UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Distribuições discretas para duas observações inflacionadas

Luisa Hebling

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

ICMC USP
SÃO CARLOS

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Luisa Hebling

Distribuições discretas para duas observações inflacionadas

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestra em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientadora: Profª. Dra. Katiane Silva Conceição

USP – São Carlos
Agosto de 2021

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

Hebling, Luisa
H443d Distribuições discretas para duas observações inflacionadas / Luisa Hebling; orientadora Katiane Silva Conceição. -- São Carlos, 2021.
93 p.

Dissertação (Mestrado - Programa Interinstitucional de Pós-graduação em Estatística) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2021.

1. Dados de contagem. 2. Dados inflacionados. 3. Distribuição Série de Potência. 4. Distribuição hurdle. 5. Medidas de evidências. I. Conceição, Katiane Silva, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Luisa Hebling

Discrete distributions for two inflated observations

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics.
FINAL VERSION

Concentration Area: Statistics

Advisor: Profª. Dra. Katiane Silva Conceição

USP – São Carlos
August 2021

Este trabalho é dedicado à todos que direta ou indiretamente fizeram parte da minha trajetória.

Em especial, aos meus familiares, amigos e orientadora.

AGRADECIMENTOS

Agradeço a Deus por me dar saúde e força para superar as dificuldades e concluir esta etapa da minha vida.

Aos meus pais Nilton e Elisa, por todo amor, incentivo e apoio incondicional. Me fizeram entender que o futuro é feito a partir da constante dedicação no presente. Às minhas irmãs Flavia e Patricia, pelo amor, amizade e por me escutarem, principalmente, nos momentos mais difíceis. A minha avó Ines, pelo amor incondicional e por todos seus ensinamentos. E um agradecimento especial ao meu avô Nilton (*in memoriam*), que sempre foi meu exemplo de sabedoria, determinação, honestidade e humildade.

À todos meus amigos com quem tive o prazer de conviver e que, de alguma forma, fizeram parte dessa minha jornada.

À minha orientadora Profª Dra. Katiane, por todo conhecimento, oportunidade, apoio, suporte, incentivo e, principamente, por toda credibilidade concedida.

Ao Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC - USP) e ao Departamento de Estatística da Universidade Federal de São Carlos (DEs - UFSCar), juntamente com seu corpo docente, direção e administração, pela oportunidade de expandir meus conhecimentos e por proporcionar o melhor dos ambientes para que esse trabalho fosse realizado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“All our dreams can come true
if we have the courage to pursue them.”
(Walt Disney)

RESUMO

HEBLING, L. **Distribuições discretas para duas observações inflacionadas**. 2021. 92 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Dados de contagem são frequentemente encontrados em muitas aplicações reais e algumas observações podem ocorrer no conjunto de dados em uma quantidade excessiva. Em muitos problemas reais é bastante comum o conjunto de dados contenha excessos de observações com valores zero e um. Em um contexto mais geral, define-se k_1 e k_2 como observações de um particular conjunto de dados que apresentam discrepância (excesso) nas suas frequências, tornando a modelagem a partir de distribuições discretas tradicionais inadequada. Assim, o principal objetivo deste trabalho é propor a família de distribuições Série de Potência k_1 e k_2 Inflacionada, visando modelar adequadamente conjuntos de dados que apresentam discrepância nas observações k_1 e k_2. Para estimação dos parâmetros consideramos a abordagem clássica, com o método da máxima verossimilhança, utilizando a versão *hurdle* das distribuições. Algumas aplicações envolvendo conjuntos de dados reais serão apresentadas.

Palavras-chave: Dados de contagem, Dados inflacionados, Distribuição Série de Potência, Distribuição *hurdle*, Medidas de evidências.

ABSTRACT

HEBLING, L. **Discrete distributions for two inflated observations**. 2021. 92 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Count data is often found in many real applications and some observations may occur in the data set in an excessive amount. In many real problems it is quite common for the data set to contain excesses of zero and one observations. In a more general context, k_1 and k_2 are defined as observations of a particular data set that have discrepancy (excess) in their frequencies, making modeling from traditional discrete distributions inappropriate. Thus, the main objective of this work is to propose the k_1 and k_2 Inflated Power Series family of distributions, aiming to model data sets that present such discrepancy in the observations k_1 and k_2. In order to estimate the parameters we consider classical approach, with the maximum likelihood method, using a hurdle version of distributions. Some applications considering real data sets will be presented.

Keywords: Count data, Inflated data, Power Series Distribution, Hurdle distribuition, Evidence measures.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

$\boldsymbol{k}$-IPS	Série de Potência k_1 e k_2 Inflacionada
$\boldsymbol{k}$-SPS	Série de Potência k_1 e k_2 Subtraída
k-DPS	Série de Potência k Deflacionada
k-IPS	Série de Potência k Inflacionada
k-MPS	Série de Potência k Modificada
k-SPS	Série de Potência k Subtraída
B	Binomial
BN	Binomial Negativa
EM	expectativa-maximização
EMV	estimador de máxima verossimilhança
FMP	função massa de probabilidade
G	Geométrica
P	Poisson
PS	Série de Potência

SUMÁRIO

CAPÍTULO

1

INTRODUÇÃO

Na estatística aplicada e em várias áreas do conhecimento é muito comum algumas observações assumirem valores naturais (inteiros positivos), os quais são chamados de dados de contagem. Geralmente, a distribuição Poisson é muito utilizada para analisá-los. No entanto, em algumas situações, a variância pode ser maior ou menor que a média, sendo os dados referenciados como super-dispersos ou sub-dispersos, respectivamente, em relação a distribuição Poisson. Consequentemente, distribuições discretas alternativas têm sido propostas, como a família de distribuições Série de Potência, que constitui uma extensa família de distribuições para dados de contagem, que são amplamente utilizadas.

As pesquisas voltadas à estas distribuições vêm sendo desenvolvidas há um tempo por diversos pesquisadores. Khatri (1959) apresentou uma generalização da distribuição Série de Potência e determinou algumas propriedades relacionadas a esta. Seguindo o mesmo contexto, Gupta (1974) definiu a distribuição Série de Potência Modificada e apresentou algumas de suas propriedades envolvendo recorrência entre momentos, apresentando uma aplicação da teoria proposta utilizando a distribuição Binomial Negativa. Jani (1978), por sua vez, abordou o método bayesiano para a estimação dos parâmetros da distribuição Série de Potência Modificada e apresentou alguns exemplos envolvendo conjunto de dados reais. Em Cordeiro, Andrade e Castro (2009) já é possível encontrarmos uma extensão da teoria de modelos de regressão voltada as distribuições Série de Potência e uma aplicação para tal.

Em situações práticas pode acontecer que uma determinada observação do conjunto de dados ocorra com uma frequência maior ou menor do que a esperada ao considerar uma determinada distribuição discreta, o que torna a suposição desta distribuição inadequada. Isso tem ocorrido na prática principalmente quando a discrepância ocorre na classe de contagem de zero. Quando a frequência da observação zero no conjunto de dados é maior (ou menor) do que a esperada, ao considerar uma distribuição discreta, o conjunto de dados é dito ser zero inflacionado (ou zero deflacionado). Em Conceição *et al.* (2017) são apresentadas as distribuições Série de

Potência Zero Modificada, as quais modelam adequadamente a discrepância na frequência de zero, sem que seja necessário qualquer conhecimento prévio sobre o tipo de modificação (inflação ou deflação) presente nos dados. Ainda considerando o trabalho destes autores, um estudo sobre conjuntos de dados reais inflacionados e deflacionados no zero também podem ser encontrados.

Em um contexto mais geral, existem situações em que a discrepância entre a frequência observada e a esperada, segundo uma distribuição discreta, pode ocorrer em alguma observação diferente de zero, como por exemplo em uma observação k (identificado previamente). Visando esse conceito, Murat e Szynal (1998) generalizaram as distribuições apresentadas inicialmente por Gupta, Gupta e Tripathi (1995), estendendo as distribuições discretas inflacionadas para qualquer ponto k. Carvalho (2017), por sua vez, também estendeu esta ideia, fornecendo uma ampla família de distribuições discretas, denominada Série de Potência k Modificada, a qual inclui k-inflação e k-deflação e tem como caso particular a família de distribuição Série de Potência Zero Modificada (quando $k = 0$).

Por outro lado, um conjunto de dados pode apresentar discrepância na frequência de duas observações, ditas k_1 e k_2. Conjuntos de dados deflacionados de algum valor k_1 e k_2 são obtidos em experimentos em que estes valores são observados com uma frequência significativamente mais baixa do que a esperada, com base em uma distribuição discreta usual. Em contrapartida, dados inflacionados de algum valor k_1 e k_2 são obtidos quando são observados com uma frequência significativamente mais alta do que a esperada. No geral, a inflação (ou deflação) de uma determinada observação pode ser causada por alguma peculiaridade do processo que provoca o ganho (ou perda) deste valor específico.

Na literatura, a inflação de observações vem sendo muito estudada por diversos pesquisadores. Melkersson e Olsson (1999) estenderam a ideia da distribuição Poisson Zero e Um Inflacionados, apresentando uma aplicação de um conjunto de dados reais para posteriormente comparar os resultados obtidos entre esta distribuição e a Poisson Zero Inflacionada. Saito e Rodrigues (2005) apresentaram o método bayesiano para a estimação dos parâmetros da distribuição Poisson Zero e Um Inflacionados, a partir dos dados ampliados e, por fim, mostra um exemplo da metodologia proposta. Por sua vez, Alshkaki (2016) utiliza os métodos clássicos de estimação, máxima verossimilhança e momentos, os quais são comparados em algumas aplicações.

Buscando estender a ideia da inflação em dois pontos, iremos propor a família de distribuições discretas para duas observações inflacionadas, ditas k_1 e k_2, denominada por distribuição Série de Potência k_1 e k_2 Inflacionada, que é capaz de modelar conjuntos de dados que apresentam ou não uma alta frequência das observações k_1 e k_2, simultaneamente, em que k_1 e k_2 são diferentes entre si. Além disso, iremos utilizar uma abordagem clássica, via método da máxima verossimilhança, para a estimação dos parâmetros de interesse desta distribuição. As distribuições discretas para duas observações inflacionadas que iremos considerar são Poisson (P), Geométrica (G) e Binomial (B), isto é, estudaremos as seguintes distribuições: Poisson k_1 e k_2 Inflacionada, Geométrica k_1 e k_2 Inflacionada e Binomial k_1 e k_2 Inflacionada.

Dessa forma, este trabalho tem como objetivo estender a ideia de modificação (inflação) na probabilidade de uma única observação (k) para duas observações (k_1 e k_2), a qual a denominamos família de distribuições Série de Potência k_1 e k_2 Inflacionada.

Para completar a metodologia, descrevemos e apresentamos algumas propriedades e características importantes desta nova família de distribuições. Para fins de comparações, demonstramos as relações entre a família de distribuições proposta (Série de Potência k_1 e k_2 Inflacionada) e as distribuições Série de Potência.

Para a estimação dos parâmetros consideramos o método da máxima verossimilhança e utilizamos a versão *hurdle* da distribuição para nos auxiliar e facilitar no desenvolvimento dos cálculos para a obtenção dos estimadores. Verificamos ainda seus desempenhos e avaliamos as propriedades assintóticas a partir de um estudo de simulação.

Por fim, para medir a eficácia e a qualidade do ajuste das distribuições Série de Potência k_1 e k_2 Inflacionada, analisamos sete conjuntos de dados reais com altas frequências em duas observações distintas entre si: em três deles utilizamos a distribuição Poisson Zero e Um Inflacionados, em um deles ajustamos a Geométrica Zero e Um Inflacionados e outro a Geométrica Zero e Oito Inflacionados, por fim, nos dois demais, utilizamos a Binomial Um e Dois Inflacionados e a Binomial Zero e Dois Inflacionados. Em cada aplicação, apresentamos as estimativas dos parâmetros obtidas via método clássico, realizamos um estudo comparativo entre os ajustes das distribuições mais simples (distribuições Série de Potência e as distribuições Série de Potência k Modificada) via Teste de Aderência Kolmogorov-Smirnov e pelas medidas de evidências Distância Euclidiana (*DE*), Divergência de Kullback-Leibler (*KL*) e Divergência de Kullback-Leibler Simétrica (*KLS*).

Este trabalho é organizado da seguinte forma: no Capítulo 2 apresentamos os conceitos e notações preliminares, como a família de distribuições Série de Potência e a família de distribuições Série de Potência k Modificada, dando um foco maior para o caso particular da k-inflação, isto é, a distribuição Série de Potência k Inflacionada, juntamente com suas propriedades e a reparametrização para a versão *hurdle*; no Capítulo 3 definimos a família de distribuições Série de Potência k_1 e k_2 Inflacionada, as propriedades e a versão *hurdle* das distribuições, além de apresentar uma análise do comportamento probabilístico; no Capítulo 4 apresentamos o procedimento utilizado para a estimação dos parâmetros, considerando a abordagem clássica (via método da máxima verossimilhança); no Capítulo 5 apresentamos o estudo de simulação e aplicações envolvendo dados artificiais; no Capítulo 6 apresentamos as aplicações de alguns conjuntos de dados reais para as distribuições Poisson, Geométrica e Binomial k_1 e k_2 Inflacionada. Por fim, apresentamos no Capítulo 7 algumas considerações finais a respeito da metodologia proposta, bem como as conclusões estabelecidas neste trabalho.

O recurso computacional utilizado nesta dissertação foi o *Software* R (*versão 1.2.5019*) (R Core Team, 2019), incluindo implementação de códigos computacionais executados em um computador com processador Intel(R) Core(TM) i5-3337U CPU @ 1,80GHz e 4GB de RAM e

sistema operacional Windows 8.1 de 64 bits.

CAPÍTULO

2

CONCEITOS E NOTAÇÕES PRELIMINARES

Neste capítulo apresentamos alguns conceitos e notações que são necessários para o desenvolvimento desta dissertação, como a família de distribuições Série de Potência e a família de distribuições Série de Potência k Modificada. Iremos ainda destacar o caso particular da k-inflação, a qual é chamada de distribuição Série de Potência k Inflacionada, e apresentar sua versão *hurdle*, juntamente com as suas propriedades.

2.1 Distribuição Série de Potência Uniparamétrica

Considere X uma variável aleatória inteira e não negativa. Se X tem distribuição Série de Potência (PS) [1] uniparamétrica com parâmetro de média μ ($\mu > 0$), então sua função massa de probabilidade (FMP) é dada por:

$$\pi_{PS}(x;\mu) = \frac{a(x)[g(\mu)]^x}{f(\mu)},\ \forall\, x \in \mathcal{A}_s, \tag{2.1}$$

em que $\mathcal{A}_s$ é o suporte do subconjunto dos inteiros $\{s, s+1, \ldots\}$, com $s \geq 0$; $a(x)$ é uma função positiva; $f(\mu)$ e $g(\mu)$ também são funções positivas, finitas e duas vezes diferenciáveis, sendo $f(\mu) = \sum_{x\in\mathcal{A}_s} a(x)[g(\mu)]^x$.

A Tabela 1 exibe as funções a, g e f das distribuições da família PS utilizadas nesta dissertação, cujo suporte se inicia em $s = 0$.

[1] Do inglês "*Power Series*"

Tabela 1 – Algumas distribuições da família PS.

PS	Distribuição	$a(x)$	$g(\mu)$	$f(\mu)$	$\mathcal{A}_s$	$\mathcal{M}$
P	Poisson	$\frac{1}{x!}$	μ	e^{μ}	$\{0,1,\ldots\}$	$\mu > 0$
B	Binomial	$\binom{m}{x}$	$\frac{\mu}{m-\mu}$	$\left(\frac{m}{m-\mu}\right)^m$	$\{0,1,\ldots,m\}$	$0 < \mu < m$
G	Geométrica	1	$\frac{\mu}{1+\mu}$	$1+\mu$	$\{0,1,\ldots\}$	$\mu > 0$

Fonte: Adaptada de Carvalho (2017).

2.1.1 Algumas propriedades da distribuição PS

Iremos apresentar algumas propriedades da variável aleatória X com distribuição PS, isto é, $X \sim \text{PS}(\mu)$.

Primeiramente, a esperança matemática de X pode ser obtida a partir dos seguintes cálculos:

$$\begin{aligned}
f(\mu) &= \sum_{x\in\mathcal{A}_s} a(x)[g(\mu)]^x \\
f'(\mu) &= \sum_{x\in\mathcal{A}_s} a(x)\, x\, [g(\mu)]^{x-1} \\
\frac{f'(\mu)}{g'(\mu)} &= \sum_{x\in\mathcal{A}_s} x\, \frac{a(x)[g(\mu)]^x}{g(\mu)}\, \frac{f(\mu)}{f(\mu)} \\
\frac{f'(\mu)g(\mu)}{f(\mu)g'(\mu)} &= \sum_{x\in\mathcal{A}_s} x\, \frac{a(x)[g(\mu)]^x}{f(\mu)},
\end{aligned}$$

em que $f'(\mu)$ e $g'(\mu)$ são as derivadas das funções em relação a μ (para mais detalhes ver Gupta (1974)). Portando, a média da variável aleatória X é dada por:

$$\mathbb{E}(X) = \mu = \frac{f'(\mu)g(\mu)}{f(\mu)g'(\mu)}.$$

De maneira similar, obtemos

$$E(X^2) = \frac{\big(f(\mu)+\mu f'(\mu)\big)g(\mu)}{f(\mu)g'(\mu)},$$

e assim, a variância é dada por:

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - \big\{\mathbb{E}(X)\big\}^2 \\
&= \frac{\big(f(\mu)+\mu f'(\mu)\big)g(\mu)}{f(\mu)g'(\mu)} - \left(\frac{f'(\mu)g(\mu)}{f(\mu)g'(\mu)}\right)^2 \\
&= \left(\frac{f(\mu)+\mu f'(\mu)}{f'(\mu)}\right)\mu - \mu^2 \\
&= \frac{f(\mu)}{f'(\mu)}\mu,
\end{aligned}$$

ou seja, a variância da variável aleatória X é:

$$\mathbb{V}(X) \;=\; \sigma^2 \;=\; \frac{g(\mu)}{g'(\mu)}.$$

Generalizando o cálculo do valor esperado da variável aleatória X, podemos obter os momentos populacionais de ordem r, isto é, $\mu_r = \mathbb{E}(X^r)$. Baseando-se em Gupta (1974), a relação de recorrência entre os momentos populacionais pode ser obtida calculando:

$$\begin{aligned}
\mu_r &= \mathbb{E}(X^r) = \sum_{x\in\mathcal{A}_s} x^r \,\frac{a(x)[g(\mu)]^x}{f(\mu)} \\
\mu_r f(\mu) &= \sum_{x\in\mathcal{A}_s} x^r a(x)[g(\mu)]^x \\
\mu_r' f(\mu) + \mu_r f'(\mu) &= \sum_{x\in\mathcal{A}_s} x^{r+1} a(x)[g(\mu)]^x \,\frac{g'(\mu)}{g(\mu)} \\
\frac{g(\mu)}{g'(\mu)f(\mu)}\left(\mu_r' f(\mu) + \mu_r f'(\mu)\right) &= \sum_{x\in\mathcal{A}_s} x^{r+1} \,\frac{a(x)[g(\mu)]^x}{f(\mu)},
\end{aligned}$$

em que μ_r' é a derivada de μ_r em relação a μ.

