



Universidade Federal de Campina Grande  
Centro de Ciências e Tecnologia  
Programa de Pós-Graduação em Matemática  
Curso de Mestrado em Matemática

Rafaella Santos Vitorino <sup>†</sup>

# A distribuição gama inversa com zeros ajustados

Campina Grande - PB  
2024

---

<sup>†</sup> Este trabalho contou com o apoio financeiro parcial da FAPESQ

**Rafaella Santos Vitorino**

**A distribuição gama inversa com  
zeros ajustados**

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática da Universidade Federal de Campina Grande, pertencente à área de concentração Probabilidade e Estatística como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Dr. Manoel Ferreira dos Santos Neto

**Campina Grande - PB**

**2024**

V845d

Vitorino, Rafaella Santos.

A distribuição gama inversa com zeros ajustados / Rafaella Santos Vitorino. – Campina Grande, 2024.

35 f. : il.

Dissertação (Mestrado em Matemática) – Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia, 2024.

"Orientação: Prof. Dr. Manoel Ferreira dos Santos Neto".

Referências.

1. Probabilidade e Estatística. 2. Distribuição Gama Inversa. 3. Zeros Ajustados. 4. Simulações de Monte Carlo. I. Santos Neto, Manoel Ferreira dos. II. Título.

CDU 519.2(043)

# A distribuição gama inversa com zeros ajustados

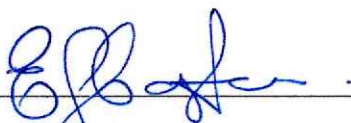
por

**Rafaella Santos Vitorino**

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de Concentração: Probabilidade e Estatística

Aprovada por: 25 de abril de 2024



Prof. Dr. Eliardo Guimarães da Costa - UFRN



Prof. Dr. Juvêncio Santos Nobre - UFC



Prof. Dr. Manoel Ferreira dos Santos Neto - UFCG

Orientador

Universidade Federal de Campina Grande  
Centro de Ciências e Tecnologia  
Programa de Pós-Graduação em Matemática  
Curso de Mestrado em Matemática

Abril - 2024

# Resumo

Neste trabalho, apresentamos uma nova classe de distribuições mistas, composta pela distribuição gama inversa como componente contínuo e uma distribuição Bernoulli como componente discreto. Exploramos várias propriedades dessa distribuição, incluindo a apresentação de resultados inéditos, e obtenção dos estimadores de máxima verossimilhança. Além disso, conduzimos estudos de simulação para avaliar o desempenho dos estimadores propostos em diferentes cenários, nos quais a precisão varia e as proporções de zeros são distintas. Por fim, apresentamos um exemplo com dados reais no qual a distribuição proposta apresentou um ajuste superior à distribuição gama com zeros ajustados.

**Palavras chaves:** Distribuição Gama Inversa; Zeros Ajustados; Simulações de Monte Carlo.

# Abstract

In this work, we introduce a new class of semi-continuous distributions, composed of the inverse gamma distribution as the continuous component and a Bernoulli distribution as the discrete component. We explore various properties of this distribution, including presenting novel results, and derive the maximum likelihood estimators. Additionally, we conduct simulation studies to assess the performance of the proposed estimators in different scenarios, where precision varies and zero proportions differ. Finally, we provide an example with real data in which the proposed distribution demonstrates a superior fit compared to the gamma distribution with adjusted zeros.

**Keywords:** Inverse gamma distribution; Zero Adjusted; Monte Carlo Simulations.

# Agradecimentos

Primeiramente, expresso minha gratidão a Deus por tudo que Ele fez e continua fazendo em minha vida. Sou infinitamente grata pela Sua bondade e misericórdia, por me conceder saúde, pessoas e força para prosseguir e persistir neste projeto; sem Ele, nada disso seria possível. Reconheço que esta jornada não foi fácil, foi um dos períodos mais desafiadores que já vivi. Enfrentei a pandemia, problemas de saúde e, mais doloroso ainda, o falecimento de meu pai. Durante meses, carreguei uma dor intensa durante as aulas do mestrado, sem compartilhar com meus colegas e professores o que estava acontecendo, para não interferir no andamento das atividades. Somente no último dia de aula do semestre compartilhei minha situação, mas sempre mantive minha confiança na bondade de Deus.

Agradeço imensamente à minha mãe, Eurides Vitorino, e ao meu pai, Benício Vitorino (embora não esteja mais presente, sei o quanto ele desejava estar), por estarem sempre ao meu lado em todas as fases da minha vida. Agradeço também à minha irmã por seu constante apoio e incentivo. Sem eles, nada disso seria possível. Agradeço às minhas tias, Diana e Giselle, ao meu noivo Diego, e aos demais familiares e amigos que sempre torceram por mim e estiveram ao meu lado nesta jornada acadêmica.

Expresso minha profunda gratidão ao meu orientador, Professor Dr. Manoel Ferreira dos Santos Neto, que me auxiliou, apoiou e incentivou a persistir. Suas palavras de sabedoria, ensinamentos e até mesmo seus sermões foram fundamentais para o desenvolvimento deste projeto e para minha vida pessoal. Agradeço por me orientar na elaboração deste trabalho, por sua dedicação, paciência e compreensão em todas as etapas deste mestrado e da realização desta dissertação. Que Deus o abençoe infinitamente.

Agradeço também aos pesquisadores Eliardo Guimarães da Costa e Juvêncio Santos Nobre pelos conhecimentos transmitidos e por aceitarem o convite para participarem da banca examinadora. Seus insights e sugestões de correção foram muito valiosos.

Ao colega de mestrado, Adenilson Borba, expresso minha profunda gratidão. Sem sua ajuda, nada disso teria sido possível. Que Deus o abençoe ricamente.

# Dedicatória

”Não me preocupo absolutamente com o futuro. Estou certa de que o bom Deus fará a sua vontade.”

Santa Terezinha



# Sumário

<b>1</b>	<b>Introdução</b>	<b>6</b>
1.1	Objetivos . . . . .	7
1.2	Apresentação dos capítulos . . . . .	7
1.3	Suporte computacional . . . . .	8
1.4	Preliminares . . . . .	8
1.4.1	Distribuições recíprocas . . . . .	8
1.4.2	Distribuição gama inversa . . . . .	10
<b>2</b>	<b>Distribuição gama inversa com zeros ajustados</b>	<b>12</b>
2.1	Definição . . . . .	12
2.2	Inferência . . . . .	14
2.3	Pacote gamlss.inf . . . . .	16
2.4	Simulações . . . . .	20
2.4.1	Cenário 1: Sensibilidade aos Parâmetros . . . . .	21
2.4.2	Cenário 2: Desempenho em Dados com Zeros Ajustados . . . . .	22
<b>3</b>	<b>Aplicação</b>	<b>24</b>
<b>4</b>	<b>Conclusões</b>	<b>27</b>
	<b>Appendix A</b>	<b>28</b>
	<b>A Códigos R</b>	<b>28</b>
	<b>Referências</b>	<b>33</b>

# Lista de Figuras

2.1	Gráfico (a) Função de Densidade, (b) Cumulativa e (c) Quantilica da distribuição. . . . .	20
3.1	Histograma do montante da reivindicação. . . . .	25

# Lista de Tabelas

2.1	Estimativas do viés e raiz do erro quadrático médio do estimadores de máxima verossimilhança da distribuição GIRZA sob o cenário 1. . . . .	22
2.2	Estimativas do viés e raiz do erro quadrático médio do estimadores de máxima verossimilhança da distribuição GIRZA sob o cenário 2. . . . .	23
3.1	Estatística descritivas para a variável <b>claimcst0</b> . . . . .	25
3.2	Estimativas dos parâmetros das distribuições GIRZA e GAZA para a variável <b>claimcst0</b> . . . . .	26

# Capítulo 1

## Introdução

Em contextos, é comum encontrar dados positivos juntamente com zeros amostrais. Chamamos de “zeros amostrais” aqueles casos em que o valor zero é observado de fato, ou seja, há um mecanismo probabilístico gerador. É trabalhado, por exemplo em conjuntos dados de precipitação, dados biométricos, dados ecológicos e de seguros. Uma situação hipotética seria o tempo em que uma pessoa fica parada em um semáforo, por exemplo, se o sinal luminoso estiver verde o tempo de espera foi zero (existe uma probabilidade positiva do evento ocorrer), caso contrário será registrado um valor positivo.

O trabalho pioneiro de [Aitchison \(1955\)](#) introduziu a primeira distribuição com zeros ajustados. Neste trabalho, o autor apresentou uma mistura entre uma distribuição log-normal e uma distribuição degenerada zero. Inicialmente, essa distribuição não recebeu um nome específico; no entanto, em [Aitchison & Brown \(1957\)](#), os autores a chamam de distribuição Delta.

