inflacionado em zero utilizando um conjunto de dados reais. Para tal fim, foi usada como variável resposta a proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados, contendo 5565 observações referentes aos municípios brasileiros em 2010. Selecionamos potenciais variáveis regressoras a partir de uma análise de correlação. Na Tabela 4.20 estão apresentadas algumas medidas resumo da variável resposta, que contém 434 valores iguais a zero. Note, na Tabela 4.20, que a variável proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados apresenta comportamento assimétrico, em que 75% da amostra é menor ou igual a 0.1302.

**Tabela 4.20:** Medidas resumo da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
0.0000	0.0053	0.0326	0.0920	0.1302	0.8536

A Figura 4.1 contém o histograma e o boxplot da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados. Percebe-se o inflacionamento



**Figura 4.1:** Histograma (painel esquerdo) e boxplot (painel direito) da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados.

de zeros através do círculo que aparece no histograma, representando a frequência de zeros na amostra, bem como se nota no boxplot que esses valores iguais a zero não são considerados como *outliers*.

Para escolher covariáveis a fim de ajustar o modelo de regressão Kumaraswamy inflacionado em zero, inicialmente foram calculados os coeficientes de correlação de postos de Spearman entre a variável resposta e covariáveis encontradas na base de dados. A princípio foram consideradas como regressoras as 10 variáveis que estão listadas no Apêndice A.

Utilizando as variáveis regressoras descritas no Apêndice A, foi ajustado o modelo de regressão Kumaraswamy inflacionado em zero com dispersão variável dado por

$$\begin{cases} h(\lambda_i) = \pi_0 + z_{i1}\pi_1 + z_{i2}\pi_2 + z_{i3}\pi_3 + z_{i4}\pi_4 + z_{i5}\pi_5 + z_{i6}\pi_6 + z_{i7}\pi_7 + z_{i8}\pi_8 + z_{i9}\pi_9 \\ \quad + z_{i10}\pi_{10} \\ f(\omega_i) = \gamma_0 + x_{i1}\gamma_1 + x_{i2}\gamma_2 + x_{i3}\gamma_3 + x_{i4}\gamma_4 + x_{i5}\gamma_5 + x_{i6}\gamma_6 + x_{i7}\gamma_7 + x_{i8}\gamma_8 + x_{i9}\gamma_9 \\ \quad + x_{i10}\gamma_{10} \\ d(\tau_i) = \varsigma_0 + q_{i1}\varsigma_1 + q_{i2}\varsigma_2 + q_{i3}\varsigma_3 + q_{i4}\varsigma_4 + q_{i5}\varsigma_5 + q_{i6}\varsigma_6 + q_{i7}\varsigma_7 + q_{i8}\varsigma_8 + q_{i9}\varsigma_9 \\ \quad + q_{i10}\varsigma_{10}, \end{cases}$$

em que  $i = 1, \dots, 5565$ . Os parâmetros do modelo foram estimados através da maximização numérica da função de log-verossimilhança utilizando o método quasi-Newton BFGS (Nocedal, 1980) com gradiente analítico. As dez covariáveis foram utilizadas nos três submodelos, totalizando assim 33 parâmetros, incluindo os três interceptos referente a cada submodelo. Os valores estimados para cada coeficiente, juntamente com os erros-padrão associados, bem como as estatísticas de teste da razão de verossimilhanças com os respectivos  $p$ -valores estão expostos na Tabela 4.21.

Com base nos resultados contidos na Tabela 4.21 pode-se notar que os parâmetros  $\pi_2$ ,  $\pi_4$  e  $\pi_8$  não são significativos para explicar a variável resposta, ao nível de 5% de significância. Logo, T\_FLSUPER, T\_FUND18A24 e T\_CRIFUNDIN\_TODOS não parecem ser relevantes para explicar o comportamento do parâmetro de mistura do modelo Kumaraswamy inflacionado em zero ajustado. De forma análoga, observando os  $p$ -valores