Dessa forma, conseguimos obter a seguinte relação:

$$\begin{aligned}
\mu_{r+1} &= \mu_r' \frac{g(\mu)}{g'(\mu)} + \mu_r \frac{f'(\mu)g(\mu)}{f(\mu)g'(\mu)} \\
&= \mu_r' \frac{g(\mu)}{g'(\mu)} + \mu_r \mu_1 \\
&= \mu_r' \sigma^2 + \mu_r \mu,
\end{aligned}$$

com $r = 1, 2, \ldots$.

A função geradora de probabilidades de X, por definição, é dada por

$$\mathbb{G}(t) \;=\; \mathbb{E}(t^X) \;=\; \sum_{x\in\mathcal{A}_s} t^x \,\frac{a(x)[g(\mu)]^x}{f(\mu)} \;=\; \sum_{x\in\mathcal{A}_s} \frac{a(x)[t\; g(\mu)]^x}{f(\mu)}.$$

Com base em Gupta (1982), esta mesma função pode ser obtida supondo a existência de uma função $h(\mu)$, tal que $h(\mu)\cdot g(\mu) = \mu$, e utilizando a média da fórmula de Lagrange:

$$\begin{aligned}
\mu &= \sum_{v=1}^{\infty} \frac{1}{v!}\left(\frac{d^{\,v-1}}{d\mu}[h(\mu)]^v \,\Big|_{\mu=0}\right)\left(\frac{\mu}{h(\mu)}\right)^v \\
&= \sum_{v=1}^{\infty} \frac{1}{v!}\left(\frac{d^{\,v-1}}{d\mu}\left[\frac{\mu}{g(\mu)}\right]^v \,\Big|_{\mu=0}\right)\left(g(\mu)\right)^v \\
&= \psi(g(\mu)).
\end{aligned}$$

Ao considerarmos a existência de uma outra função $\dfrac{h(\mu)}{t}$, tal que $\dfrac{h(\mu)}{t}\cdot t\cdot g(\mu) = \mu$, temos:

$$\mu \;=\; \sum_{v=1}^{\infty} \frac{1}{v!}\left(\frac{d^{\,v-1}}{d\mu}\left[\frac{\mu}{g(\mu)}\right]^v \,\Big|_{\mu=0}\right)\left(t\; g(\mu)\right)^v \;=\; \psi(t\; g(\mu)),$$

e, portanto, a função geradora de probabilidades da variável aleatória X pode ser obtida por:

$$\frac{f(\psi(t\ g(\mu)))}{f(\psi(g(\mu)))} = \sum_{x\in\mathcal{A}_s} \frac{a(x)\,[t\ g(\mu)]^x}{f(\mu)} = \mathbb{G}(t).$$

Já a função geradora de momentos de X pode ser facilmente obtida a partir da definição:

$$\begin{aligned}\mathbb{M}(t) &= \mathbb{E}(e^{tX}) \\ &= \sum_{x\in\mathcal{A}_s} e^{tx}\,\frac{a(x)[g(\mu)]^x}{f(\mu)} \\ &= \sum_{x\in\mathcal{A}_s} \frac{a(x)\,[e^t\ g(\mu)]^x}{f(\mu)},\end{aligned}$$

ou ainda pode ser vista como um caso particular da função geradora de probabilidades quando $t = e^t$, isto é, $\mathbb{G}(e^t) = \mathbb{E}(e^{tX}) = \mathbb{M}(t)$. Assim, de maneira similar, iremos considerar uma função $\frac{h(\mu)}{e^t}$, tal que $e^t \cdot g(\mu) \cdot \frac{h(\mu)}{e^t} = \mu$. E, pela média da fórmula de Lagrange, temos:

$$\begin{aligned}\mu &= \sum_{v=1}^{\infty} \frac{1}{v!}\left(\frac{d^{\,v-1}}{d\mu}[h(\mu)]^v\ \Big|_{\mu=0}\right)\left(\frac{\mu}{h(\mu)}\right)^v \\ &= \sum_{v=1}^{\infty} \frac{1}{v!}\left(\frac{d^{\,v-1}}{d\mu}\left[\frac{\mu}{g(\mu)}\right]^v\ \Big|_{\mu=0}\right)\left(\frac{\mu e^t}{h(\mu)}\right)^v \\ &= \sum_{v=1}^{\infty} \frac{1}{v!}\left(\frac{d^{\,v-1}}{d\mu}\left[\frac{\mu}{g(\mu)}\right]^v\ \Big|_{\mu=0}\right)\left(e^t\ g(\mu)\right)^v \\ &= \psi^*(e^t\ g(\mu)),\end{aligned}$$

e, portanto, a função geradora de momentos também pode ser obtida por:

$$\frac{f(\psi^*(e^t\ g(\mu)))}{f(\psi^*(g(\mu)))} = \sum_{x\in\mathcal{A}_s} \frac{a(x)\,[e^t\ g(\mu)]^x}{f(\mu)} = \mathbb{M}(t).$$

Vale ressaltar que a função $\mathbb{M}(t)$ pode ser utilizada para obter os momentos populacionais de ordem r, $\mu_r = \mathbb{E}(X^r)$, fazendo a derivada de $r-$ésima ordem de $\mathbb{M}(t)$ em relação a t e, em seguida, atribuir valor zero para t ($t = 0$).

2.2 Distribuição Série de Potência k Modificada

Em muitos problemas práticos, podemos observar casos em que a frequência de uma observação qualquer, digamos k, em um conjunto de dados é maior, menor ou igual do que a esperada ao considerar uma distribuição discreta tradicional. Para essas situações, Carvalho (2017) propôs a família de distribuições discretas k modificada, capaz de modelar conjuntos de dados que apresentam ou não algum tipo de modificação (discrepância) na frequência desta observação k.

Sendo assim, seja Y uma variável aleatória discreta que possui distribuição Série de Potência k Modificada (k-MPS) [2], para algum $k \in \mathcal{A}_s$ e $k \geq s$, com parâmetros μ ($\mu \in \mathcal{M} \subseteq \mathbb{R}^+$) e θ ($\theta \in \Theta$). A FMP de Y é dada por:

$$\pi_{k-MPS}(y;\mu,\theta) \;=\; \theta \mathbb{I}_{\{k\}}(y) \;+\; (1-\theta)\pi_{PS}(y;\mu),\; \forall\, y \in \mathcal{A}_s, \tag{2.2}$$

em que $\mathcal{A}_s$ é o suporte formado pelo subconjunto dos inteiros $\{s, s+1, \ldots\}$; $\pi_{PS}(y;\mu) = \dfrac{a(y)[g(\mu)]^y}{f(\mu)}$ é a função massa de probabilidade da distribuição PS associada; θ é o parâmetro responsável pela modificação das probabilidades em relação à distribuição PS tradicional, satisfazendo a restrição:

$$-\frac{\pi_{PS}(k;\mu)}{1-\pi_{PS}(k;\mu)} \leq \theta \leq 1, \tag{2.3}$$

ou seja, $\Theta = \left[-\frac{\pi_{PS}(k;\mu)}{1-\pi_{PS}(k;\mu)}, 1\right]$; e $\mathbb{I}_{\{k\}}(y)$ é uma função indicadora, tal que:

$$\mathbb{I}_{\{k\}}(y) = \begin{cases} 1, & \text{se } y = k \\ 0, & \text{se } y \neq k \end{cases}.$$

Diferentes valores de θ levam a diferentes distribuições k-MPS. Podemos observar esse fato quando consideramos a proporção de observações adicionais ou em escasso de k, da seguinte forma:

$$\begin{aligned} \pi_{k-MPS}(k;\mu,\theta) - \pi_{PS}(k;\mu) &= \theta - \theta\pi_{PS}(k;\mu) \\ &= \theta\big(1-\pi_{PS}(k;\mu)\big). \end{aligned} \tag{2.4}$$

A partir da Equação (2.4), podemos claramente notar que o parâmetro θ controla essencialmente a probabilidade de ocorrência da observação k, produzindo os seguintes casos particulares (CARVALHO, 2017):

(i) Quando $\theta = -\frac{\pi_{PS}(k;\mu)}{1-\pi_{PS}(k;\mu)}$ na Equação (2.4), temos $\pi_{k-MPS}(k;\mu,\theta) = 0$. Sendo assim, a Equação (2.2) é a distribuição Série de Potência k Subtraída (k-SPS) [3], cuja função massa de probabilidade é dada por:

$$\pi_{k-SPS}(y;\mu) = \frac{\pi_{PS}(y;\mu)}{1-\pi_{PS}(k;\mu)}\left\{1-\mathbb{I}_{\{k\}}(y)\right\},\; \forall\, y \in \mathcal{A}_s,$$

a qual implica em,

$$\pi_{k-SPS}(k;\mu) = 0.$$

(ii) Para todo $-\frac{\pi_{PS}(k;\mu)}{1-\pi_{PS}(k;\mu)} < \theta < 0$ na Equação (2.4), temos $\theta(1-\pi_{PS}(k;\mu)) < 0$. Sendo assim, $\pi_{k-MPS}(k;\mu,\theta) < \pi_{PS}(k;\mu)$ e a Equação (2.2) é a distribuição Série de Potência k Deflacionada (k-DPS) [4], que possui uma proporção faltante de observação k .

[2] Do inglês "*k Modified Power Series*"
[3] Do inglês "*k Subtracted Power Series*"
[4] Do inglês "*k Deflated Power Series*"

(iii) Quando $\theta = 0$ na Equação (2.4), temos $\pi_{k-MPS}(k;\mu,\theta) - \pi_{PS}(k;\mu) = 0$. Sendo assim, $\pi_{k-MPS}(k;\mu,\theta) = \pi_{PS}(k;\mu)$ e a Equação (2.2) é a distribuição PS tradicional.

(iv) Para todo $0 < \theta < 1$ na Equação (2.4), temos $\theta\big(1-\pi_{PS}(k;\mu)\big) > 0$. Sendo assim, $\pi_{k-MPS}(k;\mu,\theta) > \pi_{PS}(k;\mu)$ e a Equação (2.2) é a distribuição Série de Potência k Inflacionada (k-IPS) [5], que possui uma proporção adicional de observação k .

(v) Quando $\theta = 1$ na Equação (2.4), temos $\pi_{k-MPS}(k;\mu,\theta) = 1$. Sendo assim, a Equação (2.2) é a distribuição degenerada com toda a massa em k.

Dessa forma, a distribuição k-MPS tem como casos particulares as distribuições k-SPS, k-DPS, PS e k-IPS. Assim, a ampla distribuição k-MPS pode ser ajustada a qualquer conjunto de dados de contagem sem o conhecimento prévio da existência ou não de discrepância na frequência da observação k. O diagrama apresentado na Figura 1 ilustra de forma simplificada os casos particulares apresentados para a distribuição k-MPS.

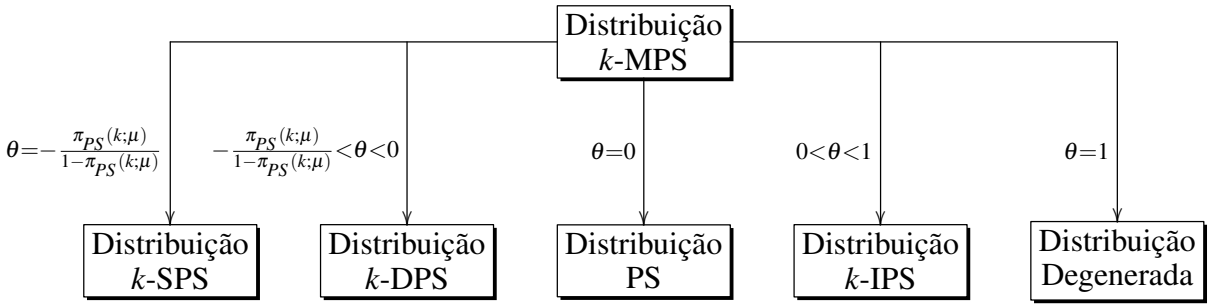


Figura 1 – Diagrama dos casos particulares da distribuição k-MPS.

Fonte: Adaptada de Carvalho (2017).

2.3 Um Caso Particular da k-MPS: Distribuição Série de Potência k Inflacionada

Considerando um caso particular da distribuição k-MPS, temos a distribuição inflacionada de uma observação qualquer. Murat e Szynal (1998) estudaram a família de distribuições discretas k inflacionadas, a qual modela adequadamente conjuntos de dados que apresentam uma alta frequência da observação k, isto é, característica de k-inflação.

Dessa forma, considere Y uma variável aleatória discreta que possui distribuição k-IPS, para algum $k \in \mathcal{A}_s$ e $k \geq s$, com parâmetros μ ($\mu \in \mathcal{M} \subseteq \mathbb{R}^+$) e θ ($\theta \in \Theta \subseteq (0,1)$). Logo, a FMP de Y é dada por:

$$\pi_{k-IPS}(y;\mu,\theta) \;=\; \theta\mathbb{I}_{\{k\}}(y) \;+\; (1-\theta)\pi_{PS}(y;\mu),\ \forall\, y \in \mathcal{A}_s, \tag{2.5}$$

[5] Do inglês "*k Inflated Power Series*"

em que $\mathcal{A}_s$ é o suporte formado pelo subconjunto dos inteiros $\{s, s+1, \ldots\}$; $\pi_{PS}(y;\mu)$ é a FMP da distribuição PS associada, dada em (2.1); θ é o parâmetro responsável pela modificação das probabilidades em relação à distribuição PS tradicional e, principalmente, pelo aumento da probabilidade da observação k (já que deve satisfazer $0 < \theta < 1$); e $\mathbb{I}_{\{k\}}(y)$ é uma função indicadora, tal que:

$$\mathbb{I}_{\{k\}}(y) = \begin{cases} 1, & \text{se } y = k \\ 0, & \text{se } y \neq k \end{cases}.$$

Podemos ainda reescrever a FMP da seguinte forma:

$$\pi_{k-IPS}(y;\mu,\theta) = \begin{cases} \theta + (1-\theta)\pi_{PS}(k;\mu), & \text{se } y = k \\ (1-\theta)\pi_{PS}(y;\mu), & \forall\, y \neq k, \text{ sendo que } y \in \mathcal{A}_s \\ 0, & \text{caso contrário} \end{cases}. \tag{2.6}$$

2.3.1 Algumas propriedades da distribuição k-IPS

Apresentaremos, a seguir, algumas propriedades da variável aleatória Y com distribuição k-IPS, isto é, $Y \sim k\text{-IPS}(\mu,\theta)$. Para isso, vamos considerar também uma variável aleatória X com distribuição PS associada a Y $\big(X \sim \text{PS}(\mu)\big)$.

A esperança matemática da variável aleatória Y, por definição, é dada por:

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{y\in\mathcal{A}_s} y\Big\{\theta\mathbb{I}_{\{k\}}(y) + (1-\theta)\pi_{PS}(y;\mu)\Big\} \\ &= \sum_{y\in\mathcal{A}_s} y(1-\theta)\pi_{PS}(y;\mu) + k\theta \\ &= (1-\theta)\mathbb{E}(X) + k\theta. \end{aligned}$$

Portanto, a média da variável aleatória Y é dada por:

$$\mathbb{E}(Y) = \mu_{k-IPS} = (1-\theta)\mu + k\theta.$$

Para calcular a variância da variável aleatória Y, primeiramente iremos obter $\mathbb{E}(Y^2)$:

$$\begin{aligned} \mathbb{E}(Y^2) &= \sum_{y\in\mathcal{A}_s} y^2\Big\{\theta\mathbb{I}_{\{k\}}(y) + (1-\theta)\pi_{PS}(y;\mu)\Big\} \\ &= \sum_{y\in\mathcal{A}_s} y^2(1-\theta)\pi_{PS}(y;\mu) + k^2\theta \\ &= (1-\theta)\mathbb{E}(X^2) + k^2\theta \\ &= (1-\theta)\big[\mathbb{V}(X) + \big\{\mathbb{E}(X)\big\}^2\big] + k^2\theta. \end{aligned}$$

Em seguida, obtemos:

$$\begin{aligned} \mathbb{V}(Y) &= \mathbb{E}(Y^2) - \big\{\mathbb{E}(Y)\big\}^2 \\ &= (1-\theta)\mathbb{V}(X) + (1-\theta)\big\{\mathbb{E}(X)\big\}^2 + k^2\theta - \big\{(1-\theta)\mathbb{E}(X) + k\theta\big\}^2, \end{aligned}$$

ou seja, a variância de Y, $\mathbb{V}(Y) = \sigma^2_{k-IPS}$, é dada por:

$$\begin{aligned}\mathbb{V}(Y) \;=\; \sigma^2_{k-IPS} &= (1-\theta)\left[\mathbb{V}(X) + \theta\big(k - \mathbb{E}(X)\big)^2\right] \\ &= (1-\theta)\left[\sigma^2 + \theta(k-\mu)^2\right].\end{aligned}$$

A função geradora de probabilidades da variável aleatória Y é calculada por:

$$\begin{aligned}\mathbb{G}_{k-IPS}(t) \;=\; \mathbb{E}(t^Y) &= \sum_{y\in\mathcal{A}_s} t^y\left\{\theta\mathbb{I}_{\{k\}}(y) \;+\; (1-\theta)\pi_{PS}(y;\mu)\right\} \\ &= (1-\theta)\sum_{y\in\mathcal{A}_s} t^y\pi_{PS}(y;\mu) + t^k\theta \\ &= (1-\theta)\mathbb{G}(t) + t^k\theta.\end{aligned}$$

Por fim, a função geradora de momentos de Y pode ser facilmente calculada por:

$$\begin{aligned}\mathbb{M}_{k-IPS}(t) \;=\; \mathbb{E}(e^{tY}) &= \sum_{y\in\mathcal{A}_s} e^{ty}\left\{\theta\mathbb{I}_{\{k\}}(y) \;+\; (1-\theta)\pi_{PS}(y;\mu)\right\} \\ &= (1-\theta)\sum_{y\in\mathcal{A}_s} e^{ty}\pi_{PS}(y;\mu) + e^{tk}\theta \\ &= (1-\theta)\mathbb{M}(t) + e^{tk}\theta.\end{aligned}$$

Além destas propriedades, destacamos a seguir uma importante proposição, que apresenta especificidade das distribuições k inflacionadas.