Na literatura existem vários trabalhos na mesma linha de [Aitchison \(1955\)](#). Destacaremos alguns deles, como o de [Ospina & Ferrari \(2010\)](#) propuseram modelos de regressão beta inflacionadas, com a distribuição beta em zero e/ou um, para modelar dados nos intervalos  $[0, 1)$ ,  $(0, 1]$  e  $[0, 1]$ . [Tong et al. \(2013\)](#) retratam um modelo de distribuição gama com zeros ajustados para analisar a perda de empréstimos hipotecários em casos de inadimplência. [Leiva et al. \(2016\)](#) retratam uma metodologia de logística de estoque, modelados por meio da distribuição Birnbaum-Saunders reparametrizada com zeros ajustados, para permitir a mensuração de dados que contenham zeros. [Nobre](#)

et al. (2017) utilizaram a distribuição gama inflacionadas para avaliar o comportamento de pessoas que praticam atividades físicas e prever quantas horas por dia elas gastam treinando. A pesquisa mostrou que há situações em que pessoas não praticam nenhuma atividade física, assim, contendo zeros. Tomazella et al. (2019) retratam uma distribuição mista com um ponto de concentração no zero para a distribuição Birnbaum-Saunders reparametrizada. Cribari-Neto & Santos (2019), retratam distribuições Kumaraswamy inflacionadas em zeros e/ou uns, considerando o intervalo unitário padrão, fechado em zero ou em um, ou ainda, em zero e em um como uma possibilidade em relação ao uso da distribuição beta inflacionadas. Entre outros exemplos.

Neste presente trabalho, será aplicada a distribuição gama inversa, absolutamente contínua, de uma família de dois parâmetros, representados por  $\mu > 0$  e  $\phi > 0$ , respectivamente, os parâmetros de localização e escala. A distribuição gama inversa é obtida através de uma transformação da variável aleatória que faz parte da distribuição gama. Devido o banco de dados apresentar excesso de zeros, há a necessidade de aplicar uma distribuição capaz de modelar os zeros presentes na parte contínua. Na literatura, uma maneira bastante utilizada para modelar dados não negativos, são os modelos como zeros ajustados ou inflacionadas de zeros.

## 1.1 Objetivos

O objetivo geral deste trabalho é introduzir uma nova classe de distribuições, baseada na combinação da classe gama inversa com uma distribuição degenerada no zero. Os objetivos específicos desta pesquisa podem ser descritos da seguinte forma:

- Propor a classe de distribuição gama inversa com zeros ajustados e apresentar a estimação pontual e intervalar;
- Realizar estudos de simulação;
- Aplicar a proposta a um conjunto de dados reais.

## 1.2 Apresentação dos capítulos

Esta dissertação está estruturada em quatro capítulos. No segundo capítulo, é introduzida a classe de distribuições gamma inversa com zeros ajustados, acompanhada

pela discussão da estimação pontual por máxima verossimilhança e a estimação intervalar de seus parâmetros. Em seguida, conduzimos um estudo de simulação para avaliar as propriedades dos estimadores de máxima verossimilhança dos parâmetros. No terceiro capítulo, apresentamos uma aplicação na qual comparamos a distribuição proposta com aquelas previamente descritas na literatura. Por fim, no quarto capítulo, são apresentadas as conclusões finais em relação à proposta desenvolvida.

## 1.3 Suporte computacional

No que diz respeito aos aspectos computacionais, este trabalho foi escrito utilizando o **Overleaf**, que é um editor de texto  $\text{\LaTeX}$  online, baseado em nuvem, utilizado para redigir, editar e publicar documentos, como artigos, documentos e relatórios. No entanto, as simulações, a aplicação e os gráficos foram realizados utilizando a linguagem de programação R, que está disponível para download gratuitamente na plataforma: <https://www.r-project.org/>. Para um estudo mais detalhado sobre a linguagem R, consulte o manual do R desenvolvido por [R Core Team \(2024\)](#).

## 1.4 Preliminares

Nas seções a seguir, apresentamos alguns resultados importantes para uma melhor compreensão da dissertação.

### 1.4.1 Distribuições recíprocas

Todos os resultados abaixo foram retirados de [Escobar & Moreno-Jiménez \(2000\)](#).

**Definição 1.1** *Uma variável aleatória  $X$ , tem suporte em  $\mathfrak{R}^+$ , é recíproca em relação ao ponto  $c$  (RRP)( $c > 0$ ), quando  $X/c$  and  $c/X$  são identicamente distribuídas, ou seja (Escobar & Moreno-Jiménez 2000)*

$$\Pr[X \leq cx] = \Pr[X \geq c/x], \quad \forall x \in \mathfrak{R}^+,$$

em que  $\Pr[\cdot]$  é a medida de probabilidade induzida por  $X$ .

**Teorema 1.2** *Uma variável aleatória  $X$ , tem suporte em  $\mathfrak{R}^+$ , é RRP  $c$  ( $c > 0$ ) se, e somente se,  $F_X(cx) = 1 - F_X(c/x) + \Pr[X = c/x]$ ,  $\forall x \in \mathfrak{R}^+$  (Escobar & Moreno-Jiménez 2000).*

**Prova.** Ver [Escobar & Moreno-Jiménez \(2000\)](#). ■

**Corolário 1.3 :** (i) Uma variável aleatória discreta  $X$ , tem suporte em  $\mathfrak{R}^+$ , é RRP  $c(c > 0)$  se e somente se  $p(cx) = p(c/x)$ ,  $\forall x \in \mathfrak{R}^+$ , em que  $p(x) = \Pr[X = x]$  é a função de probabilidade de  $X$ , e (ii) Uma variável aleatória contínua  $X$ , tem suporte nos  $\mathfrak{R}^+$ , é RRP  $c(c > 0)$  se e somente se  $f_X(cx) = (1/x^2)f_X(c/x)$ ,  $\forall x \in \mathfrak{R}^+$ , em que  $f_X(x)$  é a função densidade de probabilidade de  $X$ .

**Observação 1.1** Um interessante caso é obtido quando  $c = 1$ . Uma variável aleatória  $X$ , tem suporte em  $\mathfrak{R}^+$ , é recíproca (R)(ou recíproca em relação ao ponto  $c = 1$ ), quando  $X$  e  $1/X$  são identicamente distribuídos, ou seja,  $\Pr[X \leq x] = \Pr[X \geq 1/x]$   $\forall x \in \mathfrak{R}^+$ . [Fichtner \(1984\)](#) e [Crawford & Williams \(1985\)](#) chamam este tipo de variável aleatória como “recíproca simétrica” ([Escobar & Moreno-Jiménez 2000](#)): (i) Uma variável aleatória discreta  $X$ , tem suporte em  $\mathfrak{R}^+$ , é recíproca se e somente se  $p(x) = p(1/x)$ ,  $\forall x \in \mathfrak{R}^+$ , e (ii) Uma variável aleatória contínua  $X$ , tem suporte em  $\mathfrak{R}^+$ , é recíproca se e somente se  $f_X(x) = (1/x^2)f_X(1/x)$ ,  $\forall x \in \mathfrak{R}^+$ .

**Definição 1.4** Duas variáveis aleatórias  $X$  e  $Y$ , têm suporte nos  $\mathfrak{R}^+$ , são recíprocas entre si em relação a um ponto  $c$  (RCHRP)( $c > 0$ ), quando  $X/c$  and  $c/Y$  são identicamente distribuídas, ou seja ([Escobar & Moreno-Jiménez 2000](#))

$$\Pr[X \leq cx] = \Pr[Y \geq c/x], \quad \forall x \in \mathfrak{R}^+.$$

**Teorema 1.5** Duas variáveis aleatórias  $X$  e  $Y$ , têm suporte nos  $\mathfrak{R}^+$ , são RCHRP  $c(c > 0)$  se e somente se  $F_X(cx) = 1 - G_Y(c/x) + \Pr[Y = c/x]$ ,  $\forall x \in \mathfrak{R}^+$  ([Escobar & Moreno-Jiménez 2000](#)).

**Prova.** Ver [Escobar & Moreno-Jiménez \(2000\)](#). ■

**Corolário 1.6 :** (i) Duas variáveis aleatórias discretas  $X$  e  $Y$ , têm suporte nos  $\mathfrak{R}^+$ , são RCHRP  $c(c > 0)$  se e somente se  $p_X(cx) = p_Y(c/x)$ ,  $\forall x \in \mathfrak{R}^+$ , em que  $p_X(x)$  e  $p_Y(y)$  são as funções de probabilidade de  $X$  e  $Y$ , respectivamente., e (ii) Duas variáveis aleatórias contínuas  $X$  e  $Y$ , têm suporte nos  $\mathfrak{R}^+$ , são RCHRP  $c(c > 0)$  se e somente se  $f_X(cx) = (1/x^2)g_Y(c/x)$ ,  $\forall x \in \mathfrak{R}^+$ , em que  $f_X(x)$  e  $g_Y(y)$  são as funções densidade de probabilidade de  $X$  e  $Y$ , respectivamente.