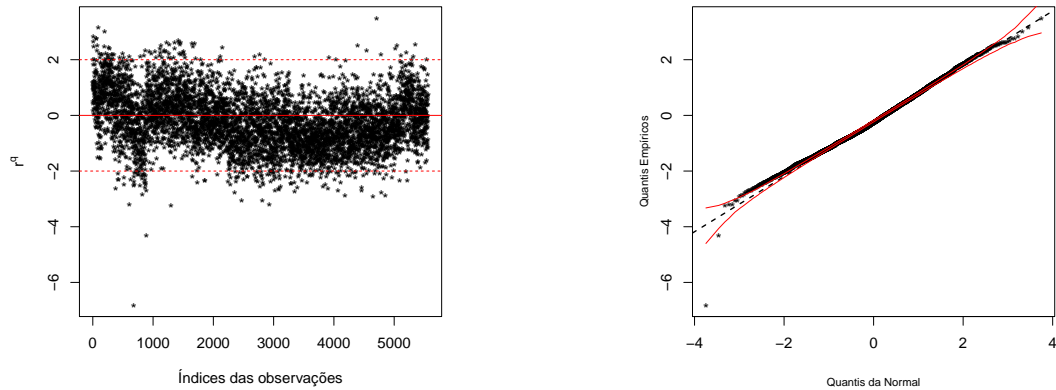
**Tabela 4.21:** Estimativas, erros-padrão, estatística da razão de verossimilhanças e  $p$ -valor - modelo KI em zero.

Parâmetro	Estimativa	EP	$RV$	$p$ -valor
$\pi_0$	-1.0978	0.7356	-	-
$\pi_1$	14.0478	1.2748	4.6706	0.0307
$\pi_2$	0.0098	0.9783	8.0701e-005	0.9928
$\pi_3$	-9.8616	1.0002	4.5661	0.0326
$\pi_4$	-1.6182	0.9928	1.3763	0.2407
$\pi_5$	3.8303	0.8579	13.8400	0.0002
$\pi_6$	-6.3493	0.9970	12.7950	0.0003
$\pi_7$	-15.5016	1.0613	75.1240	4.4200e-018
$\pi_8$	0.1817	0.8092	0.0451	0.8317
$\pi_9$	-1.9962	0.9770	4.6125	0.0317
$\pi_{10}$	-150.4010	1.0082	70.3680	4.9209e-017
$\gamma_0$	-1.2127	0.4907	-	-
$\gamma_1$	2.5603	0.9631	9.5917	0.0019
$\gamma_2$	-4.3376	0.7184	67.0030	2.7103e-016
$\gamma_3$	3.0788	0.9543	18.7510	1.4897e-005
$\gamma_4$	-1.5187	0.6320	11.5920	0.0007
$\gamma_5$	-1.5806	0.4925	16.7250	4.3198e-005
$\gamma_6$	-7.4114	0.9329	66.6770	3.1983e-016
$\gamma_7$	4.6309	0.2214	500.1300	8.9060e-111
$\gamma_8$	-2.2722	0.3965	57.9100	2.7439e-014
$\gamma_9$	2.0291	0.2281	95.0840	1.8246e-022
$\gamma_{10}$	-0.2101	0.4496	0.36286	0.5469
$\varsigma_0$	0.2736	0.2810	-	-
$\varsigma_1$	0.2715	1.0466	0.20371	0.65174
$\varsigma_2$	1.1846	0.3876	13.9270	0.0002
$\varsigma_3$	-1.8607	0.8730	11.9920	0.0005
$\varsigma_4$	-0.1092	0.4427	0.1179	0.7313
$\varsigma_5$	-0.0517	0.2867	0.0394	0.8426
$\varsigma_6$	-0.2362	0.6520	0.1966	0.6575
$\varsigma_7$	-0.8661	0.1960	31.1670	2.3670e-008
$\varsigma_8$	-0.0731	0.1572	0.1157	0.7337
$\varsigma_9$	-0.5187	0.1653	11.0210	0.0009
$\varsigma_{10}$	1.1167	0.3205	18.8950	1.3813e-005

verifica-se que apenas a variável T\_SLUZ não é significativa para explicar variações na mediana da proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados. Por fim, analisando o terceiro submodelo onde o parâmetro de dispersão é variável, observa-se que os regressores T\_ANALF18A24, T\_FUND18A24, T\_MED18A24, T\_SUPER25M e T\_CRIFUNDIN\_TODOS não são significativos para explicar o parâmetro de dispersão do modelo ajustado nessa aplicação.

Na Figura 4.2 estão apresentados, o gráfico dos resíduos quantis aleatorizados e o envelope simulado dos resíduos do modelo Kumaraswamy inflacionado em zero ajustado para a variável proporção de pessoas em domicílios com abastecimento de água e esgoto sanitário inadequados, considerando as covariáveis significativas. Nota-se que o modelo ajustado não é adequado. Com base no teste RESET também é possível verificar que há alguma má especificação no modelo, uma vez que  $p$ -valor = 5.5711e - 067.





**Figura 4.2:** Resíduos (painel esquerdo) e envelope simulado dos resíduos (painel direito) do modelo de regressão Kumaraswamy inflacionado em zero ajustado.