Proposição 2.1. *Considere as distribuições k-IPS e sua PS associada. Para todo $y \in \mathcal{A}_s$, as probabilidades $\pi_{k-IPS}(y;\mu,\theta)$ e $\pi_{PS}(y;\mu)$ satisfazem:*

$$\begin{cases} \pi_{k-IPS}(y;\mu,\theta) > \pi_{PS}(y;\mu), & \textit{somente para } y = k \\ \pi_{k-IPS}(y;\mu,\theta) < \pi_{PS}(y;\mu), & \forall\, y \neq k, \textit{ sendo que } y \in \mathcal{A}_s \end{cases}.$$

Demonstração. Para provar que ambas relações são satisfeitas, consideramos a seguinte relação:

$$\pi_{k-IPS}(y;\mu,\theta) - \pi_{PS}(y;\mu) \;=\; \theta\big(1 - \pi_{PS}(y;\mu)\big),\ \forall\, y \in \mathcal{A}_s.$$

Se $y = k$, então $\pi_{k-IPS}(k;\mu,\theta) - \pi_{PS}(k;\mu) = \theta\big(1-\pi_{PS}(k;\mu)\big)$. Como $0 < \theta < 1$, temos que $\theta\big(1-\pi_{PS}(k;\mu)\big) > 0$ e, consequentemente, $\pi_{k-IPS}(k;\mu,\theta) > \pi_{PS}(k;\mu)$.

Se $y \neq k$, então $\pi_{k-IPS}(y;\mu,\theta) - \pi_{PS}(y;\mu) = -\theta\pi_{PS}(y;\mu)$. Como $0 < \theta < 1$, temos que $-\theta\pi_{PS}(k;\mu) < 0$ e, consequentemente, $\pi_{k-IPS}(y;\mu,\theta) < \pi_{PS}(y;\mu)$. □

2.3.2 Versão hurdle da distribuição k-IPS

A FMP apresentada na Equação (2.5), correspondente a distribuição k-IPS, pode ser reescrita como:

$$\begin{aligned}
\pi_{k-IPS}(y;\mu,\theta) &= \left[\pi_{k-IPS}(y;\mu,\theta)\right]\mathbb{I}_{\{k\}}(y) + \left[\pi_{k-IPS}(y;\mu,\theta)\right](1-\mathbb{I}_{\{k\}}(y)) \\
&= \left[\theta+(1-\theta)\pi_{PS}(k;\mu)\right]\mathbb{I}_{\{k\}}(y) + \\
&\quad (1-\theta)(1-\pi_{PS}(k;\mu))\frac{\pi_{PS}(y;\mu)}{1-\pi_{PS}(k;\mu)}\,(1-\mathbb{I}_{\{k\}}(y)) \\
&= \left[\theta+(1-\theta)\pi_{PS}(k;\mu)\right]\mathbb{I}_{\{k\}}(y) + (1-\theta)\left[1-\pi_{PS}(k;\mu)\right]\pi_{k-SPS}(y;\mu) \\
&= \left[\theta(1-\pi_{PS}(k;\mu))+\pi_{PS}(k;\mu)\right]\mathbb{I}_{\{k\}}(y) + \\
&\quad \left[1-\left(\theta(1-\pi_{PS}(k;\mu))+\pi_{PS}(k;\mu)\right)\right]\pi_{k-SPS}(y;\mu),
\end{aligned}$$

e a versão *hurdle* (ver Dalrymple, Hudson e Ford (2003)) da distribuição k-IPS pode ser obtida ao considerar $\lambda = \theta(1-\pi_{PS}(k;\mu))+\pi_{PS}(k;\mu)$ e é dada por:

$$\pi_{k-IPS}(y;\mu,\lambda) = \lambda\,\mathbb{I}_{\{k\}}(y) + (1-\lambda)\,\pi_{k-SPS}(y;\mu),\ \forall\, y\in\mathcal{A}_s, \tag{2.7}$$

em que $\pi_{k-SPS}(y;\mu)$ corresponde a FMP da distribuição k-SPS, dada por:

$$\pi_{k-SPS}(y;\mu) = \frac{\pi_{PS}(y;\mu)}{1-\pi_{PS}(k;\mu)}\left\{1-\mathbb{I}_{\{k\}}(y)\right\},\ \forall\, y\in\mathcal{A}_s,$$

a qual implica em,

$$\pi_{k-SPS}(k;\mu) = 0.$$

Uma vez que $0<\theta<1$, podemos afirmar que o espaço paramétrico de λ, dito Ξ, é $0<\lambda<1-\pi_{PS}(k;\mu)$ ($\lambda\in\Xi\subseteq(0,1-\pi_{PS}(k;\mu))$).

A vantagem dessa parametrização é que os parâmetros λ e μ são ortogonais, permitindo estimar λ sem depender de μ.

Ressaltamos que é possível calcular todas as propriedades anteriormente citadas utilizando a versão *hurdle*, a partir da substituição de θ por $\frac{\lambda-\pi_{PS}(k;\mu)}{1-\pi_{PS}(k;\mu)}$ ou, equivalentemente, substituir o parâmetro $(1-\theta)$ por $\frac{1-\lambda}{1-\pi_{PS}(k;\mu)}$. Resumidamente, a Tabela 2 apresenta de maneira comparativa algumas propriedades da distribuição k-IPS considerando as FMPs parametrizadas em θ e em λ.

Tabela 2 – Comparação de algumas propriedades da distribuição k-IPS parametrizada em θ e em λ.

Característica	Função	
	k-IPS$(y;\mu,\theta)$	k-IPS$(y;\mu,\lambda)$
Função Geradora de Probabilidades	$\mathbb{G}_{k-IPS}(t) = (1-\theta)\mathbb{G}(t)+t^k\theta$	$\mathbb{G}_{k-IPS}(t) = \frac{(1-\lambda)\mathbb{G}(t)+t^k(\lambda-\pi_{PS}(k;\mu))}{1-\pi_{PS}(k;\mu)}$
Função Geradora de Momentos	$\mathbb{M}_{k-IPS}(t) = (1-\theta)\mathbb{M}(t)+e^{tk}\theta$	$\mathbb{M}_{k-IPS}(t) = \frac{(1-\lambda)\mathbb{M}(t)+e^{tk}(\lambda-\pi_{PS}(k;\mu))}{1-\pi_{PS}(k;\mu)}$
Média	$\mu_{k-IPS} = (1-\theta)\mu + k\theta$	$\mu_{k-IPS} = \frac{(1-\lambda)\mu+k(\lambda-\pi_{PS}(k;\mu))}{1-\pi_{PS}(k;\mu)}$
Variância	$\sigma^2_{k-IPS} = (1-\theta)\left[\sigma^2+\theta(k-\mu)^2\right]$	$\sigma^2_{k-IPS} = \frac{1-\lambda}{1-\pi_{PS}(k;\mu)}\left[\sigma^2+\frac{(\lambda-\pi_{PS}(k;\mu))(k-\mu)^2}{1-\pi_{PS}(k;\mu)}\right]$

Fonte: Elaborada pelo autor.

CAPÍTULO

3

DISTRIBUIÇÃO SÉRIE DE POTÊNCIA k_1 E k_2 INFLACIONADA

Conjuntos de dados de contagem podem também apresentar altas frequências em duas observações, digamos k_1 e k_2. Em muitos problemas reais, é muito comum a discrepância das frequências ocorrer nas observações zero e um: $k_1 = 0$ e $k_2 = 1$ (ver Alshkaki (2016), Saito e Rodrigues (2005), Melkersson e Olsson (1999)). Em um contexto mais geral, para estas situações, é recomendado a utilização de uma distribuição que explique adequadamente o comportamento dos dados que apresentam esta característica e, dessa forma, neste Capítulo propomos a família de distribuição Série de Potência k_1 e k_2 Inflacionada ($\boldsymbol{k}$-IPS) [1], em que $\boldsymbol{k} = (k_1, k_2)$, juntamente com sua versão *hurdle*.

Para isso, considere uma variável aleatória Z definida sobre os inteiros não negativos.

Teorema 3.1. *Seja a função* $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})$ *dada por:*

$$\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}) = \theta_1 \mathbb{I}_{\{k_1\}}(z) + \theta_2 \mathbb{I}_{\{k_2\}}(z) + \theta_0 \pi_{PS}(z;\mu),\ \forall\, z \in \mathcal{A}_s, \tag{3.1}$$

em que $\mathcal{A}_s$ *é o suporte formado pelo subconjunto dos inteiros* $\{s, s+1, \ldots\}$*;* $\mathbb{I}_{\{k_j\}}(z)$*, com* $j = 1,2$*, é uma função indicadora, tal que:*

$$\mathbb{I}_{\{k_j\}}(z) = \begin{cases} 1, & se\ z = k_j \\ 0, & se\ z \neq k_j \end{cases};$$

e $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$ *é o vetor de parâmetros responsáveis pela modificação das probabilidades (principalmente nas observações* k_1 *e* k_2*) em relação à distribuição PS tradicional associada, obedecendo às condições:*

[1] Do inglês "$\boldsymbol{k}$ *Inflated Power Series*"

C1. $\theta_i \in \Theta \subset (0,1),\ \forall\ i = 0,1,2$;

C2. $\theta_0 = 1 - \theta_1 - \theta_2$, *tal que* $\sum_{i:\ \forall\ i=0,1,2} \theta_i = 1.$

Então, a função $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})$ *é uma função de distribuição de probabilidade.*

Demonstração. Para provar que (3.1) é uma função de distribuição de probabilidade, precisamos mostrar que:

(i) $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}) > 0,\ \forall\ z \in \mathcal{A}_s$;

(ii) $\sum_{z:\ z\in\mathcal{A}_s} \pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}) = 1.$

Caso (i):

$(i.1)$ Se $z = k_1$, temos:

$$\pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) = \theta_1 + \theta_0\pi_{PS}(k_1;\mu).$$

Como sabemos que $0 < \theta_1 < 1$ e $0 < \theta_0 < 1$, então $\pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) > 0$.

$(i.2)$ Se $z = k_2$, temos que:

$$\pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) = \theta_2 + \theta_0\pi_{PS}(k_2;\mu).$$

Novamente, sabemos que $0 < \theta_2 < 1$ e $0 < \theta_0 < 1$, então $\pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) > 0$.

$(i.3)$ Para todo $z \in \mathcal{A}_s$, sendo $z \neq k_1$ e $z \neq k_2$, temos:

$$\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}) = \theta_0\pi_{PS}(z;\mu).$$

E, dado que $0 < \theta_0 < 1$, então $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}) > 0$.

Logo, $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}) > 0,\ \forall\ z \in \mathcal{A}_s$.

Caso (*ii*):

$$\begin{aligned}\sum_{z\in\mathcal{A}_s}\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}) &= \sum_{z\in\mathcal{A}_s}\Big[\theta_1\mathbb{I}_{\{k_1\}}(z)+\theta_2\mathbb{I}_{\{k_2\}}(z)+\theta_0\pi_{PS}(z;\mu)\Big]\\ &= \theta_1+\theta_2+\theta_0\sum_{z\in\mathcal{A}_s}\pi_{PS}(z;\mu)\\ &= \theta_1+\theta_2+\theta_0\\ &= 1.\end{aligned}$$

Verificado os itens (*i*) e (*ii*) que se fez necessário, temos, portanto, que $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})$ é uma função de distribuição de probabilidade.

□

Definição 3.1. *A função $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})$ dada pela Equação (3.1) é uma FMP e denominamos a distribuição por Série de Potência k_1 e k_2 Inflacionada ($\boldsymbol{k}$-IPS).*

Notação: $\boldsymbol{k}$-IPS$(\mu,\boldsymbol{\theta})$.

Sendo assim, podemos reescrever a FMP (3.1) da seguinte forma:

$$\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})=\begin{cases}\theta_1+(1-\theta_1-\theta_2)\pi_{PS}(k_1;\mu), & \text{se } z=k_1\\ \theta_2+(1-\theta_1-\theta_2)\pi_{PS}(k_2;\mu), & \text{se } z=k_2\\ (1-\theta_1-\theta_2)\pi_{PS}(z;\mu), & \forall\, z\neq k_1 \text{ e } z\neq k_2, \text{ com } z\in\mathcal{A}_s\\ 0, & \text{caso contrário}\end{cases}, \tag{3.2}$$

uma vez que $\theta_0=1-\theta_1-\theta_2$.

Podemos destacar algumas particularidades desta distribuição:

- Se $\theta_2\longrightarrow 0$ (ou $\theta_1\longrightarrow 0$), então temos a distribuição k_1-IPS(μ,θ_1) (ou k_2-IPS(μ,θ_2)), isto é, a modificação consiste em um único ponto do suporte (k-inflação, com $k=k_1$ ou $k=k_2$);
- Se $\theta_1\longrightarrow 0$ e $\theta_2\longrightarrow 0$, então temos a distribuição PS(μ) tradicional.

3.1 Algumas Propriedades da Distribuição k-IPS

A seguir, apresentamos algumas propriedades da variável aleatória Z com distribuição $\boldsymbol{k}$-IPS, isto é, $Z\sim\boldsymbol{k}$-IPS$(\mu,\boldsymbol{\theta})$. Para isso, vamos considerar as variáveis aleatórias X e Y_i, com $i=1,2$, tais que $X\sim$ PS(μ) e $Y_i\sim k_i$-IPS(μ,θ_i), distribuições associadas a distribuição $\boldsymbol{k}$-IPS.

Por definição, a esperança matemática da variável aleatória Z é dada por:

$$\begin{aligned}
\mathbb{E}(Z) \;=\; \mu_{\boldsymbol{k}-IPS} &= \sum_{z\in\mathcal{A}_s} z\Big\{\theta_1\mathbb{I}_{\{k_1\}}(z)+\theta_2\mathbb{I}_{\{k_2\}}(z)+\theta_0\pi_{PS}(z;\mu)\Big\}\\
&= \sum_{z\in\mathcal{A}_s} z(1-\theta_1-\theta_2)\pi_{PS}(z;\mu) \;+\; k_1\theta_1 \;+\; k_2\theta_2\\
&= (1-\theta_1-\theta_2)\mathbb{E}(X) \;+\; k_1\theta_1 \;+\; k_2\theta_2\\
&= \mu+\theta_1(k_1-\mu) \;+\; \theta_2(k_2-\mu)\\
&= (1-\theta_1)\mu \;+\; k_1\theta_1 \;+\; \theta_2(k_2-\mu)\\
&= (1-\theta_2)\mu \;+\; k_2\theta_2 \;+\; \theta_1(k_1-\mu),
\end{aligned}$$

podendo ser reescrita por:

$$\begin{aligned}
\mathbb{E}(Z) \;=\; \mu_{\boldsymbol{k}-IPS} &= (1-\theta_1-\theta_2)\mathbb{E}(X) \;+\; k_1\theta_1 \;+\; k_2\theta_2\\
&= \mathbb{E}_{k_1}(Y_1) \;+\; \theta_2(k_2-\mathbb{E}(X))\\
&= \mu_{k_1-IPS} \;+\; \theta_2(k_2-\mu),
\end{aligned}$$

ou ainda,

$$\begin{aligned}
\mathbb{E}(Z) \;=\; \mu_{\boldsymbol{k}-IPS} &= (1-\theta_1-\theta_2)\mathbb{E}(X) \;+\; k_1\theta_1 \;+\; k_2\theta_2\\
&= \mathbb{E}_{k_2}(Y_2) \;+\; \theta_1(k_1-\mathbb{E}(X))\\
&= \mu_{k_2-IPS} \;+\; \theta_1(k_1-\mu),
\end{aligned}$$

em que $\mathbb{E}_{k_1}(Y_1)=\mu_{k_1-IPS}$ e $\mathbb{E}_{k_2}(Y_2)=\mu_{k_2-IPS}$ correspondem as médias das variáveis aleatórias com distribuições k_1-IPS(μ,θ_1) e k_2-IPS(μ,θ_2), respectivamente.