**Observação 1.2** Um importante caso é obtido quando  $c = 1$ . Duas variáveis aleatórias  $X$  e  $Y$ , têm suporte nos  $\mathfrak{R}^+$ , são recíprocas entre si (RCH)(ou recíproca entre si em relação ao ponto  $c = 1$ ), quando  $X$  e  $1/Y$  são identicamente distribuídos (por analogia, quando  $1/X$  e  $Y$  são identicamente distribuídos), ou seja,  $\Pr[X \leq x] = \Pr[Y \geq 1/x]$   $\forall x \in \mathfrak{R}^+$ : (i) Duas variáveis aleatórias discretas  $X$  e  $Y$ , têm suporte nos  $\mathfrak{R}^+$ , são recíprocas entre si se e somente se  $p_X(x) = p_Y(1/x)$ ,  $\forall x \in \mathfrak{R}^+$ , e (ii) Duas variáveis aleatórias contínuas  $X$  e  $Y$ , têm suporte nos  $\mathfrak{R}^+$ , são recíprocas entre si se e somente se  $f_X(x) = (1/x^2)g_Y(1/x)$ ,  $\forall x \in \mathfrak{R}^+$ .

## 1.4.2 Distribuição gama inversa

As distribuições gama e gama inversa com parametros  $\alpha$  e  $\beta$ , a função de densidade são dadas, respectivamente por:

$$f_X(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, \quad x > 0, \alpha > 0, \beta > 0,$$

e

$$f_Z(z | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \exp\left\{-\frac{\beta}{z}\right\}, \quad z > 0, \alpha > 0, \beta > 0.$$

Neste trabalho consideramos a versão da distribuição gamma inversa (GI) apresentada em [Bourguignon & Gallardo \(2020\)](#). Nesta versão na distribuição, substituiremos  $\alpha = \phi + 2$  e  $\beta = \mu(1 + \phi)$ . Logo, a sua função de densidade de probabilidade (FDP) e função de distribuição acumulada (FDA) são, respectivamente,

$$f_Y(y | \mu, \phi) = \frac{[\mu(1 + \phi)]^{\phi+2} y^{-(\phi+3)}}{\Gamma(\phi + 2)} \exp\left\{-\frac{\mu(1 + \phi)}{y}\right\}, \quad y > 0, \mu > 0, \phi > 0, \quad (1.1)$$

e

$$F_Y(y | \mu, \phi) = \begin{cases} 0, & \text{se } y \leq 0, \\ \frac{\gamma(\phi+2, \frac{\mu(1+\phi)}{y})}{\Gamma(\phi+2)}, & \text{se } y > 0, \end{cases} \quad (1.2)$$

em que, no numerador a função  $\gamma(a, z) = \int_0^z t^{a-1} \exp\{-t\} dt$  representa a função gama incompleta inferior e no denominador  $\Gamma(a) = \int_0^\infty t^{a-1} \exp\{-t\} dt$  é a função gama. O quociente entre estas funções é conhecido como função gama regularizada.

Outra função importante na caracterização de uma distribuição é a função quantílica (FQ). A FQ é definida como sendo o menor valor de  $y$ , tal que a probabilidade acumulada de  $Y$  até  $y$  seja maior ou igual a  $p$ . Em outras palavras, é o inverso da função de distribuição acumulada. Matematicamente temos que:

$$Q_Y(p) = \inf\{y \in \mathfrak{R} : F_Y(y) \geq p\}.$$

Pontanto, a FQ da distribuição GI pode ser escrita como:

$$Q_Y(p | \mu, \phi) = \begin{cases} 0, & \text{se } p \leq 0; \\ \frac{\mu(1+\phi)}{\gamma^{-1}(\phi+2, \Gamma(\phi+2)p)}, & \text{se } p \in (0, 1), \end{cases}$$

em que  $\gamma^{-1}(\cdot)$  representa a inversa da função gama incompleta inferior  $\gamma(a)$ .

A média e a variância de  $Y$  são respectivamente:

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{\mu^2}{\phi}.$$



Esta distribuição é chamada em [Bourguignon & Gallardo \(2020\)](#) de Gamma Inversa Reparametrizada (GIR). Desta, maneira iremos usar a notação  $Y \sim \text{GIR}(\mu, \phi)$  para indicar que  $Y$  é uma variável aleatória seguindo uma distribuição GIR com média  $\mu > 0$  e parâmetro de precisão  $\phi > 0$ .

**Teorema 1.7** *Se  $Y$  é uma variável aleatória com distribuição GIR, então a mesma não é recíproca simétrica.*

**Prova.** De acordo com a observação (1), uma variável aleatória é recíproca simétrica se e somente se  $Y$  e  $1/Y$  são identicamente distribuídas. Se  $Y$  é uma variável aleatória com distribuição GIR, então  $1/Y$  terá distribuição gama como apresentado em Bourguignon. Portanto, não satisfaz a definição de variável aleatória recíproca simétrica. ■

**Teorema 1.8** *Se  $Y$  é uma variável aleatória com distribuição GIR com média  $\mu$  e precisão  $\phi$  e  $X$  é uma variável aleatória com distribuição gama com parâmetro de forma  $\phi + 2$  e parâmetro de escala  $\mu(1 + \phi)$ , então  $X$  e  $Y$  são variáveis aleatórias recíprocas entre.*

**Prova.** Seja  $X$  uma variável aleatória com distribuição gama com parâmetro de forma  $\phi + 2$  e parâmetro de escala  $\mu(1 + \phi)$ , então

$$\begin{aligned} f_X(y \mid \mu, \phi) &= \frac{[\mu(1 + \phi)]^{\phi+2} y^{\phi+1}}{\Gamma(\phi + 2)} \exp\{-\mu(1 + \phi)y\} \\ &= \frac{y^2}{y^2} \times \frac{[\mu(1 + \phi)]^{\phi+2} y^{\phi+1}}{\Gamma(\phi + 2)} \exp\{-\mu(1 + \phi)y\} \\ &= \frac{1}{y^2} \frac{[\mu(1 + \phi)]^{\phi+2} y^{\phi+3}}{\Gamma(\phi + 2)} \exp\{-\mu(1 + \phi)y\} \\ &= \frac{1}{y^2} f_Y(1/y \mid \mu, \phi), \end{aligned}$$

em que  $f_Y(y \mid \mu, \phi)$  é a função densidade de probabilidade de uma variável aleatória com distribuição GIR com parâmetros  $\mu$  e  $\phi$ . ■

Os resultados apresentados acima complementam os resultados de [Bourguignon & Gallardo \(2020\)](#), pois deixam claro qual tipo de reciprocidade é satisfeita.

## Capítulo 2

# Distribuição gama inversa com zeros ajustados

Neste capítulo, introduzimos uma nova classe de distribuições mistas, composta pela classe gamma inversa como componente contínuo e uma distribuição Bernoulli como componente discreto. Exploramos diversas propriedades dessas distribuições e obtemos os estimadores de máxima verossimilhança. Além disso, conduzimos estudos de simulação com o intuito de avaliar o desempenho dos estimadores obtidos.

O propósito deste capítulo é oferecer uma nova ferramenta na análise de variáveis de distribuições mistas, especialmente aquelas que apresentam assimetria e valores zero.

### 2.1 Definição

Em situações práticas, uma determinada medição pode resultar em um número não negativo. Por exemplo, no caso da precipitação pluviométrica, corresponde à chuva. Quando ocorre chuva, será registrado um número real positivo e quando não ocorre, será registrado zero. Nessa situação, uma distribuição contínua com suporte nos números reais positivos não seria adequada. Se os dados contêm zeros, um modelo natural consiste em adicionar à distribuição contínua um ponto de massa em zero. Desta forma, podemos obter modelos para medições no intervalo  $[0, \infty)$ . Neste contexto, vamos supor que o componente contínuo dos dados é modelado pela distribuição gama inversa reparametrizada. Já o componente discreto, ou seja, o ponto de massa no zero, será

modelado através de uma distribuição degenerada em zero.

Podemos escrever a FDP de uma distribuição com zeros ajustados como:

$$f_Y(y | \mu, \phi, p) = [(1 - p)f_{Y_1}(y)]^{1-I_{\{0\}}(y)} p^{I_{\{0\}}(y)}, \quad y \geq 0.$$

em que  $f_{Y_1}(\cdot)$  é a FDP de uma distribuição contínua com suporte positivo. Portanto, se uma variável aleatória  $Y$  segue uma distribuição Gama inversa com zeros ajustados (GIRZA), com notação  $Y \sim \text{GIRZA}(\mu, \phi, p)$ , sua FDP pode ser expressa como:

$$f_Y(y | \mu, \phi, p) = \left\{ \frac{(1 - p) [\mu(1 + \phi)]^{(\phi+2)} y^{-\phi-3} \exp\left\{-\frac{\mu(1+\phi)}{y}\right\}}{\Gamma(\phi + 2)} \right\}^{1-I_{\{0\}}(y)} \times p^{I_{\{0\}}(y)},$$

com  $y \geq 0, \phi > 0, \mu > 0$  e  $0 < p < 1$ . Além disso, a FDA de  $Y$  é dado por:

$$F_Y(y | \mu, \phi, p) = \begin{cases} p, & \text{se } y = 0; \\ p + (1 - p) \frac{\gamma((\phi+2), \frac{\mu(1+\phi)}{y})}{p(\phi+2)}, & \text{se } y > 0. \end{cases} \quad (2.1)$$

A FQ de  $Y \sim \text{GIRZA}(\mu, \phi, p)$  pode ser obtido pela inversa de (2.1), resultando em:

$$Q_Y(u; \mu, \phi, p) = \begin{cases} 0, & \text{se } p \geq u; \\ Q_{Y_1}((u - p)/(1 - p); \mu, \phi), & \text{se } p < u, \end{cases} \quad (2.2)$$

onde  $0 < u < 1$  e  $Q_{Y_1}(\cdot)$  é a FQ da distribuição GIR dada em (1.1). O gerador de números aleatórios de  $Y \sim \text{ZAGIR}(\mu, \phi, p)$  é dado pelo Algoritmo abaixo.