#### 4.8.2 Modelo de regressão Kumaraswamy inflacionado em zero e um

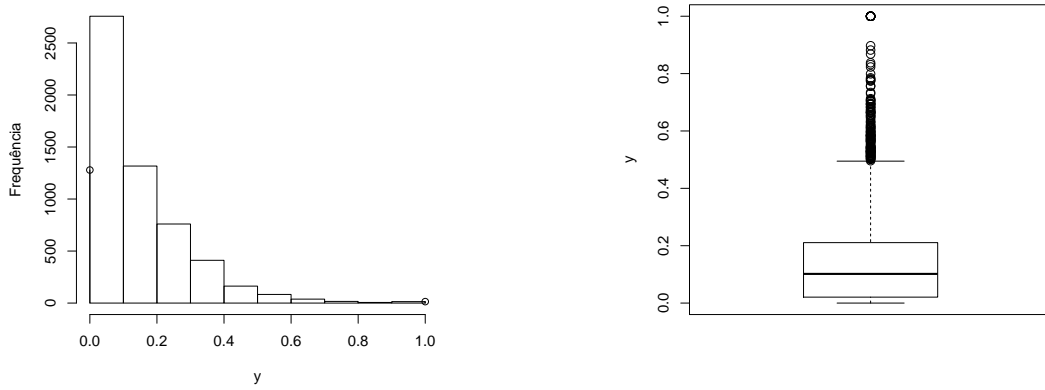
Nessa segunda aplicação foi utilizada como variável resposta a taxa de desocupação (desemprego) da população de 10 a 14 anos de idade. Essa variável é dada pela proporção da população economicamente ativa (PEA) nessa faixa etária que estava desocupada, ou seja, que não estava ocupada na semana anterior à data do censo, mas havia procurado trabalho ao longo do mês anterior à data dessa pesquisa. Esse conjunto de dados contém 5565 observações. Na Tabela 4.22 encontram-se medidas resumo da variável resposta, que contém 1279 valores iguais a zero e 14 valores iguais a um. O objetivo dessa aplicação é ajustar o modelo de regressão Kumaraswamy inflacionado em zero e um à proporção de pessoas de 10 a 14 anos desocupadas, mas que haviam procurado trabalho, nos municípios brasileiros em 2010. Observando a Tabela 4.22, nota-se que a variável resposta apresenta comportamento assimétrico, sendo 75% da amostra menor ou igual a 0.2107.

**Tabela 4.22:** Medidas resumo da proporção de pessoas de 10 a 14 anos desocupadas no Brasil em 2010, mas que havia procurado trabalho ao longo do mês anterior à data dessa pesquisa.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
0.0000	0.0207	0.1017	0.1395	0.2107	1.0000

A Figura 4.3 contém o histograma e o boxplot da proporção de pessoas de 10 a 14 anos desocupadas no Brasil em 2010. Mais uma vez nota-se o comportamento assimétrico dos dados e a presença de zeros e uns, como mencionado anteriormente.

Observando os coeficientes de correlação de Spearman entre a variável resposta e algumas potenciais covariáveis encontradas na base de dados, inicialmente foram considerados os 9 regressores listados no Apêndice B. Utilizando as variáveis regressoras descritas anteriormente, foi ajustado o modelo de regressão Kumaraswamy inflacionado em zero



**Figura 4.3:** Histograma (painel esquerdo) e boxplot (painel direito) da proporção de pessoas de 10 a 14 anos desocupadas no Brasil em 2010.

com dispersão variável, expresso em (4.4), que ficou da seguinte forma:

$$\begin{cases} h(\xi_{0i}) &= \varphi_0 + z_{0i1}\varphi_1 + z_{0i2}\varphi_2 + z_{0i3}\varphi_3 + z_{0i4}\varphi_4 + z_{0i5}\varphi_5 + z_{0i6}\varphi_6 + z_{0i7}\varphi_7 + z_{0i8}\varphi_8 \\ &+ z_{0i9}\varphi_9 \\ l(\xi_{1i}) &= v_0 + z_{1i1}v_1 + z_{1i2}v_2 + z_{1i3}v_3 + z_{1i4}v_4 + z_{1i5}v_5 + z_{1i6}v_6 + z_{1i7}v_7 + z_{1i8}v_8 \\ &+ z_{1i9}v_9 \\ f(\omega_i) &= \gamma_0 + x_{i1}\gamma_1 + x_{i2}\gamma_2 + x_{i3}\gamma_3 + x_{i4}\gamma_4 + x_{i5}\gamma_5 + x_{i6}\gamma_6 + x_{i7}\gamma_7 + x_{i8}\gamma_8 + x_{i9}\gamma_9 \\ d(\tau_i) &= \varsigma_0 + q_{i1}\varsigma_1 + q_{i2}\varsigma_2 + q_{i3}\varsigma_3 + q_{i4}\varsigma_4 + q_{i5}\varsigma_5 + q_{i6}\varsigma_6 + q_{i7}\varsigma_7 + q_{i8}\varsigma_8 + q_{i9}\varsigma_9, \end{cases}$$

em que  $i = 1, \dots, 5565$ .