Para calcular a variância da variável aleatória Z, encontramos primeiramente $\mathbb{E}(Z^2)$:

$$\begin{aligned}
\mathbb{E}(Z^2) &= \sum_{z\in\mathcal{A}_s} z^2\Big\{\theta_1\mathbb{I}_{\{k_1\}}(z)+\theta_2\mathbb{I}_{\{k_2\}}(z)+\theta_0\pi_{PS}(z;\mu)\Big\}\\
&= \sum_{z\in\mathcal{A}_s} z^2(1-\theta_1-\theta_2)\pi_{PS}(z;\mu) \;+\; k_1^2\theta_1 \;+\; k_2^2\theta_2\\
&= (1-\theta_1-\theta_2)\mathbb{E}(X^2) \;+\; k_1^2\theta_1 \;+\; k_2^2\theta_2\\
&= (1-\theta_1-\theta_2)\big[\mathbb{V}(X)+\big\{\mathbb{E}(X)\big\}^2\big] \;+\; k_1^2\theta_1 \;+\; k_2^2\theta_2.
\end{aligned}$$

Em seguida, calculamos:

$$\begin{aligned}
\mathbb{V}(Z) &= \mathbb{E}(Z^2) \;-\; \big\{\mathbb{E}(Z)\big\}^2\\
&= (1-\theta_1-\theta_2)\big[\mathbb{V}(X)+\mathbb{E}(X)^2\big] \;+\; k_1^2\theta_1 \;+\; k_2^2\theta_2 \;-\\
&\quad \big\{(1-\theta_1-\theta_2)\mathbb{E}(X)+k_1\theta_1+k_2\theta_2\big\}^2,
\end{aligned}$$

ou seja, a variância da variável aleatória Z é dada por:

$$\begin{aligned}\mathbb{V}(Z) \;=\; \sigma^2_{\boldsymbol{k}-IPS} \;&=\; (1-\theta_1-\theta_2)\big[\mathbb{V}(X)+\mathbb{E}(X)^2(\theta_1+\theta_2)\big] \;+\\ &\quad k_1\theta_1\big[k_1(1-\theta_1) \;-\; 2\mathbb{E}(X)(1-\theta_1-\theta_2)\big] \;+\\ &\quad k_2\theta_2\big[k_2(1-\theta_2) \;-\; 2\mathbb{E}(X)(1-\theta_1-\theta_2)\big] \;-\; 2k_1k_2\theta_1\theta_2\\ &=\; (1-\theta_1-\theta_2)\big[\sigma^2+\mu^2(\theta_1+\theta_2)\big] \;+\\ &\quad k_1\theta_1\big[k_1(1-\theta_1) \;-\; 2\mu(1-\theta_1-\theta_2)\big] \;+\\ &\quad k_2\theta_2\big[k_2(1-\theta_2) \;-\; 2\mu(1-\theta_1-\theta_2)\big] \;-\; 2k_1k_2\theta_1\theta_2.\end{aligned}$$

podendo ser reescrita por:

$$\mathbb{V}(Z) \;=\; \sigma^2_{\boldsymbol{k}-IPS} \;=\; \sigma^2_{\boldsymbol{k_1}-IPS}+\theta_2(1-\theta_2)(k_2-\mu)^2-2\theta_1\theta_2(k_1-\mu)(k_2-\mu)-\sigma^2\theta_2,$$

ou ainda,

$$\mathbb{V}(Z) \;=\; \sigma^2_{\boldsymbol{k}-IPS} \;=\; \sigma^2_{\boldsymbol{k_2}-IPS}+\theta_1(1-\theta_1)(k_1-\mu)^2-2\theta_1\theta_2(k_1-\mu)(k_2-\mu)-\sigma^2\theta_1,$$

em que $\mathbb{V}_{k_1}(Y_1)=\sigma^2_{k_1-IPS}$ e $\mathbb{V}_{k_2}(Y_2)=\sigma^2_{k_2-IPS}$ correspondem as variâncias das variáveis aleatórias com distribuições k_1-IPS(μ,θ_1) e k_2-IPS(μ,θ_2), respectivamente.

A função geradora de probabilidades da variável aleatória Z é calculada por:

$$\begin{aligned}\mathbb{G}_{\boldsymbol{k}-IPS}(t) \;=\; \mathbb{E}(t^Z) \;&=\; \sum_{z\in\mathcal{A}_s} t^z\Big\{\theta_1\mathbb{I}_{\{k_1\}}(z)+\theta_2\mathbb{I}_{\{k_2\}}(z)+\theta_0\pi_{PS}(z;\mu)\Big\}\\ &=\; (1-\theta_1-\theta_2)\sum_{z\in\mathcal{A}_s} t^z\pi_{PS}(z;\mu) \;+\; t^{k_1}\theta_1 \;+\; t^{k_2}\theta_2\\ &=\; (1-\theta_1-\theta_2)\mathbb{G}(t) \;+\; t^{k_1}\theta_1 \;+\; t^{k_2}\theta_2\\ &=\; (1-\theta_1)\mathbb{G}(t) \;+\; t^{k_1}\theta_1 \;+\; \theta_2(t^{k_2}-\mathbb{G}(t))\\ &=\; (1-\theta_2)\mathbb{G}(t) \;+\; t^{k_2}\theta_2 \;+\; \theta_1(t^{k_1}-\mathbb{G}(t))\\ &=\; \mathbb{G}(t)+\theta_1(t^{k_1}-\mathbb{G}(t)) \;+\; \theta_2(t^{k_2}-\mathbb{G}(t)),\end{aligned}$$

podendo ainda ser reescrita por:

$$\begin{aligned}\mathbb{G}_{\boldsymbol{k}-IPS}(t) \;&=\; (1-\theta_1-\theta_2)\mathbb{G}(t) \;+\; t^{k_1}\theta_1 \;+\; t^{k_2}\theta_2\\ &=\; \mathbb{G}_{k_1-IPS}(t) \;+\; \theta_2(t^{k_2}-\mathbb{G}(t))\\ &=\; \mathbb{G}_{k_1}(Y_1) \;+\; \theta_2(t^{k_2}-\mathbb{G}(t)),\end{aligned}$$

ou ainda,

$$\begin{aligned}\mathbb{G}_{\boldsymbol{k}-IPS}(t) \;&=\; (1-\theta_1-\theta_2)\mathbb{G}(t) \;+\; t^{k_1}\theta_1 \;+\; t^{k_2}\theta_2\\ &=\; \mathbb{G}_{k_2-IPS}(t) \;+\; \theta_1(t^{k_1}-\mathbb{G}(t))\\ &=\; \mathbb{G}_{k_2}(Y_2) \;+\; \theta_1(t^{k_1}-\mathbb{G}(t)),\end{aligned}$$

em que $\mathbb{G}_{k_1}(Y_1) = \mathbb{G}_{k_1-IPS}$ e $\mathbb{G}_{k_2}(Y_2) = \mathbb{G}_{k_2-IPS}$ correspondem, respectivamente, as funções geradoras de probabilidades das variáveis aleatórias com distribuições k_1-IPS(μ, θ_1) e k_2-IPS(μ, θ_2).

Por fim, a função geradora de momentos da variável aleatória Z pode ser facilmente calculada por:

$$\begin{aligned}
\mathbb{M}_{\boldsymbol{k}-IPS}(t) \;=\; \mathbb{E}(e^{tZ}) &= \sum_{z\in\mathcal{A}_s} e^{tz}\left\{\theta_1\mathbb{I}_{\{k_1\}}(z) + \theta_2\mathbb{I}_{\{k_2\}}(z) + \theta_0\pi_{PS}(z;\mu)\right\} \\
&= (1-\theta_1-\theta_2)\sum_{z\in\mathcal{A}_s} e^{tz}\pi_{PS}(z;\mu) \;+\; e^{tk_1}\theta_1 \;+\; e^{tk_2}\theta_2 \\
&= (1-\theta_1-\theta_2)\mathbb{M}(t) \;+\; e^{tk_1}\theta_1 \;+\; e^{tk_2}\theta_2 \\
&= (1-\theta_1)\mathbb{M}(t) \;+\; e^{tk_1}\theta_1 \;+\; \theta_2(e^{tk_2} - \mathbb{M}(t)) \\
&= (1-\theta_2)\mathbb{M}(t) \;+\; e^{tk_2}\theta_2 \;+\; \theta_1(e^{tk_1} - \mathbb{M}(t)) \\
&= \mathbb{M}(t) + \theta_1(e^{tk_1} - \mathbb{M}(t)) \;+\; \theta_2(e^{tk_2} - \mathbb{M}(t)),
\end{aligned}$$

que pode ser reescrita da seguinte forma:

$$\begin{aligned}
\mathbb{M}_{\boldsymbol{k}-IPS}(t) &= (1-\theta_1-\theta_2)\mathbb{M}(t) + e^{tk_1}\theta_1 + e^{tk_2}\theta_2 \\
&= \mathbb{M}_{k_1-IPS}(t) + \theta_2(e^{tk_2} - \mathbb{M}(t)) \\
&= \mathbb{M}_{k_1}(Y_1) + \theta_2(e^{tk_2} - \mathbb{M}(t)),
\end{aligned}$$

ou ainda,

$$\begin{aligned}
\mathbb{M}_{\boldsymbol{k}-IPS}(t) &= (1-\theta_1-\theta_2)\mathbb{M}(t) + e^{tk_1}\theta_1 + e^{tk_2}\theta_2 \\
&= \mathbb{M}_{k_2-IPS}(t) + \theta_1(e^{tk_1} - \mathbb{M}(t)) \\
&= \mathbb{M}_{k_2}(Y_2) + \theta_1(e^{tk_1} - \mathbb{M}(t)),
\end{aligned}$$

em que $\mathbb{M}_{k_1}(Y_1) = \mathbb{M}_{k_1-IPS}$ e $\mathbb{M}_{k_2}(Y_2) = \mathbb{M}_{k_2-IPS}$ correspondem, respectivamente, as funções geradoras de momentos das variáveis aleatórias com distribuições k_1-IPS(μ, θ_1) e k_2-IPS(μ, θ_2).

3.2 Comportamento da Função $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})$

Estudar o comportamento probabilístico de uma variável aleatória é de suma importância em problemas práticos. Este comportamento é completamente especificado pela FMP. Sendo assim, nesta Seção descrevemos este comportamento para algumas distribuições da família $\boldsymbol{k}$-IPS e as comparamos com as distribuições PS, k_1-IPS e k_2-IPS associadas. Em outras palavras, temos interesse em analisar e verificar possíveis tendências das funções, considerando diferentes valores dos parâmetros μ, θ_1 e θ_2, juntamente com diferentes valores dos pontos de modificações k_1 e k_2. Para essa descrição, optamos por uma visualização gráfica comparativa das probabilidades $\pi_{PS}(z;\mu)$, $\pi_{k_1-IPS}(z;\mu,\theta_1)$, $\pi_{k_2-IPS}(z;\mu,\theta_2)$ e $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})$, para alguns valores de z ($z \in \mathcal{A}_s$). De

antemão, destacamos alguns comentários relevantes que contribuirão para a melhor compreensão da análise do comportamento das funções.

Sobre este estudo, é importante ressaltar que consideramos a k_1-inflação (k_1-IPS(μ,θ_1)) e k_2-inflação (k_2-IPS(μ,θ_2)), proposto por Carvalho (2017), os quais são casos particulares da k-modificação, uma vez que os parâmetros θ_1 e θ_2 assumem valores entre 0 e 1 e que também compõem a distribuição $\boldsymbol{k}$-IPS$(\mu,\boldsymbol{\theta})$.

De uma forma resumida, ao considerar a distribuição k_1-IPS(μ,θ_1), temos a inflação da observação k_1 e, consequentemente, $\pi_{k_1-IPS}(k_1;\mu,\theta_1) > \pi_{PS}(k_1;\mu)$, implicando que as demais observações do suporte da variável (inclusive k_2) são deflacionadas ($\pi_{k_1-IPS}(z;\mu,\theta_1) < \pi_{PS}(z;\mu)$, $\forall\ z \in \{A_s - k_1\}$), garantindo que $\pi_{k_1-IPS}(k_1;\mu,\theta_1)$ é uma FMP com a condição $\sum_{z\in\mathcal{A}_s} \pi_{k_1-MPS}(z;\mu,\theta_1) = 1$ satisfeita. Equivalentemente, ao considerar a distribuição k_2-IPS(μ,θ_2), teremos a inflação de k_2 e, consequentemente, $\pi_{k_2-IPS}(k_2;\mu,\theta_2) > \pi_{PS}(k_2;\mu)$, implicando que as demais observações do suporte da variável (incluindo k_1) são deflacionadas ($\pi_{k_2-IPS}(z;\mu,\theta_2) < \pi_{PS}(z;\mu)$, $\forall\ z \in \{A_s - k_2\}$), garantindo que $\pi_{k_2-IPS}(k_2;\mu,\theta_2)$ é uma FMP com a condição $\sum_{z\in\mathcal{A}_s} \pi_{k_2-MPS}(z;\mu,\theta_2) = 1$ satisfeita.

Por outro lado, ao considerar os cálculos de probabilidade das observações k_1 e k_2 considerando a distribuição com a inflação simultânea nestes dois pontos (a distribuição $\boldsymbol{k}$-IPS$(\mu,\boldsymbol{\theta})$), o comportamento da FMP pode nos indicar as três possíveis situações:

i. $\pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) > \pi_{PS}(k_1;\mu)$ e $\pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) > \pi_{PS}(k_2;\mu)$;

ii. $\pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) > \pi_{PS}(k_1;\mu)$ e $\pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) < \pi_{PS}(k_2;\mu)$;

iii. $\pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) < \pi_{PS}(k_1;\mu)$ e $\pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) > \pi_{PS}(k_2;\mu)$.

Com isso, podemos afirmar que o conhecimento sobre a existência de inflação das observações k_1 e k_2 com probabilidades calculadas a partir da distribuição $\boldsymbol{k}$-IPS$(\mu,\boldsymbol{\theta})$ não é garantia de valores de probabilidades maiores para estas observações quando calculadas com a distribuição PS(μ) associada. Em suma, para as situações (*ii*) e (*iii*), podemos ainda afirmar que a $\boldsymbol{k}$-inflação é válida e de fato existe. Vejamos:

- Para a situação (*ii*), é razoável a suposição de inflação apenas da observação k_1 com cálculos de probabilidades a partir de $\pi_{k_1-IPS}(k_1;\mu,\theta_1)$. Sob esta suposição, temos deflação de todas as observações pertencentes ao suporte da variável que são diferentes de k_1, ou seja, $\pi_{k_1-IPS}(z;\mu,\theta_1) < \pi_{PS}(z;\mu)$, $\forall\ z \in \{A_s - k_1\}$, incluindo k_2. No entanto, apesar de haver uma deflação no ponto k_2, temos que $\pi_{k_2-IPS}(k_2;\mu,\theta_2) > \pi_{PS}(k_2;\mu)$, o que faz de k_2 uma observação também inflacionada;

- Equivalentemente, para a situação (*iii*), sob a suposição de inflação apenas da observação k_2 com cálculos de probabilidades a partir de $\pi_{k_2-IPS}(k_2;\mu,\theta_2)$, temos deflação de todas

as observações pertencentes ao suporte da variável que são diferentes de k_2, ou seja, $\pi_{k_2-IPS}(z;\mu,\theta_2) < \pi_{PS}(z;\mu)$, $\forall\ z \in \{A_s - k_2\}$, incluindo k_1. No entanto, apesar de haver uma deflação no ponto k_1, temos que $\pi_{k_1-IPS}(k_1;\mu,\theta_1) > \pi_{PS}(k_1;\mu)$, fazendo de k_1 uma observação também inflacionada.

Atentos a estes comentários, realizamos o estudo do comportamento da funções considerando inicialmente as observações zero ($k_1 = 0$) e um ($k_2 = 1$) como pontos de modificação ($\boldsymbol{k}$-inflação). Os valores considerados para os parâmetros da distribuição $\boldsymbol{k}$-IPS$(\mu,\boldsymbol{\theta})$ foram: $\mu = 0,50$ e 1,50; $\theta_1 = 0,49,\ 0,80$ e 0,40; e $\theta_2 = 0,49,\ 0,10$ e 0,50. Gráficos de barras com os comportamentos simultâneos das funções $\pi_{PS}(z;\mu)$, $\pi_{k_1-IPS}(z;\mu,\theta_1)$, $\pi_{k_2-IPS}(z;\mu,\theta_2)$ e $\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})$ foram construídos considerando algumas combinações de valores atribuídos aos parâmetros μ e $\boldsymbol{\theta}$ e a análise destes comportamentos foi feita de maneira conjunta para fins de comparação. Para este estudo foram consideradas as seguintes distribuições PS associadas: Poisson, Geométrica e Binomial.

As Figuras 2-4 apresentam os gráficos de barras com os comportamentos das FMPs das distribuições PS, k_1-IPS(μ,θ_1), k_2-IPS(μ,θ_2) e $\boldsymbol{k}$-IPS$(\mu,\boldsymbol{\theta})$ e, ao analisá-los, notamos comportamentos similares para as diferentes distribuições associadas. Descrevemos os comportamentos a seguir:

- A medida que μ aumenta e considerando $\theta_1 \approx \theta_2$, observamos que:
  - no ponto de modificação k_1 temos:

$$\pi_{k_1-IPS}(k_1;\mu,\theta_1) > \pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) > \pi_{PS}(k_1;\mu) > \pi_{k_2-IPS}(k_1;\mu,\theta_2);$$

  - no ponto de modificação k_2 temos:

$$\pi_{k_2-IPS}(k_2;\mu,\theta_2) > \pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) > \pi_{PS}(k_2;\mu) > \pi_{k_1-IPS}(k_2;\mu,\theta_1);$$

  - nas observações diferentes de k_1 e k_2 temos:

$$\pi_{PS}(z;\mu) > \pi_{k_1-IPS}(z;\mu,\theta_1) \approx \pi_{k_2-IPS}(z;\mu,\theta_2) > \pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta});$$

- A medida que μ aumenta e considerando $\theta_1 \gg \theta_2$, observamos que:
  - no ponto de modificação k_1 temos:

$$\pi_{k_1-IPS}(k_1;\mu,\theta_1) \approx \pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) \gg \pi_{PS}(k_1;\mu) \approx \pi_{k_2-IPS}(k_1;\mu,\theta_2);$$

  - no ponto de modificação k_2 temos:

$$\pi_{k_2-IPS}(k_2;\mu,\theta_2) > \pi_{PS}(k_2;\mu) > \pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) > \pi_{k_1-IPS}(k_2;\mu,\theta_1);$$

  - nas observações diferentes de k_1 e k_2 temos:

$$\pi_{PS}(z;\mu) \approx \pi_{k_2-IPS}(z;\mu,\theta_2) > \pi_{k_1-IPS}(z;\mu,\theta_1) \approx \pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta});$$

- A medida que θ_1 aumenta e considerando μ fixo, observamos que:
  - no ponto de modificação k_1 temos:

$$\pi_{k_1-IPS}(k_1;\mu,\theta_1) \approx \pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) > \pi_{PS}(k_1;\mu) \approx \pi_{k_2-IPS}(k_1;\mu,\theta_2);$$

  - no ponto de modificação k_2 temos:

$$\pi_{k_2-IPS}(k_2;\mu,\theta_2) > \pi_{PS}(k_2;\mu) > \pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) > \pi_{k_1-IPS}(k_2;\mu,\theta_1);$$

  - nas observações diferentes de k_1 e k_2 temos:

$$\pi_{PS}(z;\mu) \approx \pi_{k_2-IPS}(z;\mu,\theta_2) > \pi_{k_1-IPS}(z;\mu,\theta_1) \approx \pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta});$$

- A medida que θ_2 aumenta e considerando μ fixo, observamos que:
  - no ponto de modificação k_1 temos:

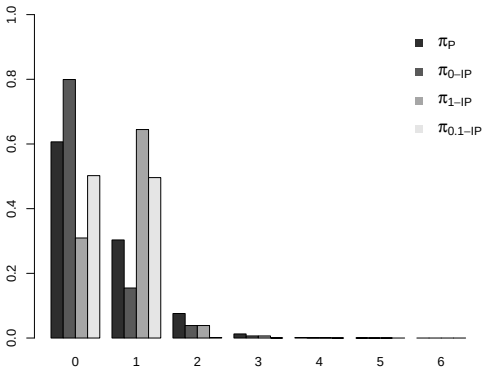
$$\pi_{k_1-IPS}(k_1;\mu,\theta_1) > \pi_{PS}(k_1;\mu) > \pi_{\boldsymbol{k}-IPS}(k_1;\mu,\boldsymbol{\theta}) > \pi_{k_2-IPS}(k_1;\mu,\theta_2);$$