---

**Algorithm 1** Gerador de números aleatórios para a distribuição GIRZA

---

- 1: Gerar valores uniformes  $u$  de  $U \sim U(0, 1)$ .
  - 2: Definir valores para  $\mu, \phi$  de  $p$  de  $Y \sim \text{GIRZA}(\mu, \phi, p)$ ;
  - 3: Gerar o número pseudoaleatório  $y$  de  $Y \sim \text{GIRZA}(\mu, \phi, p)$  usando (2.2), ou seja,
    - 3.1: Se  $u \leq p$ , então  $y = 0$ ;
    - 3.2: Caso contrário,  $y = Q_{Y_1}([u - p]/[1 - p]; \mu, \phi)$ ;
  - 4: Repita as etapas 1 a 3 até que a quantidade necessária de números pseudoaleatórios seja obtida.
- 

A distribuição GIRZA possui algumas propriedades interessantes. A seguir, destacamos algumas delas.

**Teorema 2.1** *Temos que se  $Y \sim \text{GIRZA}(\mu, \phi, p)$ , então a média e a variância são dadas, respectivamente, por:*

$$E(Y) = (1 - p)\mu \quad e \quad \text{Var}(Y) = p(1 - p)\mu^2 + (1 - p)\frac{\mu^2}{\phi}.$$

**Prova.** Para calcular  $E(Y)$  e  $\text{Var}(Y)$  iremos usar os seguintes resultados de probabilidade: (i)  $E(Y) = E[E(Y | X)]$  e  $\text{Var}(Y) = E[\text{Var}(Y | X)] + \text{Var}[E(Y | X)]$ . Se  $X \stackrel{d}{=} I_{\{0\}}(y)$ , então

$$E(Y | I_{\{0\}}(y)) = \begin{cases} 0, & \text{c.p. } p, \\ \mu, & \text{c.p. } 1 - p, \end{cases}$$

e

$$\text{Var}(Y | I_{\{0\}}(y)) = \begin{cases} 0, & \text{c.p. } p, \\ \frac{\mu^2}{\phi}, & \text{c.p. } 1 - p. \end{cases}$$

Para o caso do valor esperado é fácil ver que:

$$E(Y) = E[E(Y | I_{\{0\}}(y))] = (1 - p)\mu.$$

No caso da variância iremos fazer por partes. Primeiro, note que

$$E[\text{Var}(Y | I_{\{0\}}(y))] = (1 - p)\frac{\mu^2}{\phi},$$

enquanto que

$$\begin{aligned} \text{Var}[E(Y | I_{\{0\}}(y))] &= (1 - p)\mu^2 - [(1 - p)\mu]^2 \\ &= (1 - p)\mu^2 - (1 - p)^2\mu^2 \\ &= (1 - p)\mu^2[1 - 1 + p] = p(1 - p)\mu^2. \end{aligned}$$

Desta maneira, temos que

$$\text{Var}(Y) = E[\text{Var}(Y | X)] + \text{Var}[E(Y | X)] = (1 - p)\frac{\mu^2}{\phi} + p(1 - p)\mu^2.$$

■

**Teorema 2.2** Se  $Y \sim \text{GIRZA}(\mu, \phi, p)$ , então  $V = aY \sim \text{GIRZA}(a\mu, \phi, p)$ , para  $a > 0$ .

**Prova.** Basta notar que, para  $a > 0$ ,  $\Pr(V \leq v) = \Pr(aY \leq v) = \Pr(Y \leq v/a) = F_Y(v/a | \mu, \phi, p) = F_Y(v | a\mu, \phi, p)$ . ■

## 2.2 Inferência

Se  $Y \sim \text{GIRZA}(\mu, \phi, p)$  o logaritmo da função densidade pode ser expressa como

$$\ell(\mu, \phi, p | y) = \ell(p | y) + \ell(\mu, \phi | y),$$

em que

$$\begin{aligned}\ell(p | \mathbf{y}) &= I_{\{0\}}(\mathbf{y}) \log(p) + [1 - I_{\{0\}}(\mathbf{y})] \log(1 - p), \\ \ell(\mu, \phi | \mathbf{y}) &= [1 - I_{\{0\}}(\mathbf{y})] \left[ (\phi + 2) \log(\mu(1 + \phi)) - (\phi + 3) \log(\mathbf{y}) - \frac{\mu(1 + \phi)}{\mathbf{y}} - \log(\Gamma(\phi + 2)) \right].\end{aligned}\quad (2.3)$$

Como pode ser observado em (2.3), o logaritmo da função densidade de probabilidade fatora em um termo que depende apenas de  $p$  e outro que depende de  $\phi$  e  $\mu$ . Conseqüentemente, é feito o método de máxima verossimilhança (MV).

Agora considere a amostra aleatória,  $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ , oriunda de uma distribuição GIRZA( $\mu, \phi, p$ ). Além disso, defina  $B_0 = \{y_i \in \mathbf{y} : y_i = 0, i = 1, 2, \dots, n\}$ ,  $B_+ = \{y_i \in \mathbf{y} : y_i > 0, i = 1, 2, \dots, n\}$ ,  $n_0 = \#B_0$ ,  $n_+ = \#B_+$  e  $n = n_0 + n_+$ . O logaritmo da função de verossimilhança pode ser obtido facilmente a partir de (2.3) da seguinte maneira:

$$\begin{aligned}\mathcal{L}(\mu, \phi, p | \mathbf{y}) &= \sum_{y_i \in B_0} \ell(\mu, \phi, p | y_i) + \sum_{y_i \in B_+} \ell(\mu, \phi, p | y_i) \\ &= \sum_{i=1}^{n_0} \log(p) + \sum_{i=1}^{n_+} \log(1 - p) + (\phi + 2) \sum_{i=1}^{n_+} \log(\mu(1 + \phi)) \\ &\quad - (\phi + 3) \sum_{i=1}^{n_+} \log(y_i) - \mu(1 + \phi) \sum_{i=1}^{n_+} \frac{1}{y_i} - \sum_{i=1}^{n_+} \log(\Gamma(\phi + 2)).\end{aligned}\quad (2.4)$$

Portanto, as derivadas da função log-verossimilhança dada em (2.4) em relação aos parâmetros desconhecidos são:

$$\dot{\ell}_\mu = \frac{n_+(\phi + 2)}{\mu} - n_+(\phi + 1)\tilde{y}_+, \quad (2.5)$$

$$\dot{\ell}_\phi = n_+ \left\{ 1 + \log(\mu) + \log(\phi + 1) + \frac{1}{\phi + 1} - \bar{y}_{l+} - \Psi(\phi + 2) - \mu\tilde{y}_+ \right\}, \quad (2.6)$$

$$\dot{\ell}_p = \frac{n_0}{p} - \frac{n_+}{1 - p}, \quad (2.7)$$

em que  $\tilde{y}_+ = \sum_{i=1}^{n_+} \frac{1}{y_i}$ ,  $\bar{y}_{l+} = \sum_{i=1}^{n_+} \log(y_i)$  e  $\Psi(x) = \frac{d \log(\Gamma(x))}{dx}$ .

Da equação (2.7), temos que a estimativa de MV de  $p$  é  $\hat{p} = n_0/n$ , que representa a proporção de zeros na amostra. Como as soluções das equações  $\dot{\ell}_\mu = 0$  e  $\dot{\ell}_\phi = 0$  obtidas das equações (2.5) e (2.6) não têm forma fechada, a função de log-verossimilhança dada em (2.3) para  $\mu$  e  $\phi$  é maximizada usando algum algoritmo de otimização não-linear. O algoritmo de otimização não-linear denominado Cole-Green (CG) será usado neste caso. Para maiores detalhes veja [Cole & Green \(1992\)](#).