Os nove regressores foram utilizados nos quatro submodelos, totalizando assim 40 parâmetros, incluindo os quatro interceptos de cada submodelo. Os parâmetros do modelo foram estimados através da maximização numérica da função de log-verossimilhança utilizando o método quasi-Newton BFGS (Nocedal, 1980) com primeiras derivadas analíticas. Os valores estimados dos coeficientes de regressão, juntamente com os erros-padrão associados, bem como as estatísticas de teste da razão de verossimilhanças, com os respectivos  $p$ -valores, estão expostos na Tabela 4.23.

Analisando a Tabela 4.23 pode-se concluir que T\_MULCHEFEFIF014, T\_M15A17CF, T\_ATRASO\_2\_FUND, T\_FLFUND, T\_FREQ11A14 e T\_CRIFUNDIN\_TODOS são significantes para explicar a probabilidade da variável resposta ser igual a zero. De forma análoga, observando os  $p$ -valores verifica-se que apenas as variáveis T\_ANALF11A14 e T\_FREQ11A14 são significantes para explicar a probabilidade da variável resposta ser igual a um. Já a mediana da variável resposta pode ser explicada por T\_M10A14CF, T\_ATRASO\_2\_FUND, T\_FREQ11A14, PINDCRI e T\_CRIFUNDIN\_TODOS. Por fim, analisando o submodelo do parâmetro de dispersão, observa-se que os regressores T\_ATRASO\_2\_FUND, T\_FREQ11A14, PINDCRI e T\_CRIFUNDIN\_TODOS são significativos para explicar a dispersão da variável resposta utilizada.

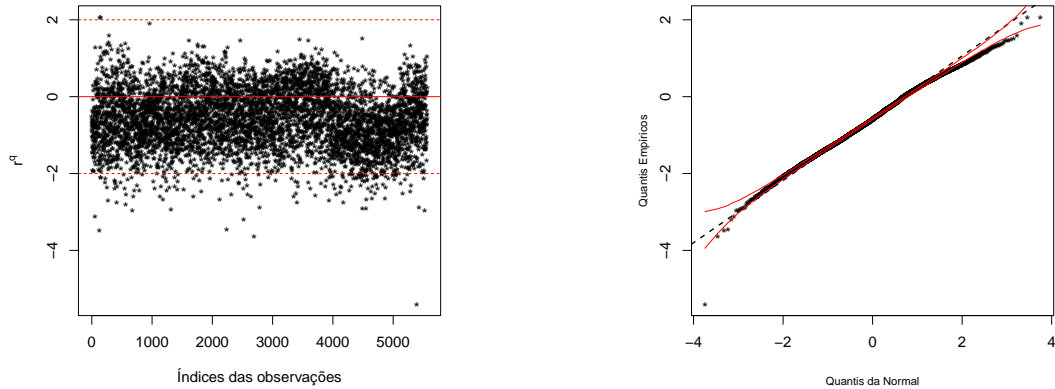
Na Figura 4.4 é possível visualizar o gráfico dos resíduos quantis aleatorizados e o envelope simulado para estes resíduos, obtidos do modelo Kumaraswamy inflacionado em zero e um ajustado para a proporção de pessoas de 10 a 14 anos desocupadas no Brasil