  - no ponto de modificação k_2 temos:

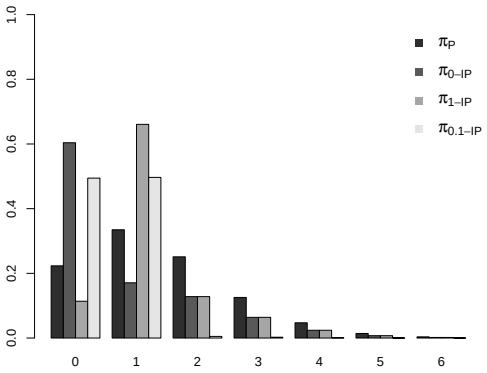
$$\pi_{k_2-IPS}(k_2;\mu,\theta_2) \approx \pi_{\boldsymbol{k}-IPS}(k_2;\mu,\boldsymbol{\theta}) > \pi_{PS}(k_2;\mu) \approx \pi_{k_1-IPS}(k_2;\mu,\theta_1);$$

  - nas observações diferentes de k_1 e k_2 temos:

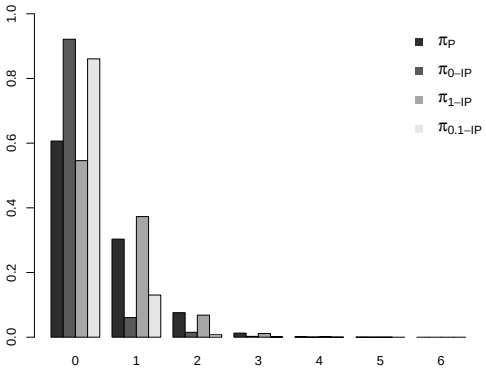
$$\pi_{PS}(z;\mu) \approx \pi_{k_1-IPS}(z;\mu,\theta_1) > \pi_{k_2-IPS}(z;\mu,\theta_2) \approx \pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}).$$



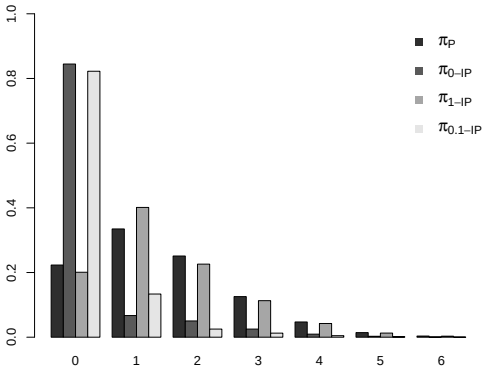
(a) $\mu = 0,50$, $\theta_1 = \theta_2 = 0,49$.



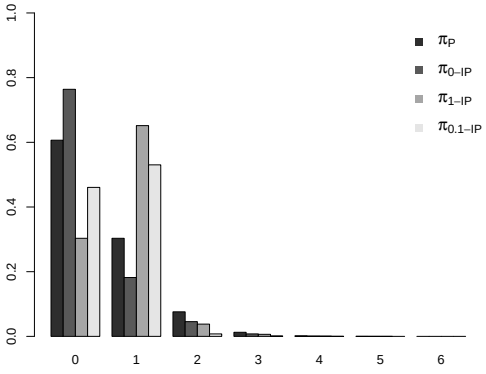
(b) $\mu = 1,50$, $\theta_1 = \theta_2 = 0,49$.



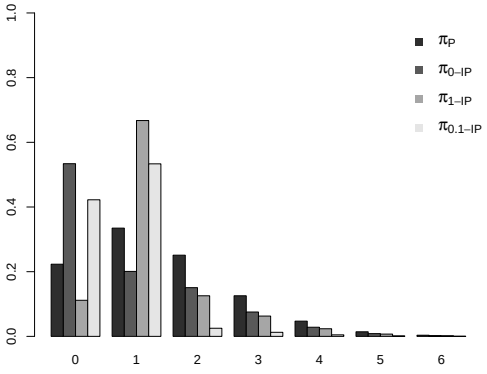
(c) $\mu = 0,50$, $\theta_1 = 0,80$ e $\theta_2 = 0,10$.



(d) $\mu = 1,50$, $\theta_1 = 0,80$ e $\theta_2 = 0,10$.



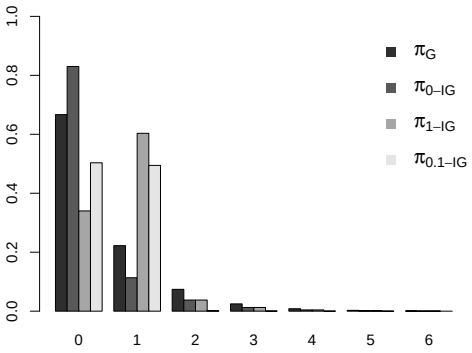
(e) $\mu = 0,50$, $\theta_1 = 0,40$ e $\theta_2 = 0,50$.



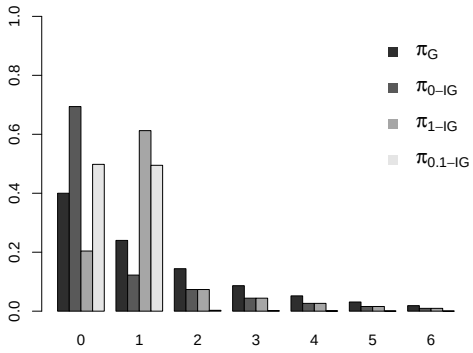
(f) $\mu = 1,50$, $\theta_1 = 0,40$ e $\theta_2 = 0,50$.

Figura 2 – Comportamento das FMPs das distribuições Poisson(μ), k_1-IP(μ, θ_1), k_2-IP(μ, θ_2) e $\boldsymbol{k}$-IP($\mu, \boldsymbol{\theta}$), com $k_1 = 0$ e $k_2 = 1$.

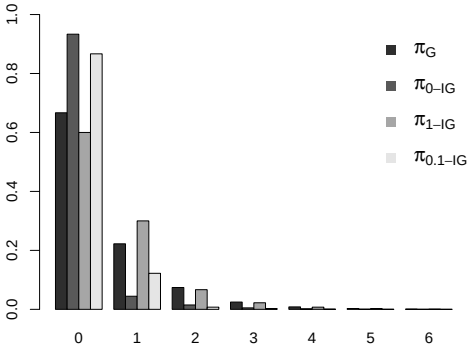
Fonte: Elaborada pelo autor.



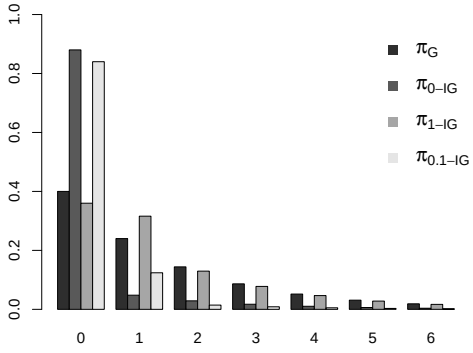
(a) $\mu = 0{,}50$, $\theta_1 = \theta_2 = 0{,}49$.



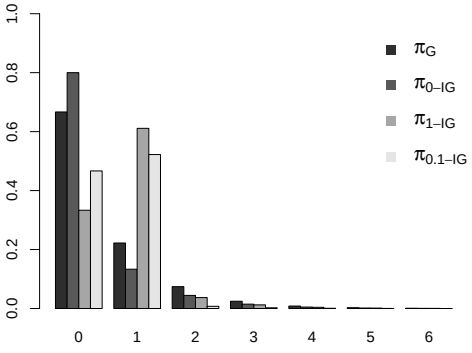
(b) $\mu = 1{,}50$, $\theta_1 = \theta_2 = 0{,}49$.



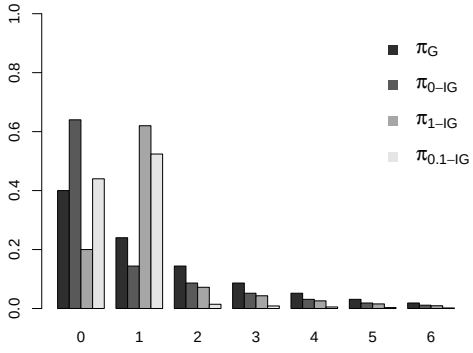
(c) $\mu = 0{,}50$, $\theta_1 = 0{,}80$ e $\theta_2 = 0{,}10$.



(d) $\mu = 1{,}50$, $\theta_1 = 0{,}80$ e $\theta_2 = 0{,}10$.



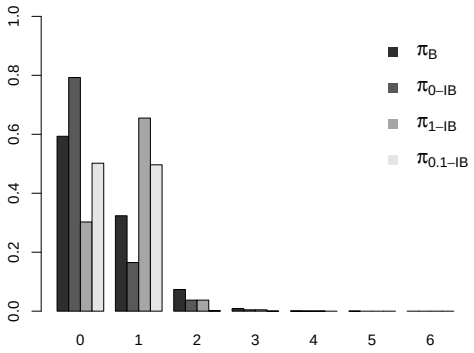
(e) $\mu = 0{,}50$, $\theta_1 = 0{,}40$ e $\theta_2 = 0{,}50$.



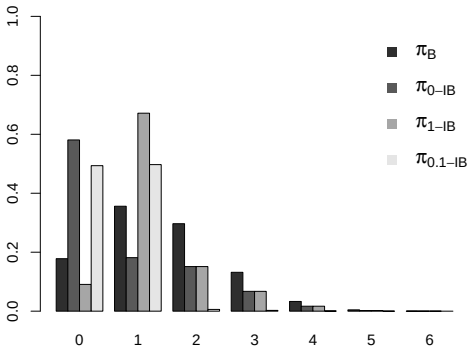
(f) $\mu = 1{,}50$, $\theta_1 = 0{,}40$ e $\theta_2 = 0{,}50$.

Figura 3 – Comportamento das FMPs das distribuições Geométrica(μ), k_1-IG(μ,θ_1), k_2-IG(μ,θ_2) e $\boldsymbol{k}$-IG($\mu,\boldsymbol{\theta}$), com $k_1 = 0$ e $k_2 = 1$.

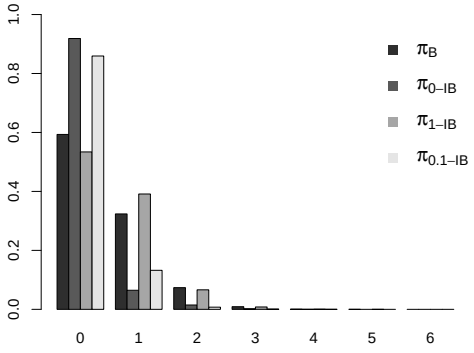
Fonte: Elaborada pelo autor.



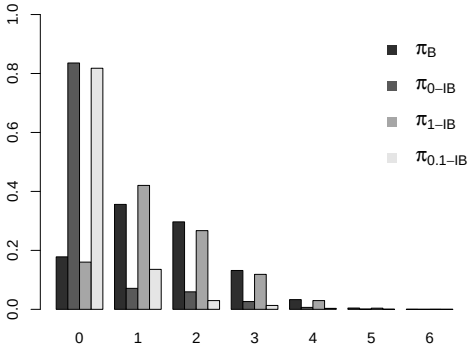
(a) $\mu = 0,50$, $\theta_1 = \theta_2 = 0,49$.



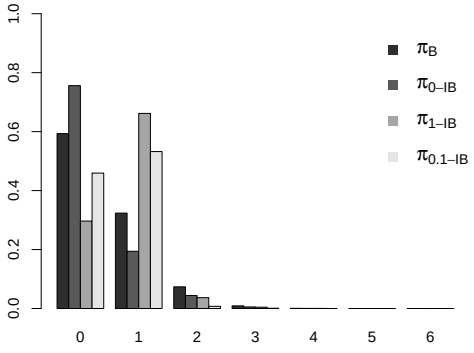
(b) $\mu = 1,50$, $\theta_1 = \theta_2 = 0,49$.



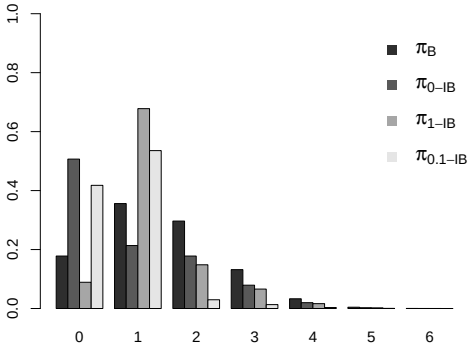
(c) $\mu = 0,50$, $\theta_1 = 0,80$ e $\theta_2 = 0,10$.



(d) $\mu = 1,50$, $\theta_1 = 0,80$ e $\theta_2 = 0,10$.



(e) $\mu = 0,50$, $\theta_1 = 0,40$ e $\theta_2 = 0,50$.



(f) $\mu = 1,50$, $\theta_1 = 0,40$ e $\theta_2 = 0,50$.

Figura 4 – Comportamento das FMPs das distribuições Binomial(μ), k_1-IB(μ, θ_1), k_2-IB(μ, θ_2) e $\boldsymbol{k}$-IB($\mu, \boldsymbol{\theta}$), com $m = 6$, $k_1 = 0$ e $k_2 = 1$.

Fonte: Elaborada pelo autor.

3.3 Versão *Hurdle* da Distribuição k-IPS

A FMP apresentada em (3.1) pode ser reescrita como:

$$\begin{aligned}
\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta}) &= \pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})\mathbb{I}_{\{k_1\}}(z)+\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})\mathbb{I}_{\{k_2\}}(z)+\\
&\quad \pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\theta})\big(1-\mathbb{I}_{\{k_1\}}(z)\big)\big(1-\mathbb{I}_{\{k_2\}}(z)\big)\\
&= \big(\pi_{k_1-IPS}(k_1;\mu,\theta_1)-\theta_2\pi_{PS}(k_1;\mu)\big)\mathbb{I}_{\{k_1\}}(z)+\\
&\quad \big(\pi_{k_2-IPS}(k_2;\mu,\theta_2)-\theta_1\pi_{PS}(k_2;\mu)\big)\mathbb{I}_{\{k_2\}}(z)+\\
&\quad \theta_0\left(\frac{1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)}{1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)}\right)\pi_{PS}(z;\mu)\big(1-\mathbb{I}_{\{k_1\}}(z)\big)\big(1-\mathbb{I}_{\{k_2\}}(z)\big)\\
&= \big(\theta_1+(1-\theta_1-\theta_2)\pi_{PS}(k_1;\mu)\big)\mathbb{I}_{\{k_1\}}(z)+\\
&\quad \big(\theta_2+(1-\theta_1-\theta_2)\pi_{PS}(k_2;\mu)\big)\mathbb{I}_{\{k_2\}}(z)+\\
&\quad (1-\theta_1-\theta_2)\big(1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)\big)\pi_{\boldsymbol{k}-SPS}(z;\mu), \qquad (3.3)
\end{aligned}$$

em que $\pi_{\boldsymbol{k}-SPS}(z;\mu)$ corresponde a FMP da distribuição Série de Potência k_1 e k_2 Subtraída ($\boldsymbol{k}$-SPS), sendo $\boldsymbol{k}=(k_1,k_2)$, a qual é dada por:

$$\pi_{\boldsymbol{k}-SPS}(z;\mu)=\frac{\pi_{PS}(z;\mu)}{1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)}\big(1-\mathbb{I}_{\{k_1\}}(z)\big)\big(1-\mathbb{I}_{\{k_2\}}(z)\big),\ \forall\, z\in\mathcal{A}_s,$$

a qual implica em,

$$\pi_{\boldsymbol{k}-SPS}(k_1;\mu)=0 \quad \text{e} \quad \pi_{\boldsymbol{k}-SPS}(k_2;\mu)=0.$$

Definindo o vetor de parâmetros $\boldsymbol{\gamma}=(\alpha,\beta)$, em que $\alpha=\theta_1+\theta_0\pi_{PS}(k_1;\mu)$ e $\beta=\theta_2+\theta_0\pi_{PS}(k_2;\mu)$, a FMP em (3.3) se reduz a

$$\pi_{\boldsymbol{k}-IPS}(z;\mu,\boldsymbol{\gamma}) = \alpha\mathbb{I}_{\{k_1\}}(z)+\beta\mathbb{I}_{\{k_2\}}(z)+(1-\alpha-\beta)\pi_{\boldsymbol{k}-SPS}(z;\mu),\ \forall\, z\in\mathcal{A}_s, \qquad (3.4)$$

que é a versão *hurdle* da distribuição $\boldsymbol{k}$-IPS.

Uma vez que $0<\theta_0<1$, equivalente a $0<(1-\theta_1-\theta_2)<1$, podemos dizer que $\theta_1<\alpha<\theta_1+\pi_{PS}(k_1;\mu)$ e $\theta_2<\beta<\theta_2+\pi_{PS}(k_2;\mu)$.

Essa nova parametrização tem como vantagem o fato de que os parâmetros μ e $\boldsymbol{\gamma}=(\alpha,\beta)$ são ortogonais, possibilitando estimar o vetor $\boldsymbol{\gamma}$ sem depender de μ e vice-versa.

Ressaltamos que é possível obter diretamente todas as propriedades anteriormente citadas ($\mu_{\boldsymbol{k}-IPS}$, $\sigma^2_{\boldsymbol{k}-IPS}$, $\mathbb{G}_{\boldsymbol{k}-IPS}(t)$ e $\mathbb{M}_{\boldsymbol{k}-IPS}(t)$) utilizando a versão *hurdle*. Para isto, basta substituirmos os parâmetros θ_1 e θ_2 pelas seguintes expressões:

$$\theta_1=\frac{\alpha\left[\big(1-\pi_{PS}(k_1;\mu)\big)\big(1-\pi_{PS}(k_2;\mu)\big)\right]-\pi_{PS}(k_1;\mu)\left[1-\pi_{PS}(k_1;\mu)-\beta\big(1-\pi_{PS}(k_1;\mu)\big)\right]}{\left[1-\pi_{PS}(k_1;\mu)\right]^2-\pi_{PS}(k_2;\mu)\left[1-\pi_{PS}(k_1;\mu)\right]}$$

e

$$\theta_2=\frac{\beta\big(1-\pi_{PS}(k_1;\mu)\big)-(1-\alpha)\pi_{PS}(k_2;\mu)}{1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)}.$$

CAPÍTULO

4

ESTIMAÇÃO DOS PARÂMETROS

Neste capítulo apresentamos o procedimento de estimação para os parâmetros das distribuições $\boldsymbol{k}$-IPS, considerando o método de máxima verossimilhança.