A matriz de informação de Fisher é dada por:

$$\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E} \left[ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] = \begin{pmatrix} \mathcal{I}_{\mu\mu} & \mathcal{I}_{\mu\phi} & 0 \\ \mathcal{I}_{\phi\mu} & \mathcal{I}_{\phi\phi} & 0 \\ 0 & 0 & \mathcal{I}_{pp} \end{pmatrix}, \quad (2.8)$$

em que

$$\begin{aligned} \mathcal{I}_{\mu\mu} &= \frac{n(1-p)(\phi+2)}{\mu^2}, \\ \mathcal{I}_{\mu\phi} &= \mathcal{I}_{\phi\mu} = \frac{n(1-p)}{\mu(\phi+1)}, \\ \mathcal{I}_{\phi\phi} &= n(1-p) \left[ \Psi'(\phi+2) - \frac{\phi}{(1+\phi)^2} \right], \\ \mathcal{I}_{pp} &= \frac{n}{p(1-p)}. \end{aligned}$$

Na matrix (2.8), observa-se que o parâmetro  $p$  é ortogonal aos parâmetros  $\mu$  e  $\phi$ . Segundo [Cox & Hinkley \(1979\)](#), sob algumas condições,  $\hat{\boldsymbol{\theta}}$  é um estimador consistente de  $\boldsymbol{\theta}$  e tem uma distribuição assintótica normal. Então,  $\sqrt{n}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] \xrightarrow{D} N_3(0, \mathcal{J}(\boldsymbol{\theta})^{-1})$ , com,  $n \rightarrow \infty$ , onde,  $\mathcal{J}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{I}(\boldsymbol{\theta})$ . Sendo  $\mathcal{I}(\boldsymbol{\theta})$  a matriz de informação de Fisher dada em (2.8) e  $\xrightarrow{D}$  indica convergência em distribuição. Observa-se que  $\hat{\mathcal{I}}(\boldsymbol{\theta})^{-1}$  é um estimador assintoticamente eficiente e consistente da matriz de variâncias e covariâncias assintótica de  $\hat{\boldsymbol{\theta}}$ . Logo, o inverso da matrix do estimador é dada como,

$$\mathcal{I}(\boldsymbol{\theta})^{-1} = \frac{1}{n} \begin{pmatrix} \frac{\mu^2[(\phi+1)^2\Psi'(\phi+2)-\phi]}{(1-p)[(\phi+2)\Psi'(\phi+2)-1]} & \frac{\mu}{(1-p)(\phi+1)[(\phi+2)\Psi(\phi+2)-1]} & 0 \\ \frac{\mu}{(1-p)(\phi+1)[(\phi+2)\Psi(\phi+2)-1]} & \frac{\mu}{(1-p)[(\phi+2)\Psi'(\phi+2)-1]} & 0 \\ 0 & 0 & p(1-p) \end{pmatrix}.$$

## 2.3 Pacote `gamlss.inf`

Este pacote possibilita o ajuste de distribuições mistas no GAMLSS ([Stasinopoulos 2017](#)), que compreendem componentes tanto contínuos quanto discretos, conforme explicado no Capítulo 5 de [Stasinopoulos \(2017\)](#). Embora existam algumas distribuições mistas na implementação do GAMLSS no R, como beta inflacionada (BEINF), beta inflacionada no zero (BEINF0), beta inflacionada no um (BEINF1) para dados definidos no intervalo  $[0,1]$ , ou gama com zeros ajustados (ZAGA) e normal inversa com zeros ajustados (ZAIG) para dados definidos em uma linha real positiva, as opções são limitadas. Este pacote expande a disponibilidade de distribuições mistas no contexto do GAMLSS.

Para os modelos de ajuste a zero em que a variável resposta está definida nos reais positivos, mas contém zeros nos dados, o ajuste efetivo pode ser alcançado por meio da estimação de dois modelos GAMLSS distintos. Um desses modelos é baseado em uma distribuição nos reais positivos, enquanto o outro é um modelo binomial aplicado aos valores zero e não zero.

A função `gamlssZadj()` é usada para realizar o ajuste das distribuições com zeros ajustados. Além disso, para gerar distribuições ajustadas no zero a partir de distribuições contínuas existentes da família GAMLSS (ou não), definidas nos reais positivos podemos usar a função `gen.Zadj()`.

Para a concepção da nova categoria de distribuição com zeros ajustados, serão empregados os códigos em R fornecidos por [Bourguignon & Gallardo \(2020\)](#) no endereço <https://acesse.dev/codigosartigo>. Os autores implementaram a distribuição IGAMMA2 na família GAMLSS. Dessa forma, a criação da distribuição gama inversa com zeros ajustados torna-se facilmente realizável mediante a utilização da função `gen.Zadj()`. No entanto, foi necessário fazer um pequeno acréscimo nos códigos originais para que a função `gen.Zadj()` funcionasse corretamente. A seguir, é apresentado o código com a parte que foi ajustada.

*Listagem 2.1: Ajuste nos códigos da família IGAMMA2.*

```

1 IGAMMA2 <-function (mu.link = "log", sigma.link = "log")
2 {
3   mstats <- checklink("mu.link", "Inverse Gamma.2", substitute(mu
4     .link),
5     c("inverse", "log", "identity", "own"))
6   dstats <- checklink("sigma.link", "Inverse Gamma.2", substitute
7     (sigma.link),
8     c("inverse", "log", "identity", "own"))
9   structure(list(family = c("IGAMMA2", "Inverse Gamma.2"),
10     parameters = list(mu = TRUE,
11     sigma = TRUE), nopar = 2, type = "Continuous", mu.link = as
12     .character(substitute(mu.link)),
13     sigma.link = as.character(substitute(sigma.link)), mu.
14     linkfun = mstats$linkfun,
15     sigma.linkfun = dstats$linkfun, mu.linkinv = mstats$linkinv
16     ,
17     sigma.linkinv = dstats$linkinv, mu.dr = mstats$mu.eta,
18     sigma.dr = dstats$mu.eta, dldm = function(y, mu, sigma) {

```

```

13     dldm <- (sigma+2)/mu - (sigma+1)/y
14     dldm
15   }, d2ldm2 = function(y, mu, sigma) {
16     d2ldm2 <- -(sigma+2)/mu^2
17     d2ldm2
18   }, dlidd = function(y, mu, sigma) {
19     dlidd <- 1+log(mu)+log1p(sigma)+1/(sigma+1)
20     -digamma(sigma+2)-log(y)-mu/y
21     dlidd
22   }, d2lidd2 = function(y, mu, sigma) {
23     d2lidd2 <- sigma/(sigma+1)^2-psigamma(sigma+2,1)
24     d2lidd2
25   }, d2ldmdd = function(y, mu, sigma) {
26     d2ldmdd <- 1/mu-1/y
27     d2ldmdd
28   }, G.dev.incr = function(y, mu, sigma, ...) -2 * dIGAMMA2(y
29     ,
30     mu, sigma, log = TRUE), rqrres = expression(rqrres(pfun =
31     "pIGAMMA2",
32     type = "Continuous", y = y, mu = mu, sigma=sigma)),
33     mu.initial = expression({
34     mu <- rep(mean(y), length(y))
35   }), sigma.initial = expression({
36     sigma <- rep(((mean(y)^2)/var(y)), length(y))
37   }), mu.valid = function(mu) all(mu > 0), sigma.valid =
38     function(sigma) all(sigma >
39     0), y.valid = function(y) all(y > 0), #modificado
40     mean = function(mu, sigma) mu, #adicionado
41     variance = function(mu, sigma) (1/sigma) *
42     mu^2), #adicionado
43     class = c("gamlss.family",
44     "family"))
45 }

```

Com o código ajustado é possível agora sem erros criar a família de distribuição gama inversa reparametrizada com zeros ajustados. No código abaixo é apresentado o procedimento de criação da distribuição no R.

*Listagem 2.2: Distribuição gama inversa reparametrizada com zeros ajustados.*

```

1 #Programa de Pos-Graduacao em Matematica
2 #Universidade Federal de Campina Grande

```



```

3 #Area de Concentracao: Probabilidade e Estatistica
4 #Orientador: Prof. Dr. Manoel Santos-Neto
5 #Autora: Rafaella Santos Vitorino
6
7 #Pacotes
8 library(gamlss.inf)
9
10 #Carregando a funcoes de Bourguignon and Gallardo (2021)
11 source("stan12221-sup-0003-supinfo.r")
12
13 gen.Zadj(family = "IGAMMA2")

```

O resultado esperado é o apresentado a seguir:

*Listagem 2.3: Saída do R.*

```

1 > #Programa de Pos-Graduacao em Matematica
2 > #Universidade Federal de Campina Grande
3 > #Area de Concentracao: Probabilidade e Estatistica
4 > #Orientador: Prof. Dr. Manoel Santos-Neto
5 > #Autora: Rafaella Santos Vitorino
6 >
7 >
8 > #Pacotes
9 > library(gamlss.inf)
10 >
11 > #Carregando a funcoes de Bourguignon and Gallardo (2021)
12 > source("stan12221-sup-0003-supinfo.r")
13 >
14 > gen.Zadj(family = "IGAMMA2")
15 A zero adjusted IGAMMA2 distribution has been generated
16 and saved under the names:
17 dIGAMMA2Zadj pIGAMMA2Zadj qIGAMMA2Zadj rIGAMMA2Zadj
18 plotIGAMMA2Zadj
19
20 > #Algumas funcao da distribuicao:
21 > dIGAMMA2Zadj() #Funcao de Densidade da Probabilidade
22 > plot(dIGAMMA2Zadj)
23 > pIGAMMA2Zadj() #Funcao de Densidade Cumulativa
24 > plot(pIGAMMA2Zadj)
25 > qIGAMMA2Zadj() #Funcao Quantilica
26 > plot(qIGAMMA2Zadj)