**Tabela 4.23:** Estimativas, erros-padrão, estatística da razão de verossimilhanças e  $p$ -valor - modelo KIZU.

Parâmetro	Estimativa	EP	$RV$	$p$ -valor
$\varphi_0$	-16.5059	1.8298	-	-
$\varphi_1$	-3.0882	0.4475	47.9580	4.3551e-012
$\varphi_2$	5.1067	4.8987	0.9581	0.3276
$\varphi_3$	-3.4762	0.8220	18.7130	1.5191e-005
$\varphi_4$	0.7220	1.4294	0.1982	0.6562
$\varphi_5$	-3.2260	0.6356	26.2560	2.9910e-007
$\varphi_6$	4.2253	1.3844	7.9940	0.0047
$\varphi_7$	11.5971	2.0212	31.5960	1.8979e-008
$\varphi_8$	0.4783	0.3802	1.4606	0.2268
$\varphi_9$	3.5603	0.4184	73.9800	7.8931e-018
$v_0$	-57.0294	19.6860	-	-
$v_1$	-5.1570	3.9610	1.5551	0.2124
$v_2$	37.6259	35.3476	0.8702	0.3509
$v_3$	-8.3592	6.9039	1.4269	0.2323
$v_4$	25.1436	11.7706	3.2156	0.0729
$v_5$	-6.8316	5.2236	1.6076	0.2048
$v_6$	0.4281	10.9542	0.0014	0.9698
$v_7$	53.8463	19.9244	7.0047	0.0081
$v_8$	0.3598	3.2753	0.0095	0.9223
$v_9$	-0.5608	3.1606	0.026	0.8711
$\gamma_0$	0.6200	0.6536	-	-
$\gamma_1$	0.1637	0.2129	0.5837	0.4448
$\gamma_2$	4.5641	2.2299	3.3842	0.0658
$\gamma_3$	0.5860	0.3796	2.3061	0.1289
$\gamma_4$	0.7857	0.8172	0.9910	0.3195
$\gamma_5$	1.5765	0.3054	26.3630	2.8297e-007
$\gamma_6$	1.0466	0.6979	2.1820	0.1396
$\gamma_7$	-2.5634	0.8521	8.0024	0.0047
$\gamma_8$	-1.1554	0.2067	32.9530	9.4437e-009
$\gamma_9$	-2.7197	0.1967	183.8000	7.1703e-042
$\varsigma_0$	-1.2624	0.5071	-	-
$\varsigma_1$	-0.1511	0.1678	0.7842	0.3758
$\varsigma_2$	-3.2596	1.7803	2.5169	0.1126
$\varsigma_3$	-0.2082	0.3241	0.4143	0.5198
$\varsigma_4$	-0.0869	0.6206	0.0205	0.8861
$\varsigma_5$	-0.7678	0.2592	8.7988	0.0030
$\varsigma_6$	-0.7422	0.5994	1.5178	0.2179
$\varsigma_7$	1.2718	0.7189	2.8209	0.0930
$\varsigma_8$	0.4347	0.1602	7.6652	0.0056
$\varsigma_9$	1.4760	0.1740	71.498	2.7752e-017

em 2010. Observa-se que há pontos influentes no ajuste, os quais devem ser avaliados, posteriormente, com mais atenção. Adicionalmente foi realizado o teste RESET proposto por Ramsey (1969) e utilizado por Oliveira e Bayer (2017) no contexto de regressão Kumaraswamy, no qual  $\hat{\eta}_i^2$  foi adicionado como regressor no submodelo do parâmetro da mediana. Vale ressaltar que o teste foi realizado utilizando apenas as covariáveis significativas em cada submodelo. Sendo assim, foi testada a hipótese de que o modelo está corretamente especificado e através da estatística da razão de verossimilhanças que foi igual a 1.9158, a hipótese nula não foi rejeitada, aos níveis usuais de significância,

uma vez que  $p\text{-valor} = 0.1663$ . Logo, pelo teste RESET, o modelo está corretamente especificado. Todavia, faz-se necessário um diagnóstico mais aprofundado para avaliar os pontos influentes e possíveis *outliers*, o que pode ser considerado como trabalhos futuros.



**Figura 4.4:** Resíduos (painel esquerdo) e envelope simulado dos resíduos (painel direito) do modelo de regressão Kumaraswamy inflacionado em zero e um.

## 4.9 Conclusões

Neste capítulo foram propostos modelos de regressão Kumaraswamy inflacionados que são úteis para modelar variáveis respostas que assumem valores nos intervalos  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$ . Estimação pontual, através do método de máxima verossimilhança, foi discutida para os modelos, bem como foram apresentados intervalos de confiança assintóticos e também testes de hipóteses.

Os desempenhos dos estimadores de máxima verossimilhança dos parâmetros dos modelos Kumaraswamy inflacionados, bem como dos intervalos de confiança para esses parâmetros, foram avaliados através de simulações de Monte Carlo. Também foram realizadas simulações de Monte Carlo para verificar os desempenhos dos testes razão de verossimilhanças e Z, comparando seus tamanhos e poderes.

A partir dos resultados encontrados pode-se concluir que os estimadores de máxima verossimilhança dos modelos Kumaraswamy inflacionados e também os testes apresentados possuem bom desempenho.

Adicionalmente, foram realizadas aplicações a dados reais, onde foi possível verificar os ajustes dos modelos Kumaraswamy inflacionados propostos.