4.1 Método da Máxima Verossimilhança

Seja $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ uma amostra aleatória formada por n realizações independentes da variável aleatória Z com distribuição $\boldsymbol{k}$-IPS$(\mu, \boldsymbol{\theta})$ escrita da forma dada pela Equação (3.2). Considere também $\boldsymbol{z} = (z_1, \ldots, z_n)$ o vetor de observações associado a $\boldsymbol{Z}$. Equivalentemente, podemos reescrever a Equação (3.2) da seguinte maneira:

$$\begin{aligned}\pi_{\boldsymbol{k}-IPS}(z_i;\mu,\boldsymbol{\theta}) \quad = \quad & \left[\theta_1 + (1-\theta_1-\theta_2)\ \pi_{PS}(k_1;\mu)\right]^{\mathbb{I}_{\{k_1\}}(z_i)} \times \\ & \left[\theta_2 + (1-\theta_1-\theta_2)\ \pi_{PS}(k_2;\mu)\right]^{\mathbb{I}_{\{k_2\}}(z_i)} \times \\ & \left[(1-\theta_1-\theta_2)\ \pi_{PS}(z_i;\mu)\right]^{\left(1-\mathbb{I}_{\{k_1\}}(z_i)\right)\left(1-\mathbb{I}_{\{k_2\}}(z_i)\right)}.\end{aligned}$$

Dessa forma, a função de verossimilhança associada a $\boldsymbol{z}$ é dada por:

$$\begin{aligned}\mathcal{L}(\mu,\boldsymbol{\theta}|\boldsymbol{z}) \quad = \quad & \prod_{i=1}^{n}\Bigg\{\left[\theta_1 + (1-\theta_1-\theta_2)\ \pi_{PS}(k_1;\mu)\right]^{\mathbb{I}_{\{k_1\}}(z_i)} \times \\ & \left[\theta_2 + (1-\theta_1-\theta_2)\ \pi_{PS}(k_2;\mu)\right]^{\mathbb{I}_{\{k_2\}}(z_i)} \times \\ & \left[(1-\theta_1-\theta_2)\ \pi_{PS}(z_i;\mu)\right]^{\left(1-\mathbb{I}_{\{k_1\}}(z_i)\right)\left(1-\mathbb{I}_{\{k_2\}}(z_i)\right)}\Bigg\}.\end{aligned}$$

Uma vez que $\prod_{i=1}^{n}(1-\theta_1-\theta_2)^{(1-\mathbb{I}_{k_1}(z_i))(1-\mathbb{I}_{k_2}(z_i))} = (1-\theta_1-\theta_2)^{(n-n_{k_1}-n_{k_2})}$, temos:

$$\begin{aligned}\mathcal{L}(\mu,\boldsymbol{\theta}|\boldsymbol{z}) &= \left[\theta_1+(1-\theta_1-\theta_2)\ \pi_{PS}(k_1;\mu)\right]^{n_{k_1}} \times \\ &\quad \left[\theta_2+(1-\theta_1-\theta_2)\ \pi_{PS}(k_2;\mu)\right]^{n_{k_2}} \times \\ &\quad (1-\theta_1-\theta_2)^{(n-n_{k_1}-n_{k_2})} \cdot \prod_{\substack{\forall z_i\neq k_1\\ \forall z_i\neq k_2}} \pi_{PS}(z_i;\mu), \end{aligned} \tag{4.1}$$

em que n_{k_1} e n_{k_2} correspondem, respectivamente, a quantidade de observações k_1 e k_2 na amostra observada $\boldsymbol{z}$.

Logo, o logaritmo natural da função de verossimilhança, conhecido como função de log-verossimilhança, é dado por:

$$\begin{aligned}\ell(\mu,\boldsymbol{\theta}|\boldsymbol{z}) &= \log\left(\mathcal{L}(\mu,\boldsymbol{\theta}|\boldsymbol{z})\right) \\ &= n_{k_1}\ \log\left(\theta_1+(1-\theta_1-\theta_2)\ \pi_{PS}(k_1;\mu)\right)\ + \\ &\quad n_{k_2}\ \log\left(\theta_2+(1-\theta_1-\theta_2)\ \pi_{PS}(k_2;\mu)\right)\ + \\ &\quad (n-n_{k_1}-n_{k_2})\ \log(1-\theta_1-\theta_2)\ + \sum_{\substack{\forall z_i\neq k_1\\ \forall z_i\neq k_2}} \log(\pi_{PS}(z_i;\mu)). \end{aligned} \tag{4.2}$$

Derivando $\ell(\mu,\boldsymbol{\theta}|\boldsymbol{z})$ em relação a cada um dos parâmetros (θ_1, θ_2 e μ) temos o vetor escore $\mathbf{U}=(U_{\theta_1},\ U_{\theta_2},\ U_{\mu})$, cujos elementos são:

$$\begin{aligned}U_{\theta_1}=\frac{\partial\ell(\mu,\boldsymbol{\theta}|\boldsymbol{z})}{\partial\theta_1} &= \frac{n_{k_1}(1-\pi_{PS}(k_1;\mu))}{\theta_1+(1-\theta_1-\theta_2)\ \pi_{PS}(k_1;\mu)}\ - \\ &\quad \frac{n_{k_2}\pi_{PS}(k_2;\mu)}{\theta_2+(1-\theta_1-\theta_2)\ \pi_{PS}(k_2;\mu)}\ -\ \frac{(n-n_{k_1}-n_{k_2})}{1-\theta_1-\theta_2};\end{aligned}$$

$$\begin{aligned}U_{\theta_2}=\frac{\partial\ell(\mu,\boldsymbol{\theta}|\boldsymbol{z})}{\partial\theta_2} &= -\ \frac{n_{k_1}\pi_{PS}(k_1;\mu)}{\theta_1+(1-\theta_1-\theta_2)\ \pi_{PS}(k_1;\mu)}\ + \\ &\quad \frac{n_{k_2}(1-\pi_{PS}(k_2;\mu))}{\theta_2+(1-\theta_1-\theta_2)\ \pi_{PS}(k_2;\mu)}\ -\ \frac{(n-n_{k_1}-n_{k_2})}{1-\theta_1-\theta_2};\end{aligned}$$

e

$$\begin{aligned}U_{\mu}=\frac{\partial\ell(\mu,\boldsymbol{\theta}|\boldsymbol{z})}{\partial\mu} &= n_{k_1}\left[\frac{(1-\theta_1-\theta_2)\ \pi'_{PS}(k_1;\mu)}{\theta_1+(1-\theta_1-\theta_2)\ \pi_{PS}(k_1;\mu)}\right]\ + \\ &\quad n_{k_2}\left[\frac{(1-\theta_1-\theta_2)\ \pi'_{PS}(k_2;\mu)}{\theta_2+(1-\theta_1-\theta_2)\ \pi_{PS}(k_2;\mu)}\right]\ +\ \sum_{\substack{\forall z_i\neq k_1\\ \forall z_i\neq k_2}}\left[\frac{\pi'_{PS}(z_i;\mu)}{\pi_{PS}(z_i;\mu)}\right],\end{aligned}$$

em que $\pi'_{PS}(*;\mu)$ corresponde a derivada de $\pi_{PS}(*;\mu)$ no ponto "$*$" em relação a μ.

O estimador de máxima verossimilhança (EMV) de cada parâmetro (θ_1, θ_2 e μ, denotado por $\hat{\theta}_1$, $\hat{\theta}_2$ e $\hat{\mu}$, respectivamente) pode ser obtido igualando cada elemento do vetor escore $\mathbf{U}$ a

zero. Isto é, se $\hat{\theta}_1$, $\hat{\theta}_2$ e $\hat{\mu}$ existem, então devem satisfazer a equação de verossimilhança:

$$\begin{cases} U_{\theta_1} = 0 \\ U_{\theta_2} = 0 \\ U_{\mu} = 0 \end{cases}.$$

No entanto, a solução do sistema de equações não pode ser obtida explicitamente, sendo necessário recorrer a procedimentos numéricos para obtenção das estimativas de θ_1, θ_2 e μ.

Ao considerar a distribuição $\boldsymbol{k}$-IPS em sua versão *hurdle*, conseguimos obter equações mais simples para a estimação dos parâmetros. Sendo assim, utilizando a expressão dada na Equação (3.4), obtemos a função de verossimilhança associada ao vetor de observação $\boldsymbol{z}$, que é dada por:

$$\begin{aligned}
\mathcal{L}(\mu, \boldsymbol{\gamma}|\boldsymbol{z}) &= \prod_{i=1}^{n} \left[\alpha \mathbb{I}_{\{k_1\}}(z_i) + \beta \mathbb{I}_{\{k_2\}}(z_i) + (1-\alpha-\beta)\pi_{\boldsymbol{k}-SPS}(z_i;\mu) \right] \\
&= \prod_{i=1}^{n} \left[\alpha^{\mathbb{I}_{\{k_1\}}(z_i)} \cdot \beta^{\mathbb{I}_{\{k_2\}}(z_i)} \cdot \left[(1-\alpha-\beta)\pi_{\boldsymbol{k}-SPS}(z_i;\mu) \right]^{\left(1-\mathbb{I}_{\{k_1\}}(z_i)\right)\left(1-\mathbb{I}_{\{k_2\}}(z_i)\right)} \right] \\
&= \alpha^{n_{k_1}} \cdot \beta^{n_{k_2}} \cdot (1-\alpha-\beta)^{n-n_{k_1}-n_{k_2}} \prod_{\substack{\forall z_i \neq k_1 \\ \forall z_i \neq k_2}} \left(\frac{\pi_{PS}(z_i;\mu)}{1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)} \right).
\end{aligned}$$

E, portanto, a função de log-verossimilhança é:

$$\begin{aligned}
\ell(\mu, \boldsymbol{\gamma}|\boldsymbol{z}) &= \log\left(\mathcal{L}(\mu, \boldsymbol{\gamma}|\boldsymbol{z})\right) \\
&= \sum_{\substack{\forall z_i \neq k_1 \\ \forall z_i \neq k_2}} \log\left(\pi_{PS}(z_i;\mu)\right) - (n-n_{k_1}-n_{k_2}) \log\left(1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)\right) + \\
&\quad n_{k_1}\log(\alpha) + n_{k_2}\log(\beta) + (n-n_{k_1}-n_{k_2})\log(1-\alpha-\beta) \\
&= \ell_1(\mu|\boldsymbol{z}) + \ell_2(\boldsymbol{\gamma}|\boldsymbol{z}), \qquad (4.3)
\end{aligned}$$

em que

$$\ell_1(\mu|\boldsymbol{z}) = \sum_{\substack{\forall z_i \neq k_1 \\ \forall z_i \neq k_2}} \log\left(\pi_{PS}(z_i;\mu)\right) - (n-n_{k_1}-n_{k_2}) \log\left(1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)\right),$$

e

$$\ell_2(\boldsymbol{\gamma}|\boldsymbol{z}) = n_{k_1}\log(\alpha) + n_{k_2}\log(\beta) + (n-n_{k_1}-n_{k_2})\log(1-\alpha-\beta).$$

De acordo com a função de log-verossimilhança dada pela Equação (4.3), $\boldsymbol{\gamma} = (\alpha, \beta)$ e μ são ortogonais, uma vez que $\ell_1(\mu|\boldsymbol{z})$ independe de $\boldsymbol{\gamma}$, da mesma forma que $\ell_2(\boldsymbol{\gamma}|\boldsymbol{z})$ independe de μ, possibilitando estimar μ independentemente de $\boldsymbol{\gamma}$ e vice-versa.

Obtendo a primeira derivada de $\ell(\mu, \boldsymbol{\gamma}|\boldsymbol{z})$ em relação a cada um dos parâmetros (α, β e μ) temos o vetor escore $\mathbf{U}^\star = (U_\alpha^\star,\ U_\beta^\star,\ U_\mu^\star)$, cujos elementos são:

$$U_\alpha^\star \ = \ \frac{\partial \ell_2(\boldsymbol{\gamma}|\boldsymbol{z})}{\partial \alpha} \ = \ \frac{n_{k_1}}{\alpha} \ - \ \frac{(n - n_{k_1} - n_{k_2})}{1-\alpha-\beta};$$

$$U_\beta^\star \ = \ \frac{\partial \ell_2(\boldsymbol{\gamma}|\boldsymbol{z})}{\partial \beta} \ = \ \frac{n_{k_2}}{\beta} \ - \ \frac{(n - n_{k_1} - n_{k_2})}{1-\alpha-\beta};$$

e

$$U_\mu^\star \ = \ \frac{\partial \ell_1(\mu|\boldsymbol{z})}{\partial \mu} \ = \ \sum_{\substack{\forall z_i \neq k_1 \\ \forall z_i \neq k_2}} \frac{\frac{\partial}{\partial \mu}\pi_{PS}(z_i;\mu)}{\pi_{PS}(z_i;\mu)} \ + \ (n - n_{k_1} - n_{k_2}) \ \frac{\frac{\partial}{\partial \mu}\pi_{PS}(k_1;\mu) + \frac{\partial}{\partial \mu}\pi_{PS}(k_2;\mu)}{1 - \pi_{PS}(k_1;\mu) - \pi_{PS}(k_2;\mu)}.$$

Ao igualarmos cada elemento do vetor escore $\mathbf{U}^\star$ a zero encontramos um sistema de equações. A partir da solução deste, obtemos os EMVs para os parâmetros α e β, denotados por $\hat{\alpha}$ e $\hat{\beta}$, e que são dados respectivamente, por:

$$\hat{\alpha} = \frac{n_{k_1}}{n} \qquad \text{e} \qquad \hat{\beta} = \frac{n_{k_2}}{n}.$$

Já para o parâmetro μ, o EMV pode ser calculado resolvendo numericamente a solução da seguinte equação:

$$\mu \ = \ \frac{\big(1 - \pi_{PS}(k_1;\mu) - \pi_{PS}(k_2;\mu)\big)}{n - n_{k_1} - n_{k_2}} \sum_{\substack{\forall z_i \neq k_1 \\ \forall z_i \neq k_2}} z_i \ + \ k_1 \pi_{PS}(k_1;\mu) + k_2 \pi_{PS}(k_2;\mu).$$

O procedimento numérico que utilizamos foi o algoritmo de expectativa-maximização (EM), apresentado por Dempster, Laird e Rubin (1977). Este método tem como finalidade encontrar as estimativas de máxima verossimilhança dos parâmetros de uma distribuição qualquer a partir de um procedimento iterativo, dividido em dois passos: E e M. O passo E consiste na estimação via esperança condicional e o passo M maximiza a função de verossimilhança sob a hipótese de interesse.

Uma vez que $Z \sim \boldsymbol{k}\text{-IPS}(\mu, \boldsymbol{\theta})$, consideramos um vetor de observações $\boldsymbol{z_s} = (z_1, ..., z_{n_s})$ de tamanho n_s que contém apenas observações diferentes de k_1 e k_2. Em seguida, adicionamos uma variável aleatória latente, N_k, que nos indica a quantidade de observações iguais a k_1 e k_2 e que é utilizada para completar nosso vetor $\boldsymbol{z_s}$. Como nosso interesse está em verificar a quantidade de observações k_1 e/ou k_2 até a ocorrência das n_s observações, dizemos que N_k possui distribuição Binomial Negativa (BN). Assim, sua FMP é dada por:

$$\pi(N_k|\boldsymbol{z_s};\mu) = \frac{\Gamma(n_k + n_s)}{\Gamma(n_s)\Gamma(n_k+1)}(1 - \pi_{PS}(k_1;\mu) - \pi_{PS}(k_2;\mu))^{n_s}(\pi_{PS}(k_1;\mu) + \pi_{PS}(k_2;\mu))^{n_k},$$

com $n_k = 1, 2, \cdots$.

Dessa forma, o número esperado de observações iguais a k_1 e/ou k_2 é:

$$\mathbb{E}[N_k|\boldsymbol{z_s};\mu] = \frac{n_s(\pi_{PS}(k_1;\mu)+\pi_{PS}(k_2;\mu))}{1-\pi_{PS}(k_1;\mu)-\pi_{PS}(k_2;\mu)}.$$

Por fim, unindo o vetor $\boldsymbol{z_s}$ com os dados n_k, obtemos um vetor completo $\boldsymbol{z_c}$ com variáveis aleatórias Z com uma distribuição PS qualquer. E, repetindo os cálculos, temos a função de verossimilhança associada a este vetor $\boldsymbol{z_c}$, dada por:

$$\begin{aligned}
\mathcal{L}(\mu|\boldsymbol{z_c}) &= \prod_{i=1}^{n}\left\{\pi_{PS}(z_i;\mu)\right\} \\
&= \prod_{i=1}^{n}\left\{\pi_{PS}(z_1;\mu)\pi_{PS}(z_2;\mu)\pi_{PS}(z_i;\mu)\right\} \\
&= \left(\pi_{PS}(k_1;\mu)\ \pi_{PS}(k_2;\mu)\right)^{n_k}\prod_{i=1}^{\infty}\left\{\left(\pi_{PS}(i;\mu)\right)^{n_i}\right\}, \qquad (4.4)
\end{aligned}$$

e a função log-verossimilhança:

$$\ell(\mu|\boldsymbol{z_c}) = n_k\log\left(\pi_{PS}(k_1;\mu)\ \pi_{PS}(k_2;\mu)\right) + \sum_{i=1}^{\infty} n_i\log\left(\pi_{PS}(i;\mu)\right).$$

Seguindo os passos do algoritmo EM, determinamos o valor esperado da função verossimilhança dada em (4.4) no chamado "Passo E"e, posteriormente, a maximizamos no chamado "Passo M". Em suma, resolvemos o seguinte sistema:

$$\begin{aligned}
\frac{\partial\mathbb{E}\left[\ell(\mu|\boldsymbol{z_c})\right]}{\partial\mu} &= \frac{\partial}{\partial\mu}n_k\log\left(\pi_{PS}(k_1;\mu)\ \pi_{PS}(k_2;\mu)\right) + \frac{\partial}{\partial\mu}\sum_{i=1}^{\infty} n_i\log\left(\pi_{PS}(i;\mu)\right) \\
&= \frac{n_k\left[(k_1-\mu)+(k_2-\mu)\right]}{\sigma^2} + \frac{\sum_{i=1}^{\infty} n_i(z_i-\mu)}{\sigma^2} \\
&= 0.
\end{aligned}$$

Por fim, este procedimento iterativo nos leva a seguinte solução analítica para $\hat{\mu}$:

$$\hat{\mu} = \frac{\sum\limits_{\substack{\forall z_i\neq k_1\\ \forall z_i\neq k_2}} z_i + \hat{n}_{k_1}k_1 + \hat{n}_{k_2}k_2}{n_t + \hat{n}_{\boldsymbol{k}}},$$

em que $\hat{n}_{\boldsymbol{k}} = \mathbb{E}[N_{\boldsymbol{k}}|\boldsymbol{z_s};\mu]$, $\hat{n}_{\boldsymbol{k}} = \hat{n}_{k_1} + \hat{n}_{k_2}$, com $\hat{n}_{k_1} = \hat{n}_{\boldsymbol{k}}/2$ e $\hat{n}_{k_2} = \hat{n}_{\boldsymbol{k}} - \hat{n}_{k_1}$. Vale ressaltar que o critério utilizado para a ponderação $\hat{n}_{k_1} = \hat{n}_{k_2} = 1/2$ segue o pensamento de proporção mínima aceitável para o cenário de inflação.