```

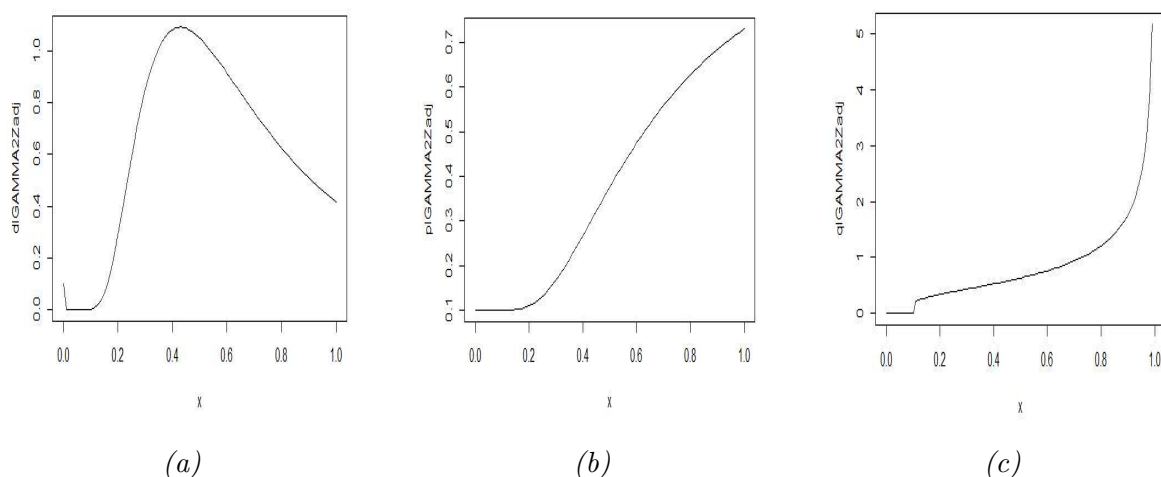


Figura 2.1: Gráfico (a) Função de Densidade, (b) Cumulativa e (c) Quantilica da distribuição.

## 2.4 Simulações

A presente seção tem como objetivo conduzir simulações para investigar o desempenho dos estimadores de máxima verossimilhança da distribuição GIRZA em diferentes cenários controlados. As simulações são uma ferramenta essencial para avaliar a robustez e a eficácia de um modelo estatístico em condições diversas, proporcionando uma compreensão mais aprofundada do comportamento da distribuição em questão.

Ao conduzir essas simulações, buscamos responder a perguntas, incluindo, mas não se limitando a:

- Qual é a sensibilidade da distribuição a diferentes configurações de parâmetros, e como essas variações afetam seus estimadores?
- Em que medida os estimadores da distribuição proposta são afetados pela presença de zeros nos dados, uma característica comum em diversas áreas de aplicação?

Essas perguntas norteiam nossos esforços na realização de simulações, visando proporcionar uma análise abrangente do desempenho da distribuição gama inversa com zeros ajustados. Os resultados obtidos contribuirão significativamente para a compreensão da aplicabilidade prática desta distribuição em contextos estatísticos diversos.

Os resultados da simulação são interpretados com base nas estimativas do: viés

(V) e da raiz do erro quadrático médio (REQM).

$$V(\hat{\theta}) = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{\theta}_i - \theta,$$

$$REQM(\hat{\theta}) = \sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \theta)^2}.$$

em que  $n_{sim}$  é o número de réplicas de Monte Carlo,  $\hat{\theta}$  é o estimador de  $\theta$  e  $\theta$  é o parâmetro de interesse.

### 2.4.1 Cenário 1: Sensibilidade aos Parâmetros

Neste cenário, a sensibilidade aos parâmetros vai avaliar a influência que os diferentes valores do parâmetro tem no resultado, ou seja, serão considerados diferentes valores do parâmetro de precisão da distribuição GIRZA sob diferentes tamanhos amostrais. Os tamanhos de amostra considerados são:  $n = 10$ ,  $n = 40$  e  $n = 160$ . Serão realizadas simulações sob precisão baixa  $\phi = 0,5$ , moderada  $\phi = 2$  e alta  $\phi = 20$ . Sem perda de generalizada iremos considerar uma média unitária, ou seja,  $\mu = 1$ . Além disso, a proporção de zeros na amostra será de 10%. Por fim, foi considerado 5000 réplicas de Monte Carlo.

No aspecto computacional iremos utilizar a função `gamlssZadj()` para a obtenção das estimativas de  $\mu$ ,  $\phi$  e  $p$ . Para a geração dos valores pseudoaleatórios da distribuição GIRZA será usado a função `rIGAMMA2Zadj()` obtida a partir do procedimento descrito no Capítulo 2 desta dissertação.

Para a realização das simulações de Monte Carlo utilizamos a função `MonteCarlo()` do pacote `MonteCarlo` (Leschinski 2019). As estimativas do viés ( $V(\hat{\theta})$ ) e da raiz do erro quadrático médio ( $REQM(\hat{\theta})$ ) foram obtidas usando a função `calc_absolute()` do pacote `simhelpers` (Joshi & Pustejovsky 2022). Os códigos das simulações dos Cenários 1 e 2 são apresentados na Listagem A.1.

Tabela 2.1: Estimativas do viés e raiz do erro quadrático médio do estimadores de máxima verossimilhança da distribuição GIRZA sob o cenário 1.

$n$	$\phi$	Estimativas					
		$V(\hat{\mu})$	$REQM(\hat{\mu})$	$V(\hat{\phi})$	$REQM(\hat{\phi})$	$V(\hat{p})$	$REQM(\hat{p})$
10	0,5	-0,107	0,271	2,020	4,020	0,054	0,093
	2,0	-0,012	0,219	2,510	6,150	0,053	0,090
	20,0	-0,001	0,077	12,800	33,800	0,052	0,090
40	0,5	-0,027	0,147	0,321	0,682	0,000	0,046
	2,0	0,004	0,116	0,325	1,120	0,000	0,045
	20,0	0,000	0,037	2,020	6,360	0,001	0,046
160	0,5	-0,002	0,087	0,059	0,285	-0,000	0,024
	2,0	0,001	0,056	0,082	0,480	0,001	0,024
	20,0	-0,000	0,019	0,478	2,690	-0,000	0,024

Na Tabela 2.1, nota-se, por exemplo, que para  $\phi = 20$  e amostra moderada  $n = 40$ , e grande  $n = 160$ , os estimadores de máxima verossimilhança de  $\mu$  e  $p$  possuem um viés ínfimo. Por outro lado, o estimador de máxima verossimilhança de  $\phi$  mostrou-se com viés elevado, principalmente para  $\phi = 20$ . Com relação à raiz do erro quadrático médio, observa-se que a mesma decai com o aumento da amostra. No entanto,  $\hat{\phi}$  apresentou os maiores valores desta medida. Além disso, é possível notar que o estimador de máxima verossimilhança de  $p$  não sofre influência do aumento da precisão. Por exemplo, para  $n = 10$ , temos que, para  $\phi = 0,5$ ,  $2,0$  e  $20,0$ , os respectivos valores do viés (raiz do erro quadrático médio) são  $0,054(0,093)$ ,  $0,053(0,090)$  e  $0,052(0,090)$ . Este último resultado é coerente com os resultados teóricos apresentados neste trabalho, mas especificamente se deve ao fato do logaritmo da função de verossimilhança ser separável, em outras palavras, o parâmetro  $p$  pode ser estimado de maneira independente dos parâmetros  $\mu$  e  $\phi$ .

## 2.4.2 Cenário 2: Desempenho em Dados com Zeros Ajustados

No Cenário 2, serão explorados distintos valores da proporção de zeros da distribuição GIRZA em diferentes tamanhos amostrais. Os tamanhos de amostra selecionados compreendem  $n = 60$ ,  $n = 120$  e  $n = 180$ . Essa abordagem difere dos valores conside-

rados no Cenário 1, pois a presença de uma quantidade substancial de zeros poderia implicar em desafios no processo de estimação.

As simulações serão conduzidas em cenários caracterizados por proporções de zeros baixas  $p = 0,1$ , moderada  $p = 0,3$  e alta  $p = 0,9$ . Sem perda de generalização, adotaremos uma média igual a 1,0. Além disso, a precisão será mantida fixa em 20. Adicionalmente, serão executadas 5000 réplicas de Monte Carlo para cada configuração simulada.

*Tabela 2.2: Estimativas do viés e raiz do erro quadrático médio dos estimadores de máxima verossimilhança da distribuição GIRZA sob o cenário 2.*

$n$	$p$	Estimativas					
		$V(\hat{\mu})$	$REQM(\hat{\mu})$	$V(\hat{\phi})$	$REQM(\hat{\phi})$	$V(\hat{p})$	$REQM(\hat{p})$
60	0.1	0,000	0,030	1,230	4,860	0,000	0,039
	0.3	-0,001	0,035	1,750	5,740	-0,001	0,059
	0.9	-0,002	0,099	27.810,000	249.276,000	-0,001	0,038
120	0.1	-0,000	0,022	0,584	3,140	0,000	0,028
	0.3	-0,001	0,025	0,844	3,730	0,000	0,043
	0.9	0,000	0,069	8,620	24,000	0,000	0,027
180	0.1	-0,000	0,018	0,466	2,560	-0,000	0,023
	0.3	-0,000	0,020	0,566	2,940	-0,000	0,035
	0.9	0,001	0,053	4,910	13,200	0,000	0,023

Na Tabela 2.2, nota-se que a principal influência do aumento da proporção de zeros, ou seja, na redução de valores positivos na amostra, é sobre o estimador de máxima verossimilhança do parâmetro  $\phi$ . Por exemplo, para  $p = 0,9$  e  $n = 60$ , temos que o viés e a raiz do erro quadrático médio de  $\hat{\phi}$  são, respectivamente, 27810,00 e 249276,00. Para os demais estimadores, o impacto maior é apenas na raiz do erro quadrático médio. Por exemplo, para  $n = 60$  e  $p = 0,1, 0,3$  e  $0,9$ , temos os respectivos valores,  $REQM(\hat{\mu}) = 0,030, 0,035$  e  $0,035$ .