Como trabalhos futuros, é possível desenvolver técnicas de diagnóstico para detectar pontos influentes, *outliers*, etc, bem como abordar outros resíduos.

## 5 CONSIDERAÇÕES FINAIS

Para estudar o comportamento de uma variável no intervalo  $(0, 1)$  é possível utilizar algumas distribuições de probabilidade, como por exemplo, a distribuição beta, a distribuição simplex, a distribuição Kumaraswamy, entre outras. A distribuição beta é bastante utilizada para ajustar dados de taxas e proporções, bem como outras variáveis que possuem suporte em  $(0, 1)$ . Contudo, a distribuição Kumaraswamy vem ganhando espaço na literatura estatística, pois possui algumas vantagens, destacando-se o fato de ter uma função quantílica simples, sem depender de funções matemáticas especiais, como é o caso da distribuição beta.

Na prática é comum que dados de taxas e proporções assumam valores iguais a zeros ou uns. Então, para acomodar os extremos do intervalo é preciso definir uma nova distribuição que resulta da mistura entre uma distribuição degenerada em zero ou em um e uma distribuição contínua. Esse foi o enfoque da distribuição beta inflacionada em zero ou em um, proposta por Ospina e Ferrari (2010). Para o caso em que as proporções são observadas no intervalo  $[0, 1]$ , a proposta é construir uma nova distribuição que resulta da mistura entre duas distribuições, uma contínua e uma discreta, permitindo assim modelar os valores zeros e uns em uma amostra de dados no intervalo  $[0, 1]$ . Ospina e Ferrari (2010) também propuseram uma distribuição inflacionada em zero e um considerando como base a distribuição beta.

No contexto de modelos de regressão, é possível avaliar o comportamento de uma variável dado o conhecimento de variáveis explicativas conhecidas. Para dados no intervalo  $(0, 1)$ , Ferrari e Cribari (2004) propuseram o modelo de regressão beta. Já para variáveis que assumem valores nos intervalos  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$ , é possível utilizar os modelos de regressão beta inflacionados propostos por Ospina e Ferrari (2012), que são baseados em distribuições beta inflacionadas (Ospina e Ferrari, 2010).

Nessa tese foram propostas a distribuição Kumaraswamy inflacionada em zero, para dados que apresentam valores no intervalo  $[0, 1)$ , a distribuição Kumaraswamy inflacionada em um, para variáveis no intervalo  $(0, 1]$ , e a distribuição Kumaraswamy inflacionada em zero e um, para dados no intervalo  $[0, 1]$ . Adicionalmente, foram propostos modelos de regressão Kumaraswamy inflacionados para modelar variáveis dependentes distribuídas nos intervalos  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$ .

Em cada cenário utilizado nessa tese, foram desenvolvidas as inferências pontuais e intervalares para cada um dos parâmetros das distribuições Kumaraswamy inflacionadas, bem como dos modelos Kumaraswamy inflacionados. Além disso, foram também apresentados testes de hipóteses para os parâmetros das distribuições e modelos propostos.

Nos Capítulos 2 e 3 abordamos as respectivas inferências para cada uma dessas distribuições, ou seja, apresentamos a estimação pontual dos parâmetros de cada distribuição proposta, através do método de máxima verossimilhança; a distribuição assintótica dos

estimadores; estimação intervalar e testes de hipóteses sobre os parâmetros. Foram realizadas simulações de Monte Carlo para verificar os desempenhos dos estimadores de máxima verossimilhança dos parâmetros de cada distribuição e também dos testes de hipóteses. Observamos que tanto os estimadores quanto os testes apresentam bons desempenhos em amostras finitas. Por fim, foram apresentadas aplicações a dados reais, nas quais verificamos o ajuste das distribuições propostas e estimamos os parâmetros com base nas amostras utilizadas.

No Capítulo 4 propomos o modelo de regressão Kumaraswamy inflacionado em zero ou um e o modelo de regressão Kumaraswamy inflacionado em zero e um. Desenvolvemos a inferência pontual para ambos os modelos e averiguamos o desempenho dos estimadores de máxima verossimilhança, dos intervalos de confiança e dos testes de hipóteses utilizados, através de simulações de Monte Carlo. Com base nos resultados encontrados, constatamos o bom desempenho dos estimadores de máxima verossimilhança dos parâmetros dos modelos Kumaraswamy inflacionados. Adicionalmente, foram avaliados os desempenhos do teste da razão de verossimilhanças e do teste  $Z$ , os quais apresentam bons desempenhos em amostras finitas. Apresentamos também critérios para seleção dos modelos, resíduos quantis aleatorizados,  $pseudo R^2$  e o teste RESET para detectar se há má especificação no modelo. Ajustamos os modelos de regressão Kumaraswamy inflacionados a dados reais, estimamos os parâmetros através da máxima verossimilhança e investigamos a bondade dos ajustes.