4.2 Medidas de Evidências

Para ajustar a distribuição $\boldsymbol{k}$-IPS é necessário fazer uma suposição dos pontos de inflações, k_1 e k_2, existentes nos dados. Em algumas situações conseguimos identificar facilmente esses pontos a partir das frequências observadas. Entretanto, encontramos conjuntos de dados em que a suposição de inflação em k_1 e k_2 não é tão evidente, necessitando assim, de métodos específicos que determinam quais observações possuem essa discrepância, de uma forma mais precisa.

Dessa maneira, medidas de identificação são propostas para comparar as distribuições empírica e teórica (ajustada). Para isso, consideramos uma amostra aleatória $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)$ com n observações independentes da variável aleatória Z com distribuição $\boldsymbol{k}$-IPS$(\mu, \boldsymbol{\theta})$. Seja ainda $P_j(\boldsymbol{z})$ e $\hat{\pi}_j(\boldsymbol{z}) = \pi_{\boldsymbol{k}-IPS}(j; \mu, \hat{\boldsymbol{\theta}})$ (equivalentemente $\pi_{\boldsymbol{k}-IPS}(j; \mu, \hat{\boldsymbol{\gamma}})$, na versão *hurdle*) a proporção amostral e a probabilidade estimada da observação j, em que j representa qualquer valor do vetor $\boldsymbol{z}$.

Assim, nessa dissertação consideramos três medidas de identificação, as quais avaliam a diferença entre a proporção amostral e a probabilidade estimada das observações. Em outras palavras, quanto menor for essa diferença, mais adequada é a distribuição ajustada e assim, os pontos de inflação considerados são identificados corretamente. São elas:

- Distância Euclidiana (DE):

$$DE(P_j, \hat{\pi}_j) = \sqrt{\sum_{j:j\in J} (P_j(\boldsymbol{z}) - \hat{\pi}_j(\boldsymbol{z}))^2}.$$

- Divergência de Kullback-Leibler (KL) (KULLBACK; LEIBLER, 1951):

$$KL(P_j, \hat{\pi}_j) = \sum_{j:j\in J} P_j(\boldsymbol{z}) \log\left(\frac{P_j(\boldsymbol{z})}{\hat{\pi}_j(\boldsymbol{z})}\right).$$

- Divergência de Kullback-Leibler Simétrica (KLS)(KULLBACK; LEIBLER, 1951):

$$\begin{aligned} KLS &= \sum_{j:j\in J} P_j(\boldsymbol{z}) \log\left(\frac{P_j(\boldsymbol{z})}{\hat{\pi}_j(\boldsymbol{z})}\right) + \sum_{j:j\in J} \hat{\pi}_j(\boldsymbol{z}) \log\left(\frac{\hat{\pi}_j(\boldsymbol{z})}{P_j(\boldsymbol{z})}\right) \\ &= KL(P_j, \hat{\pi}_j) + KL(\hat{\pi}_j, P_j). \end{aligned}$$

Por fim, é recomendado utilizar essas medidas para identificar os pontos de inflação, a partir do ajuste das distribuições $\boldsymbol{k}$-IPS com diferentes valores k_1 ek_2, até identificar o melhor ajuste dado pelos pontos considerados. É importante mencionarmos que nem sempre todas as medidas indicam para os mesmos pontos de inflação. Nestes casos, cabe ao pesquisador escolher uma das medidas, com base na descrição do problema e/ou opinião de especialista.

CAPÍTULO

5

ESTUDO DE SIMULAÇÃO

Para o estudo de simulação, consideramos conjuntos de dados em que as observações zero e um ocorrem com uma alta frequência, isto é, $k_1 = 0$ e $k_2 = 1$ inflacionadas. Geramos $N = 1.000$ conjuntos de dados de tamanho $n = 50$, 200 e 500 para cada distribuição $0,1$-IPS, em que supomos dois diferentes valores para o parâmetro μ e três diferentes valores para cada parâmetro de modificação, θ_1 e θ_2.

Após estas considerações, as respectivas distribuições foram ajustadas considerando o método da máxima verossimilhança. Assim, obtivemos as estimativas dos parâmetros e os seus respectivos intervalos *bootstrap* não-paramétrico com 95% de confiança, considerando $R = 5.000$ réplicas. Por fim, averiguamos as probabilidades de cobertura de cada parâmetro estimado.

Para avaliamos a eficiência do estimador de cada parâmetro calculamos as estimativas de Monte Carlo do erro quadrático médio ($\mathbb{MSE}$, do inglês *mean square error*), a variância do estimador ($\mathbb{V}$), a média dos vícios ($\mathcal{B}$) e o erro percentual absoluto médio ($\mathbb{MAPE}$, do inglês *mean absolute percentage error*). Sendo assim, consideramos as seguintes equações:

$$\mathbb{MSE}(\hat{\mu}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}_i - \mu)^2, \quad \mathbb{MSE}(\hat{\theta}_1) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{1i} - \theta_1)^2 \quad \text{e} \quad \mathbb{MSE}(\hat{\theta}_2) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{2i} - \theta_2)^2;$$

$$\mathbb{V}(\hat{\mu}) = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\mu}_i - \overline{\hat{\mu}})^2, \quad \mathbb{V}(\hat{\theta}_1) = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\theta}_{1i} - \overline{\hat{\theta}_1})^2 \quad \text{e} \quad \mathbb{V}(\hat{\theta}_2) = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\theta}_{2i} - \overline{\hat{\theta}_2})^2;$$

$$\mathcal{B}(\hat{\mu}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}_i - \mu), \quad \mathcal{B}(\hat{\theta}_1) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{1i} - \theta_1) \quad \text{e} \quad \mathcal{B}(\hat{\theta}_2) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{2i} - \theta_2);$$

$$\mathbb{MAPE} = 100\left(\frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{\mu}_i - \mu|}{\mu}\right), \quad \mathbb{MAPE} = 100\left(\frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{\theta}_{1i} - \theta_1|}{\theta_1}\right) \quad \text{e} \quad \mathbb{MAPE} = 100\left(\frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{\theta}_{2i} - \theta_2|}{\theta_2}\right),$$

em que $\overline{\hat{\mu}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i$, $\overline{\hat{\theta}_1} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{1i}$ e $\overline{\hat{\theta}_2} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{2i}$.

Esperamos, neste estudo de simulação, que a teoria assintótica seja satisfeita, ou seja, a medida que o tamanho do conjunto de dados aumenta, os estimadores se tornam assintoticamente não viciados e, consequentemente, a razão entre o $\mathbb{MSE}$ e $\mathbb{V}$ se aproxime de um e $\mathcal{B}$ se aproxime de zero.

Para o estudo de simulação, consideramos as distribuições Poisson Zero e Um Inflacionados $(0,1$-IP), Geométrica Zero e Um Inflacionados $(0,1$-IG) e Binomial Zero e Um Inflacionados $(0,1$-IB), com m correspondente ao número de ensaios. Os valores dos parâmetros das distribuições $0,1$-IPS utilizados para geração dos dados são apresentados na Tabela 3.

Tabela 3 – Valores dos parâmetros das distribuições $\boldsymbol{k}$-IPS utilizados no estudo de simulação.

Distribuição	Parâmetros			m
	θ_1	θ_2	μ	
0,1-IP	0,80 0,40 0,49	0,10 0,50 0,49	3,0; 6,0	-
0,1-IG	0,80 0,40 0,49	0,10 0,50 0,49	3,0; 6,0	-
0,1-IB	0,80 0,40 0,49	0,10 0,50 0,49	3,0; 6,0	10

Devemos ressaltar que os tamanhos dos conjuntos de dados considerados e os valores atribuídos aos parâmetros de cada distribuição foram escolhidos de tal forma a produzir uma quantidade razoável de observações positivas na amostra gerada, possibilitando avaliar adequadamente o desempenho dos métodos utilizados para a estimação dos parâmetros.

A seguir, apresentaremos separadamente os resultados do estudo de simulação para as distribuições $0,1$-IP, $0,1$-IG e $0,1$-IB.

5.1 Resultados da simulação para a Distribuição 0,1-IP

Geramos N conjuntos de dados para a distribuição $0,1$-IP considerando os valores dos parâmetros apresentados na Tabela 3. Com base nestes conjuntos, conseguimos obter as estimativas dadas pelo método da máxima verossimilhança dos parâmetros μ, θ_1, θ_2, os intervalos *bootstrap* não-paramétrico com 95% de confiança, juntamente com as medidas de eficiência dos estimadores. As Tabelas 4-6 apresentam os resultados do estudo de simulação para esta distribuição.

A Tabela 4 apresenta as probabilidades de cobertura dos intervalos *bootstrap* com 95% de confiança. Podemos notar que o comportamento dos intervalos *bootstrap* não-paramétrico dos parâmetros θ_1 e θ_2 estão conforme o esperado, isto é, as probabilidades de cobertura estão

próximas do valor nominal que é de 95%. Já os intervalos *bootstrap* não-paramétrico de μ apresentaram probabilidades de cobertura menores do que o esperado, principalmente quando $n = 50$. Entretanto, suas probabilidades de cobertura vão se aproximando do valor teórico a medida que o tamanho da amostra aumenta. Dessa forma, acreditamos que em amostras maiores que 500 essa probabilidade atinja um valor satisfatório.

Tabela 4 – Probabilidades de cobertura dos intervalos de confiança *bootstrap* para os parâmetros da distribuição 0,1-IP.

n	Parâmetros			PC (95%) de μ	PC (95%) de θ_1	PC (95%) de θ_2
	θ_1	θ_2	μ			
50	0,80	0,10	3,0	75,9	94,1	89,8
			6,0	78,2	94,9	89,7
	0,40	0,50	3,0	73,4	92,4	92,7
			6,0	77,0	92,9	94,9
	0,49	0,49	3,0	49,8	89,7	89,5
			6,0	43,3	92,6	90,8
200	0,80	0,10	3,0	91,2	95,0	92,9
			6,0	92,5	96,5	95,3
	0,40	0,50	3,0	90,9	95,8	94,7
			6,0	93,5	95,1	94,8
	0,49	0,49	3,0	69,0	93,9	94,3
			6,0	69,0	94,7	94,9
500	0,80	0,10	3,0	91,4	95,5	94,9
			6,0	94,0	95,5	95,6
	0,40	0,50	3,0	92,5	95,1	95,2
			6,0	94,7	94,7	93,7
	0,49	0,49	3,0	87,5	95,0	95,5
			6,0	89,6	95,6	95,5

A Tabela 5 mostra os valores médios das estimativas pontuais de cada parâmetro obtidas pelo método da máxima verossimilhança e as médias dos limites inferiores e superiores dos intervalos *bootstrap* com 95% de confiança. Podemos notar que as estimativas médias pontuais estão bem próximas dos verdadeiros valores dos parâmetros.

A Tabela 6 apresenta as estimativas de Monte Carlo das medidas de eficiência dos estimadores de máxima verossimilhança. Observamos que, no geral, as médias da razão entre $\mathbb{MSE}$ e $\mathbb{V}$ de cada estimador são bem próximas de 1, o que indica que os estimadores de μ, θ_1 e θ_2 são consistentes assintoticamente, isto é, se aproximam cada vez mais do verdadeiro valor do parâmetro. Notamos também que os vícios dos estimadores θ_1 e θ_2 vão se tornando cada vez mais próximos de zero a medida que aumentamos o tamanho do conjunto de dados. Já os valores

do $\mathbb{MAPE}$ diminuem conforme o tamanho da amostra cresce.

Tabela 5 – Média das estimativas e intervalos de confiança *bootstrap* dos parâmetros da distribuição 0,1-IP.

n	Parâmetros			$\bar{\hat{\mu}}$	IC (95%) *bootstrap* de μ	$\bar{\hat{\theta}}_1$	IC (95%) *bootstrap* de θ_1	$\bar{\hat{\theta}}_2$	IC (95%) *bootstrap* de θ_2
	θ_1	θ_2	μ						
50	0,80	0,10	3,0	2,85	(1,58; 4,11)	0,78	(0,61; 0,87)	0,08	(0,01; 0,19)
			6,0	5,96	(4,16; 7,66)	0,80	(0,68; 0,89)	0,10	(0,02; 0,18)
	0,40	0,50	3,0	2,86	(1,63; 4,13)	0,38	(0,21; 0,51)	0,48	(0,30; 0,61)
			6,0	5,98	(4,19; 7,68)	0,40	(0,26; 0,53)	0,49	(0,35; 0,63)
	0,49	0,49	3,0	2,94	(2,10; 3,76)	0,46	(0,30; 0,58)	0,47	(0,29; 0,58)
			6,0	5,91	(4,78; 6,98)	0,48	(0,33; 0,60)	0,47	(0,33; 0,59)
200	0,80	0,10	3,0	2,86	(1,75; 3,88)	0,80	(0,73; 0,85)	0,09	(0,04; 0,14)
			6,0	5,96	(4,82; 7,09)	0,80	(0,74; 0,85)	0,10	(0,06; 0,14)
	0,40	0,50	3,0	2,88	(1,75; 3,93)	0,40	(0,32; 0,47)	0,49	(0,41; 0,57)
			6,0	6,01	(4,86; 7,15)	0,40	(0,33; 0,47)	0,50	(0,43; 0,57)
	0,49	0,49	3,0	2,89	(1,70; 4,09)	0,48	(0,41; 0,55)	0,49	(0,41; 0,55)
			6,0	5,85	(4,14; 7,43)	0,49	(0,42; 0,56)	0,49	(0,42; 0,56)
500	0,80	0,10	3,0	2,88	(2,16; 3,55)	0,80	(0,76; 0,83)	0,10	(0,07; 0,13)
			6,0	5,99	(5,27; 6,71)	0,80	(0,76; 0,83)	0,10	(0,07; 0,13)
	0,40	0,50	3,0	2,90	(2,18; 3,57)	0,40	(0,35; 0,44)	0,50	(0,45; 0,54)
			6,0	6,01	(5,30; 6,73)	0,40	(0,36; 0,44)	0,50	(0,46; 0,54)
	0,49	0,49	3,0	2,89	(1,50; 4,20)	0,49	(0,44; 0,53)	0,49	(0,44; 0,53)
			6,0	5,93	(4,31; 7,51)	0,49	(0,45; 0,53)	0,49	(0,45; 0,53)

Tabela 6 – Medidas de eficiência do estimador de cada parâmetro da distribuição 0,1-IP.

n	Parâmetros			$\hat{\mu}$			$\hat{\theta}_1$			$\hat{\theta}_2$		
	θ_1	θ_2	μ	$\frac{\mathbb{MSE}(\hat{\mu})}{\mathbb{V}(\hat{\mu})}$	$\mathcal{B}(\hat{\mu})$	$\mathbb{MAPE}$	$\frac{\mathbb{MSE}(\hat{\theta}_1)}{\mathbb{V}(\hat{\theta}_1)}$	$\mathcal{B}(\hat{\theta}_1)$	$\mathbb{MAPE}$	$\frac{\mathbb{MSE}(\hat{\theta}_2)}{\mathbb{V}(\hat{\theta}_2)}$	$\mathcal{B}(\hat{\theta}_2)$	$\mathbb{MAPE}$
50	0,80	0,10	3,0	1,02	-0,15	27,60	1,10	-0,02	5,97	1,09	-0,02	43,18
			6,0	1,00	-0,04	16,22	1,01	0,00	5,40	1,00	0,00	32,69
	0,40	0,50	3,0	1,02	-0,14	28,37	1,05	-0,02	15,15	1,06	-0,02	13,15
			6,0	1,00	-0,02	15,83	1,00	0,00	13,72	1,01	-0,01	10,88
	0,49	0,49	3,0	1,00	-0,06	36,55	1,15	-0,03	12,92	1,17	-0,03	12,80
			6,0	1,00	-0,09	22,62	1,03	-0,01	11,61	1,07	-0,02	11,93
200	0,80	0,10	3,0	1,07	-0,14	15,24	1,01	0,00	3,04	1,04	-0,01	20,73
			6,0	1,00	-0,04	7,95	1,00	0,00	2,67	1,00	0,00	16,79
	0,40	0,50	3,0	1,04	-0,12	15,63	1,00	0,00	6,98	1,02	-0,01	6,21
			6,0	1,00	0,01	7,72	1,00	0,00	6,49	1,00	0,00	5,60
	0,49	0,49	3,0	1,01	-0,11	31,02	1,03	-0,01	5,85	1,01	0,00	5,77
			6,0	1,01	-0,15	17,66	1,00	0,00	5,82	1,00	0,00	5,77
500	0,80	0,10	3,0	1,10	-0,12	10,03	1,02	0,00	1,81	1,01	0,00	12,61
			6,0	1,00	-0,01	5,01	1,00	0,00	1,76	1,00	0,00	10,21
	0,40	0,50	3,0	1,08	-0,01	9,62	1,01	0,00	4,55	1,01	0,00	3,88
			6,0	1,00	0,01	4,77	1,00	0,00	4,39	1,00	0,00	3,63
	0,49	0,49	3,0	1,03	-0,14	21,48	1,00	0,00	3,59	1,01	0,00	3,64
			6,0	1,01	-0,07	11,42	1,00	0,00	3,73	1,00	0,00	3,76

5.2 Resultados da simulação para a Distribuição 0,1-IG

Em seguida, geramos N conjuntos de dados para a distribuição $0,1$-IG considerando os valores dos parâmetros apresentados na Tabela 3. A partir desses conjuntos, conseguimos obter as estimativas dadas pelo método da máxima verossimilhança dos parâmetros μ, θ_1, θ_2, os intervalos *bootstrap* não-paramétrico com 95% de confiança, juntamente com as medidas de eficiência dos estimadores. As Tabelas 7-9 apresentam os resultados do estudo de simulação para esta distribuição.