Podemos observar nas simulações que o estimador do parâmetro de precisão necessita de atenção com relação ao seu viés e à sua sensibilidade à presença de muitos zeros na amostra. Isso pode ser motivo de estudos em pesquisas futuras.

# Capítulo 3

## Aplicação

Neste Capítulo, será apresentada uma ilustração com um conjunto de dados reais. Este conjunto de dados é baseado em apólices de seguro de veículos com duração de um ano contratadas em 2004 ou 2005. O conjunto de dados foi obtido através do pacote R `insuranceData`. Para acessar os dados basta usar o comando `data("dataCar")`.

O conjunto de dados tem 67.856 observações com as seguintes 11 variáveis:

- **veh\_value**: Valor do veículo, em \$10.000
- **exposure**: 0-1
- **clm**: Ocorrência de reivindicação (0 = não, 1 = sim)
- **numclaims**: Número de reivindicações
- **claimcst0**: Montante da reivindicação (0 se não houver reivindicação)
- **veh\_body**: Tipo de veículo, codificado como BUS, CONVT, COUPE, HBACK, HDTOP, MCARA, MIBUS, PANVN, RDSTR, SEDAN, STNWG, TRUCK, UTE
- **veh\_age**: 1 (o mais jovem), 2, 3, 4
- **gender**: Um fator com os níveis F (feminino) M (masculino)
- **area**: Um fator com os níveis A, B, C, D, E, F
- **agecat**: 1 (o mais jovem), 2, 3, 4, 5, 6
- **X\_OBSTAT\_**: Um fator com os níveis 01101 0 0 0

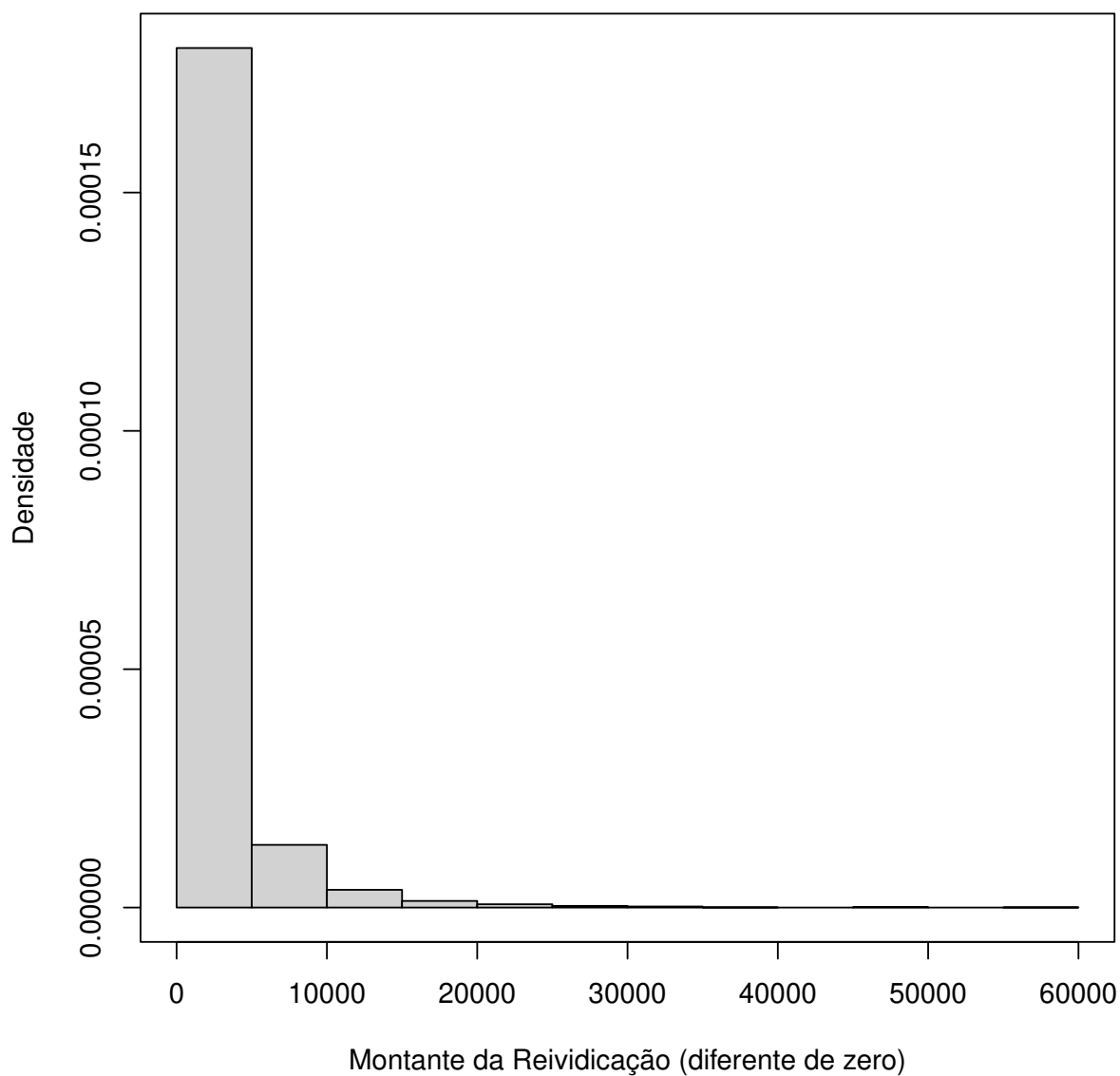
Aqui iremos estudar a variável **claimcst0**. A seguir, na Tabela 3.1, apresentamos algumas medidas descritivas para esta variável. Nota-se observando esta Tabela que a

variável é composta na sua maior parte por zeros. Na Figura 3.1 observa-se que a parte positiva da variável estudada possui forte assimetria à direita.

Tabela 3.1: Estatística descritivas para a variável *claimcst0*.

Variável	$n$	$n_0$	$n_+$	Média	Dev. Padrão	Min	Pec. 25	Pec. 75	Max
claimcst0	67856	63232	4624	137	1056	0	0	0	55922

Figura 3.1: Histograma do montante da reivindicação.



Na Tabela 3.2 são apresentadas as estimativas dos parâmetros que indexam as distribuições GIRZA e GAZA. As estimativas de máxima verossimilhança foram obtidas através do comando `gamlssZadj()` do pacote `gamlss.inf`. Os códigos da aplicação são

apresentados na Listagem A.2. Importante destacar que o comando `gamlssZadj()` por padrão retorna as estimativas de  $\mu$ ,  $\log(\phi)$  e  $\text{logit}(p)$ . Por exemplo, a estimativa de zeros  $\hat{p} = 0,9319$ . Note que se consideramos como critério para a escolha do melhor ajuste o AIC generalizado a distribuição proposta nesta dissertação apresenta um ajuste superior ao apresentado pela distribuição GAZA.

*Tabela 3.2: Estimativas dos parâmetros das distribuições GIRZA e GAZA para a variável `claimcst0`.*

Parâmetros	GIRZA	GAZA
$\mu$	1058.41*** (11.01)	2014.40*** (34.20)
$\log(\phi)$	-36.04 (1470.59)	0.14*** (0.01)
$\text{logit}(p)$	2.62*** (0.02)	2.62*** (0.02)
Num. obs.	67856	67856
AIC Generalizado	112391.82	113098.64

\*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$



# Capítulo 4

## Conclusões

Na presente dissertação, foi proposta uma nova classe de distribuições gama inversa com zeros ajustados, sendo a estrutura da nova classe composta por uma componente discreta distribuída por uma distribuição Bernoulli e uma componente contínua, sendo esta última a distribuição gama inversa. Foi apresentada a estimação pontual e intervalar dos parâmetros e um estudo de simulação sob diferentes cenários concluiu que os estimadores de máxima verossimilhança para os parâmetros da componente contínua podem ser afetados pela proporção de zeros, mas apresentam um bom desempenho para amostras grandes.

Por último, no contexto prático, foi discutido um exemplo com um conjunto de dados de seguro onde a distribuição proposta apresentou um bom ajuste aos dados. Além disso, quando comparada à distribuição GAZA, a distribuição proposta mostrou um ajuste superior.

Como trabalhos futuros, é possível criar um modelo de regressão baseado na distribuição proposta usando diferentes abordagens. Por exemplo, é possível considerar estruturas semiparamétricas. Na parte inferencial, é possível obter estimativas usando o paradigma bayesiano. Além disso, também é possível propor um modelo de regressão quantílica.