É válido ressaltar que as distribuições misturas herdaram as características da distribuição Kumaraswamy, ou seja, possuem função densidade de probabilidade bastante flexível, podendo ser bem útil na prática, uma vez que assumem diversas formas, a depender dos parâmetros que as indexam. Além disso, as distribuições Kumaraswamy inflacionadas propostas (com e sem reparametrização), também possuem uma função quantílica simples, o que facilita a modelagem da mediana da variável resposta ou de qualquer outro quantil, o que é uma vantagem em relação a distribuição beta e suas versões inflacionadas, pois a função quantílica da distribuição beta não possui forma fechada. Ademais, modelos de regressão que modelam a mediana em vez de modelar a média da variável resposta, tendem a serem mais robustos, visto que a mediana é uma mediana mais robusta do que a média.

Por fim, essa tese traz uma gama de possibilidades para trabalhos futuros, uma vez que é possível

- aprofundar as técnicas de diagnóstico dos modelos de regressão propostos;
- verificar o desempenho do teste gradiente, bem como das versões *bootstrap* dos testes utilizados na tese;
- discutir novos resíduos para averiguar a bondade do ajuste dos modelos propostos;
- avaliar o comportamento do teste RESET aplicado às distribuições e modelos propostos;

- propor modelos de regressão quantílica com base nas distribuições Kumaraswamy inflacionadas.

## REFERÊNCIAS

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716–723.
- Barreto-Souza, W., Lemonte, A. J. (2013). Bivariate Kumaraswamy distribution: properties and a new method to generate bivariate classes. *Statistics*, 47(6), 1321–1342.
- Brasil, A. (2014). Atlas do desenvolvimento humano no Brasil 2013.
- Bidram, H., Nekoukhou, V. (2013). Double bounded Kumaraswamy-power series class of distributions. *SORT-Statistics and Operations Research Transactions*, 1(2), 211–230.
- Carrasco, J. M., Ferrari, S. L., Cordeiro, G. M. (2010). A new generalized Kumaraswamy distribution. arXiv preprint arXiv:1004.0911.
- Cordeiro, G. M., Castro, M. (2011). A new family of generalized distributions. *Journal of statistical computation and simulation*, 81(7), 883–898.
- Doornik, J. A., Ooms, M. (2007). Introduction to Ox: An Object-Oriented Matrix Language.
- Ferrari, S. L., Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815.
- Frery, A. C., Cribari-Neto, F. Elementos de Estatística Computacional Usando Plataformas de Software Livre/gratuito. Instituto de Matemática Pura e Aplicada, 2005.
- Garg, M. (2009). On generalized order statistics from Kumaraswamy distribution. *Tamsui Oxford Journal of Mathematical Sciences*, 25(2), 153–166.
- Jones, M. C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, 6(1), 70–81.
- Kumaraswamy, P. (1976). Sinpower probability density function. *Journal of Hydrology*, 31(1–2), 181–184.
- Lehmann, E. L., George, C. (2006). Theory of point estimation. *Springer Science & Business Media*.
- Lemonte, A. J. (2011). Improved point estimation for the Kumaraswamy distribution. *Journal of Statistical Computation and Simulation*, 81(12), 1971–1982.



- Lucena, S.E.F. (2017). Essays on Nonnormal Regression Modeling. Tese de doutorado, Departamento de Estatística, Universidade Federal de Pernambuco.
- Mitnik, P. A., Baek, S. (2013). The Kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, 54(1), 177–192.
- McCullagh, P., Nelder, J. A. (1989). Generalized Linear Models, 2<sup>a</sup> ed. Chapman and Hall/CRC.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In: P. Zarembka, ed., *Frontiers in Econometrics*, 105–142. New York: Academic Press.
- Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691–692.
- Nocedal, J. (1980). Updating quasi-Newton matrices with limited storage. *Mathematics of computation*, 35(151), 773–782.
- Oliveira, C. R., Bayer, F. M. (2017). Modelo de regressão Kumaraswamy com função de ligação de Aranda-Ordaz. Working Paper, Universidade Federal de Santa Maria.
- Ospina, R. (2008). Modelos de regressão beta inflacionados. Tese de doutorado - Universidade de São Paulo. São Paulo, 7 de abril de 2008.
- Ospina, R., Ferrari, S. L. (2010). Inflated beta distributions. *Statistical Papers*, 51(1), 111–126.
- Ospina, R., Ferrari, S. L. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6), 1609–1623.
- Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society B*, 350–371.
- Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44(1), 50–57.
- Schwarz, G., et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Serfling, R. J. (2009). Approximation Theorems of Mathematical Statistics. John Wiley & Sons.
- Silva, R. B., Barreto-Souza, W. (2014). Beta and Kumaraswamy distributions as non-nested hypotheses in the modeling of continuous bounded data. arXiv preprint arXiv:1406.1941.