# Apêndice A

## Códigos R

*Listagem A.1: Simulações de Monte Carlo.*

```
1 #Programa de Pos-Graduacao em Matematica
2 #Universidade Federal de Campina Grande
3 #Area de Concentracao: Probabilidade e Estatistica
4 #Orientador: Prof. Dr. Manoel Santos-Neto
5 #Autora: Rafaella Santos Vitorino
6
7
8 #Pacotes
9 library(gamlss.inf)
10 library(MonteCarlo)
11 library(tidyverse)
12 library(simhelpers)
13 library(boot)
14 #Carregando a funcoes de Bourguignon and Gallardo (2021)
15 source("stan12221-sup-0003-supinfo.r")
16
17 #Gerando a distribuicao GIZA
18 gen.Zadj(family = "IGAMMA2")
19
20
21 # Funcoes auxiliares
22 my.gamlssZadj <- function(...) tryCatch(expr = gamlssZadj(...),
23   error = function(e) NA)
24
25 est <- function(n, mu, phi, p){
```

```

25
26 repeat{
27   y <- rIGAMMA2Zadj(n, mu = mu, sigma = phi, xi0 = p)
28   fit <- my.gamlssZadj (y = y, mu.formula = ~1,
29                       xi0.formula = ~1,
30                       family = IGAMMA2(mu.link = "identity",
31                                       sigma.link = "identity")
32                               )
33   if (all(is.na(fit)) == FALSE) break
34
35 }
36
37 coef_mle <- unname(c(fit$mu.coefficients, fit$sigma.coefficients
38                     , fit$xi0.coefficients))
39
40 return(list("mu0" = coef_mle[1], "phi0" = coef_mle[2], "p0" = inv
41           .logit(coef_mle[3]) ))
42 }
43
44 #Simulacao - Cenario 1
45
46 n_grid <- c(10, 40, 160)
47 mu_grid <- c(1.0)
48 phi_grid <- c(0.5, 2, 20)
49 p_grid <- c(0.1)
50 param_list <- list("n" = n_grid,
51                  "mu" = mu_grid,
52                  "phi" = phi_grid,
53                  "p" = p_grid)
54
55 MC_result <- MonteCarlo(func = est,
56                       nrep = 5000,
57                       param_list = param_list)
58
59 df <- MakeFrame(MC_result)
60 resul_mu <- df |> group_by(n, mu, phi, p) |>
61   do(calc_absolute(., estimates = mu0,
62                   true_param = mu,
63                   perfm_criteria = c("bias", "rmse")))
64
65 resul_phi <- df |> group_by(n, mu, phi, p) |>

```



*Listagem A.2: Aplicação.*

```
1 #Programa de Pos-Graduacao em Matematica
2 #Universidade Federal de Campina Grande
3 #Area de Concentracao: Probabilidade e Estatistica
4 #Orientador: Prof. Dr. Manoel Santos-Neto
5 #Autora: Rafaella Santos Vitorino
6
7
8 #Pacotes
9 library(gamlss.inf)
10 library(tidyverse)
11 library(ggpubr)
12 library(boot)
13
14 #Carregando a funcoes de Bourguignon and Gallardo (2021)
15 source("stan12221-sup-0003-supinfo.r")
16
17 #Gerando a distribuicao GIZA
18 gen.Zadj(family = "IGAMMA2")
19
20 #Importar dados
21 library(insuranceData)
22 library(vtable)
23 library(texreg)
24
25 data("dataCar")
26
27 data_fit <- data.frame(y =dataCar$claimcst0)
28
29 sumtable(data_fit,
30           align = 'p{.3\\textwidth}cccccc',
31           fit.page = '\\textwidth',
32           out = 'latex')
33
34 gghistogram(dataCar,
35             x = "claimcst0",
36             xlab = "Montante da reivindicacão",
37             ylab = "Frequencia")
38
39 ggsave("dataCar.pdf")
40
41 fit <- gamlssZadj(claimcst0,
```

```
42     mu.formula = ~1,
43     xi0.formula = ~1,
44     sigma.formula = ~1,
45     family = IGAMMA2(mu.link = "identity",
46                     sigma.link = "log"),
47     data = dataCar)
48 inv.logit(fit$xi0.coefficients)
49
50
51
52 fit1 <- gamlssZadj(claimcst0,
53                 mu.formula = ~1,
54                 xi0.formula = ~1,
55                 sigma.formula = ~1,
56                 family = GA(mu.link = "identity",
57                             sigma.link = "log"),
58                 data = dataCar)
59 inv.logit(fit1$xi0.coefficients)
60
61
62 texreg(list(fit, fit1))
```

# Referências Bibliográficas

Aitchison, J. (1955), 'On the Distribution of a Positive Random Variable Having a Discrete Probability Mass at the Origin', *Journal of the American Statistical Association* **50**(271), 901.

**URL:** <https://www.jstor.org/stable/2281175?origin=crossref>

Aitchison, J. & Brown, J. A. C. (1957), *The lognormal distribution: with special reference to its uses in economics*, number 5 in 'University of Cambridge. Dept. of Applied Economics. Monographs', University Press, Cambridge.

Bourguignon, M. & Gallardo, D. I. (2020), 'Reparameterized inverse gamma regression models with varying precision', *Statistica Neerlandica* **74**(4), 611–627.

**URL:** <https://onlinelibrary.wiley.com/doi/10.1111/stan.12221>

Cole, T. J. & Green, P. J. (1992), 'Smoothing reference centile curves: The lms method and penalized likelihood', *Statistics in Medicine* **11**(10), 1305–1319.

**URL:** <https://onlinelibrary.wiley.com/doi/10.1002/sim.4780111005>

Cox, D. & Hinkley, D. (1979), *Theoretical Statistics*, 0 edn, Chapman and Hall/CRC.

**URL:** <https://www.taylorfrancis.com/books/9781482214925>

Crawford, G. & Williams, C. (1985), 'A note on the analysis of subjective judgment matrices', *Journal of Mathematical Psychology* **29**(4), 387–405.

**URL:** <https://linkinghub.elsevier.com/retrieve/pii/0022249685900021>

Cribari-Neto, F. & Santos, J. (2019), 'Inflated Kumaraswamy distributions', *Anais da Academia Brasileira de Ciências* **91**(2), e20180955.

**URL:** [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0001-37652019000300201&lng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0001-37652019000300201&lng=en)

Escobar, M. & Moreno-Jiménez, J. (2000), ‘Reciprocal distributions in the analytic hierarchy process’, *European Journal of Operational Research* **123**(1), 154–174.

**URL:** <https://linkinghub.elsevier.com/retrieve/pii/S0377221799000867>

Fichtner, J. (1984), *Some Thoughts about the Mathematics of the Analytic Hierarchy Process*, Bericht // Hochschule der Bundeswehr München, Fachbereich Informatik, Inst. für Angewandte Systemforschung u. Operations-Research.

**URL:** <https://books.google.com.br/books?id=bCbicQAACAAJ>

Joshi, M. & Pustejovsky, J. (2022), ‘simhelpers: Helper Functions for Simulation Studies’.

**URL:** <https://meghapsimatrix.github.io/simhelpers/index.html>

Leiva, V., Santos-Neto, M., Cysneiros, F. J. A. & Barros, M. (2016), ‘A methodology for stochastic inventory models based on a zero-adjusted Birnbaum-Saunders distribution’, *Applied Stochastic Models in Business and Industry* **32**(1), 74–89.

**URL:** <https://onlinelibrary.wiley.com/doi/10.1002/asmb.2124>

Leschinski, C. H. (2019), ‘MonteCarlo: Automatic Parallelized Monte Carlo Simulations’.

**URL:** <https://CRAN.R-project.org/package=MonteCarlo>

Nobre, A. A., Carvalho, M. S., Griep, R. H., Fonseca, M. D. J. M. D., Melo, E. C. P., Santos, I. D. S. & Chor, D. (2017), ‘Multinomial model and zero-inflated gamma model to study time spent on leisure time physical activity: an example of ELSA-Brasil’, *Revista de Saúde Pública* **51**, 76.

**URL:** <https://www.revistas.usp.br/rsp/article/view/138333>

Ospina, R. & Ferrari, S. L. P. (2010), ‘Inflated beta distributions’, *Statistical Papers* **51**(1), 111–126.

**URL:** <http://link.springer.com/10.1007/s00362-008-0125-4>

R Core Team (2024), ‘R: A Language and Environment for Statistical Computing’.

**URL:** <https://www.R-project.org/>



Stasinopoulos, M. D., ed. (2017), *Flexible regression and smoothing: using GAMLSS in R*, CRC Press/Taylor & Francis Group, Boca Raton.

Tomazella, V., Pereira, G. H., Nobre, J. S. & Santos-Neto, M. (2019), ‘Zero-adjusted reparameterized Birnbaum–Saunders regression model’, *Statistics & Probability Letters* **149**, 142–145.

**URL:** <https://linkinghub.elsevier.com/retrieve/pii/S0167715219300252>

Tong, E. N., Mues, C. & Thomas, L. (2013), ‘A zero-adjusted gamma model for mortgage loan loss given default’, *International Journal of Forecasting* **29**(4), 548–562.

**URL:** <https://linkinghub.elsevier.com/retrieve/pii/S0169207013000447>