- Stasinopoulos, M., Rigby, B., Akantziliotou, C., Heller, G., Ospina, R., Motpan, N. (2012). Package ‘gamlss. dist’.
- Wang, B. X., Wang, X. K., Yu, K. (2017). Inference on the Kumaraswamy distribution. *Communications in Statistics-Theory and Methods*, 46(5), 2079–2090.
- Yee, T. W. (2008). The VGAM package. *R News*, 8(2), 28–39.

## APÊNDICE A - COVARIÁVEIS DO MODELO KUMARASWAMY INFLACIONADO EM ZERO

**Tabela 1:** Descrição das variáveis regressoras do modelo Kumaraswamy inflacionado em zero, retiradas do Atlas do Desenvolvimento Humano no Brasil 2013.

Variável	Sigla	Descrição
$z_1, x_1, q_1$	T_ANALF18A24	Taxa de analfabetismo da população de 18 a 24 anos de idade
$z_2, x_2, q_2$	T_FLSUPER	Taxa de frequência líquida <sup>1</sup> o ensino superior
$z_3, x_3, q_3$	T_FREQFUND1824	Percentual da população de 18 a 24 anos de idade frequentando o ensino fundamental
$z_4, x_4, q_4$	T_FUND18A24	Percentual da população de 18 a 24 anos com fundamental completo
$z_5, x_5, q_5$	T_MED18A24	Percentual da população de 18 a 24 anos com médio completo
$z_6, x_6, q_6$	T_SUPER25M	Percentual da população de 25 anos ou mais com superior completo
$z_7, x_7, q_7$	PINDCRI	Proporção de crianças extremamente pobres
$z_8, x_8, q_8$	T_CRIFUNDIN_TODOS	Percentual de crianças que vivem em domicílios em que nenhum dos moradores tem o ensino fundamental completo
$z_9, x_9, q_9$	T_MULCHEFEFIF014	Percentual de mães chefes de família, sem fundamental completo e com pelo menos um filho menor de 15 anos de idade
$z_{10}, x_{10}, q_{10}$	T_SLUZ	Percentual de pessoas em domicílios sem energia elétrica

<sup>1</sup>Razão entre o número de pessoas na faixa etária de 18 a 24 anos frequentando o ensino superior (graduação, especialização, mestrado ou doutorado) e a população total dessa mesma faixa etária.

## APÊNDICE B - COVARIÁVEIS DO MODELO KUMARASWAMY INFLACIONADO EM ZERO E UM

**Tabela 2:** Descrição das variáveis regressoras do modelo Kumaraswamy inflacionado em zero e um, retiradas do Atlas do Desenvolvimento Humano no Brasil 2013.

Variável	Sigla	Descrição
$z_1, x_1, q_1$	T_MULCHEFIF014	Percentual de mães chefes de família, sem fundamental completo e com pelo menos um filho menor de 15 anos de idade
$z_2, x_2, q_2$	T_M10A14CF	Percentual de mulheres de 10 a 14 anos de idade que tiveram filhos
$z_3, x_3, q_3$	T_M15A17CF	Percentual de mulheres de 15 a 17 anos de idade que tiveram filhos
$z_4, x_4, q_4$	T_ANALF11A14	Taxa de analfabetismo da população de 11 a 14 anos de idade
$z_5, x_5, q_5$	T_ATRASO_2_FUND	Percentual da população de 6 a 14 anos de idade frequentando o ensino fundamental que tem 2 anos ou mais de atraso idade-série
$z_6, x_6, q_6$	T_FLFUND	Taxa de frequência líquida ao ensino fundamental
$z_7, x_7, q_7$	T_FREQ11A14	Taxa de atendimento escolar da população de 11 a 14 anos de idade
$z_8, x_8, q_8$	PINDCRI	Proporção de crianças extremamente pobres
$z_9, x_9, q_9$	T_CRIFUNDIN_TODOS	Percentual de crianças que vivem em domicílios em que nenhum dos moradores tem o ensino fundamental completo