A Tabela 7 apresenta as probabilidades de cobertura dos intervalos *bootstrap* com 95% de confiança. Notamos que o comportamento dos intervalos *bootstrap* não-paramétrico estão conforme o predito para os parâmetros θ_1 e θ_2, ou seja, suas probabilidades de cobertura estão próximas do valor nominal (95%). Assim como no estudo de simulação anterior, os intervalos *bootstrap* não-paramétrico de μ não apresentaram probabilidades de cobertura como o esperado, principalmente quando $n = 50$, mas essa probabilidade aumenta à medida que o tamanho da amostra cresce. Mais uma vez, acreditamos que em amostras maiores que 500 essa probabilidade atinja um valor satisfatório.

Tabela 7 – Probabilidades de cobertura dos intervalos de confiança *bootstrap* para os parâmetros da distribuição 0,1-IG.

n	Parâmetros			PC (95%) de μ	PC (95%) de θ_1	PC (95%) de θ_2
	θ_1	θ_2	μ			
50	0,80	0,10	3,0	57,2	92,1	89,8
			6,0	64,4	95,5	90,5
	0,40	0,50	3,0	62,2	91,6	94,4
			6,0	63,9	93,2	93,8
	0,49	0,49	3,0	52,6	91,2	90,8
			6,0	41,4	94,1	91,9
200	0,80	0,10	3,0	90,4	95,5	94,9
			6,0	89,4	94,0	94,5
	0,40	0,50	3,0	87,9	94,6	96,6
			6,0	93,7	94,6	95,0
	0,49	0,49	3,0	68,5	94,5	94,6
			6,0	69,4	94,9	95,0
500	0,80	0,10	3,0	93,3	96,0	92,8
			6,0	95,3	94,2	94,7
	0,40	0,50	3,0	91,7	95,6	95,2
			6,0	93,3	95,3	95,1
	0,49	0,49	3,0	86,4	95,5	96,9
			6,0	89,1	95,0	95,2

Já a Tabela 8 mostra os valores médios das estimativas pontuais dos parâmetros obtidas pelo método da máxima verossimilhança e as médias dos limites inferiores e superiores dos intervalos *bootstrap* com 95% de confiança. Observamos que as estimativas médias pontuais estão bem próximas dos verdadeiros valores dos parâmetros.

A Tabela 9 apresenta as medidas de eficiência dos estimadores de máxima verossimilhança via Monte Carlo. Notamos que as médias da razão entre $\mathbb{MSE}$ e $\mathbb{V}$, em sua maioria, são bem próximas de 1, indicando que os estimadores de μ, θ_1 e θ_2 são consistentes assintoticamente. Já os vícios vão se aproximam de zero à medida que aumentamos o tamanho do conjunto de dados (exceto para alguns valores de μ). Por outro lado, os valores do $\mathbb{MAPE}$ diminuem conforme o tamanho da amostra cresce.

Tabela 8 – Média das estimativas e intervalos de confiança *bootstrap* dos parâmetros da distribuição 0,1-IG.

n	Parâmetros			$\bar{\hat{\mu}}$	IC (95%) *bootstrap* de μ	$\bar{\hat{\theta}}_1$	IC (95%) *bootstrap* de θ_1	$\bar{\hat{\theta}}_2$	IC (95%) *bootstrap* de θ_2
	θ_1	θ_2	μ						
50	0,80	0,10	3,0	3,10	(1,80; 4,85)	0,77	(0,58; 0,86)	0,09	(0,01; 0,19)
			6,0	6,06	(2,98; 9,78)	0,78	(0,63; 0,87)	0,09	(0,00; 0,18)
	0,40	0,50	3,0	3,20	(1,76; 5,11)	0,37	(0,18; 0,51)	0,48	(0,32; 0,61)
			6,0	6,05	(2,93; 9,92)	0,39	(0,22; 0,52)	0,49	(0,33; 0,62)
	0,49	0,49	3,0	2,97	(2,14; 3,77)	0,46	(0,30; 0,59)	0,46	(0,29; 0,58)
			6,0	6,02	(4,91; 7,05)	0,48	(0,33; 0,60)	0,47	(0,33; 0,59)
200	0,80	0,10	3,0	3,11	(1,58; 5,17)	0,80	(0,72; 0,86)	0,10	(0,05; 0,15)
			6,0	6,08	(3,19; 9,61)	0,80	(0,73; 0,85)	0,10	(0,05; 0,15)
	0,40	0,50	3,0	3,03	(1,54; 5,03)	0,40	(0,31; 0,47)	0,50	(0,42; 0,57)
			6,0	6,01	(4,86; 7,14)	0,40	(0,33; 0,47)	0,50	(0,43; 0,57)
	0,49	0,49	3,0	2,89	(1,73; 4,05)	0,49	(0,41; 0,55)	0,48	(0,41; 0,55)
			6,0	5,94	(4,23; 7,54)	0,49	(0,42; 0,56)	0,49	(0,42; 0,56)
500	0,80	0,10	3,0	2,90	(2,19; 3,57)	0,80	(0,76; 0,83)	0,10	(0,07; 0,13)
			6,0	6,01	(5,29; 6,72)	0,80	(0,77; 0,84)	0,10	(0,07; 0,13)
	0,40	0,50	3,0	2,91	(2,18; 3,58)	0,40	(0,36; 0,44)	0,50	(0,45; 0,54)
			6,0	5,98	(5,27; 6,69)	0,40	(0,36; 0,44)	0,50	(0,46; 0,54)
	0,49	0,49	3,0	2,84	(1,50; 4,17)	0,49	(0,44; 0,53)	0,49	(0,44; 0,53)
			6,0	6,01	(4,37; 7,59)	0,49	(0,45; 0,53)	0,49	(0,45; 0,53)

Tabela 9 – Medidas de eficiência do estimador de cada parâmetro da distribuição 0,1-IG.

n	Parâmetros			$\hat{\mu}$			$\hat{\theta}_1$			$\hat{\theta}_2$		
	θ_1	θ_2	μ	$\frac{\mathrm{MSE}(\hat{\mu})}{\mathbb{V}(\hat{\mu})}$	$\mathcal{B}(\hat{\mu})$	MAPE	$\frac{\mathrm{MSE}(\hat{\theta}_1)}{\mathbb{V}(\hat{\theta}_1)}$	$\mathcal{B}(\hat{\theta}_1)$	MAPE	$\frac{\mathrm{MSE}(\hat{\theta}_2)}{\mathbb{V}(\hat{\theta}_2)}$	$\mathcal{B}(\hat{\theta}_2)$	MAPE
50	0,80	0,10	3,0	1,00	0,10	44,17	1,21	-0,03	7,16	1,07	-0,01	41,48
			6,0	1,00	0,06	43,24	1,07	-0,02	6,17	1,03	-0,01	38,22
	0,40	0,50	3,0	1,01	0,20	45,28	1,12	-0,03	16,88	1,07	-0,02	11,85
			6,0	1,00	0,05	44,96	1,03	-0,01	15,00	1,03	-0,01	11,86
	0,49	0,49	3,0	1,00	-0,50	37,28	1,13	-0,48	12,27	1,18	-0,48	12,69
			6,0	1,00	0,02	22,55	1,03	-0,01	11,54	1,07	-0,02	11,67
200	0,80	0,10	3,0	1,01	0,11	27,21	1,01	0,00	3,29	1,01	0,00	19,27
			6,0	1,00	0,08	24,32	1,01	0,00	3,07	1,00	0,00	18,53
	0,40	0,50	3,0	1,00	0,03	27,73	1,00	0,00	7,83	1,01	0,00	6,09
			6,0	1,00	0,01	7,77	1,00	0,00	7,00	1,00	0,00	5,84
	0,49	0,49	3,0	1,01	-0,11	30,98	1,01	0,00	5,82	1,02	-0,01	5,86
			6,0	1,00	-0,06	18,24	1,00	0,00	5,31	1,00	0,00	5,34
500	0,80	0,10	3,0	1,07	-0,10	9,85	1,01	0,00	1,85	1,02	0,00	12,83
			6,0	1,00	0,01	4,90	1,00	0,00	1,78	1,00	0,00	10,87
	0,40	0,50	3,0	1,07	-0,09	9,95	1,00	0,00	4,41	1,02	0,00	3,78
			6,0	1,00	-0,02	4,93	1,00	0,00	4,32	1,00	0,00	3,56
	0,49	0,49	3,0	1,04	-0,16	21,99	1,00	0,00	3,51	1,01	0,00	3,52
			6,0	1,00	0,01	11,44	1,00	0,00	3,73	1,00	0,00	3,71

5.3 Resultados da simulação para a Distribuição 0,1-IB

Por fim, geramos N conjuntos de dados para a distribuição $0,1$-IB considerando os valores dos parâmetros apresentados na Tabela 3 e com $m = 10$. Com base nesses conjuntos, obtemos as estimativas dadas pelo método da máxima verossimilhança dos parâmetros μ, θ_1, θ_2, os intervalos *bootstrap* não-paramétrico com 95% de confiança, juntamente com as medidas de eficiência dos estimadores. As Tabelas 10-12 apresentam os resultados do estudo de simulação para esta distribuição.

A Tabela 10 mostra as probabilidades de cobertura dos intervalos *bootstrap* com 95% de confiança. Podemos notar que o comportamento dos intervalos *bootstrap* não-paramétrico dos parâmetros θ_1 e θ_2 estão como o esperado, ou seja, as probabilidades de cobertura estão próximas do valor nominal (95%). Por outro lado, quando olhamos μ, notamos que as probabilidades aumentam conforme o tamanho da amostra cresce, porém não tanto quanto o esperado. Acreditamos assim, que em amostras maiores que 500 essa probabilidade atinja um valor satisfatório.

Tabela 10 – Probabilidades de cobertura dos intervalos de confiança *bootstrap* para os parâmetros da distribuição 0,1-IB.

n	Parâmetros			PC (95%) de μ	PC (95%) de θ_1	PC (95%) de θ_2
	θ_1	θ_2	μ			
50	0,80	0,10	3,0	79,4	94,3	88,5
			6,0	73,9	94,6	88,3
	0,40	0,50	3,0	75,9	92,0	93,2
			6,0	73,4	94,2	95,5
	0,49	0,49	3,0	58,2	90,9	91,2
			6,0	39,5	93,2	93,4
200	0,80	0,10	3,0	91,7	95,3	92,9
			6,0	92,1	94,5	93,2
	0,40	0,50	3,0	92,2	95,1	94,7
			6,0	94,2	94,7	94,9
	0,49	0,49	3,0	75,4	92,9	93,5
			6,0	66,6	94,4	94,3
500	0,80	0,10	3,0	90,7	94,8	93,6
			6,0	94,8	95,4	94,6
	0,40	0,50	3,0	92,0	95,3	95,6
			6,0	95,2	94,0	94,0
	0,49	0,49	3,0	88,2	94,9	94,3
			6,0	90,8	94,8	94,9

A Tabela 11 mostra os valores médios das estimativas pontuais de cada um dos parâmetros obtidas pelo método da máxima verossimilhança e as médias dos limites inferiores e superiores dos intervalos *bootstrap* com 95% de confiança. Notamos que as estimativas médias pontuais estão bem próximas dos verdadeiros valores dos parâmetros.

Já a Tabela 12 mostra as estimativas de Monte Carlo das medidas de eficiência dos estimadores de máxima verossimilhança. Podemos observar que as médias da razão entre $\mathbb{MSE}$ e $\mathbb{V}$, em sua maioria, são bem próximas de 1 (exceto para alguns valores associados a $\mu = 3$), indicando que os estimadores de μ, θ_1 e θ_2 são consistentes assintoticamente. Os vícios se aproximam de zero à medida que o tamanho do conjunto de dados cresce, enquanto que os valores do $\mathbb{MAPE}$ diminuem à medida que aumentamos o tamanho da amostra.

Tabela 11 – Média das estimativas e intervalos de confiança *bootstrap* dos parâmetros da distribuição 0,1-IB.

n	Parâmetros			$\bar{\tilde{\mu}}$	IC (95%) *bootstrap* de μ	$\bar{\tilde{\theta}}_1$	IC (95%) *bootstrap* de θ_1	$\bar{\tilde{\theta}}_2$	IC (95%) *bootstrap* de θ_2
	θ_1	θ_2	μ						
50	0,80	0,10	3,0	2,82	(1,64; 3,91)	0,78	(0,62; 0,88)	0,08	(0,01; 0,20)
			6,0	6,01	(4,91; 7,01)	0,79	(0,67; 0,88)	0,10	(0,03; 0,19)
	0,40	0,50	3,0	2,78	(1,60; 3,86)	0,38	(0,21; 0,51)	0,48	(0,30; 0,61)
			6,0	6,00	(4,91; 7,01)	0,40	(0,26; 0,53)	0,50	(0,36; 0,63)
	0,49	0,49	3,0	2,90	(2,08; 3,65)	0,47	(0,30; 0,59)	0,46	(0,29; 0,58)
			6,0	5,96	(5,26; 6,62)	0,48	(0,34; 0,60)	0,47	(0,33; 0,60)
200	0,80	0,10	3,0	2,86	(1,82; 3,73)	0,80	(0,73; 0,85)	0,10	(0,04; 0,14)
			6,0	6,00	(5,29; 6,68)	0,80	(0,74; 0,85)	0,10	(0,06; 0,14)
	0,40	0,50	3,0	2,87	(1,82; 3,75)	0,40	(0,32; 0,47)	0,50	(0,42; 0,57)
			6,0	6,02	(5,31; 6,70)	0,40	(0,33; 0,47)	0,50	(0,43; 0,57)
	0,49	0,49	3,0	2,85	(1,69; 3,89)	0,48	(0,41; 0,55)	0,49	(0,41; 0,55)
			6,0	5,96	(4,91; 6,94)	0,49	(0,42; 0,56)	0,49	(0,42; 0,56)
500	0,80	0,10	3,0	2,88	(2,23; 3,43)	0,80	(0,76; 0,83)	0,10	(0,07; 0,13)
			6,0	6,00	(5,56; 6,43)	0,80	(0,76; 0,83)	0,10	(0,07; 0,13)
	0,40	0,50	3,0	2,89	(2,25; 3,44)	0,40	(0,35; 0,44)	0,50	(0,45 0,54)
			6,0	5,99	(5,56; 6,42)	0,40	(0,36; 0,44)	0,50	(0,46; 0,54)
	0,49	0,49	3,0	2,80	(1,51; 3,92)	0,49	(0,44; 0,53)	0,49	(0,44; 0,53)
			6,0	6,01	(5,00; 6,95)	0,49	(0,45; 0,53)	0,49	(0,45; 0,53)

Tabela 12 – Medidas de eficiência do estimador de cada parâmetro da distribuição 0,1-IB.

n	Parâmetros			$\hat{\mu}$			$\hat{\theta}_1$			$\hat{\theta}_2$		
	θ_1	θ_2	μ	$\frac{\mathbb{MSE}(\hat{\mu})}{\mathbb{V}(\hat{\mu})}$	$\mathcal{B}(\hat{\mu})$	$\mathbb{MAPE}$	$\frac{\mathbb{MSE}(\hat{\theta}_1)}{\mathbb{V}(\hat{\theta}_1)}$	$\mathcal{B}(\hat{\theta}_1)$	$\mathbb{MAPE}$	$\frac{\mathbb{MSE}(\hat{\theta}_2)}{\mathbb{V}(\hat{\theta}_2)}$	$\mathcal{B}(\hat{\theta}_2)$	$\mathbb{MAPE}$
50	0,80	0,10	3,0	1,03	-0,18	26,59	1,08	-0,02	6,23	1,09	-0,02	46,50
			6,0	1,00	0,01	9,69	1,01	-0,01	5,61	1,00	0,00	35,08
	0,40	0,50	3,0	1,05	-0,22	26,29	1,05	-0,02	15,16	1,06	-0,02	13,06
			6,0	1,00	0,00	9,88	1,00	0,00	13,00	1,00	0,00	10,69
	0,49	0,49	3,0	1,01	-0,10	31,54	1,11	-0,02	12,47	1,20	-0,03	12,46
			6,0	1,00	-0,04	13,54	1,04	-0,01	11,47	1,05	-0,02	11,50
200	0,80	0,10	3,0	1,07	-0,14	13,51	1,00	0,00	2,97	1,03	0,00	20,10
			6,0	1,00	0,00	4,74	1,00	0,00	2,81	1,00	0,00	17,38
	0,40	0,50	3,0	1,06	-0,13	13,85	1,00	0,00	7,26	1,01	0,00	5,91
			6,0	1,00	-0,01	4,60	1,00	0,00	-6,91	1,00	0,00	5,70
	0,49	0,49	3,0	1,02	-0,15	26,45	1,02	-0,01	5,86	1,01	0,00	5,93
			6,0	1,00	-0,04	11,32	1,00	0,00	5,61	1,00	0,00	5,60
500	0,80	0,10	3,0	1,15	-0,12	8,99	1,01	0,00	1,81	1,03	0,00	12,25
			6,0	1,00	0,00	2,93	1,00	0,00	1,81	1,00	0,00	11,00
	0,40	0,50	3,0	1,13	-0,11	8,36	1,01	0,00	4,38	1,01	0,00	3,85
			6,0	1,00	-0,01	2,93	1,00	0,00	4,55	1,00	0,00	3,84
	0,49	0,49	3,0	1,08	-0,08	19,54	1,00	0,00	3,69	1,01	0,00	3,77
			6,0	1,00	0,01	6,42	1,00	0,00	3,58	1,00	0,00	3,62

5.4 Uma Análise com Dados Artificiais

Nesta seção apresentamos algumas aplicações em conjuntos de dados artificiais com o intuito de verificar o bom desempenho do método da máxima verossimilhança para a estimação dos parâmetros propostos e ainda verificar se algumas medidas de evidências são capazes de escolher adequadamente o conjunto de dados com distribuição $\boldsymbol{k}$-IPS. As medidas de evidências consideram a diferença entre a probabilidade empírica e a estimada, a partir da distribuição ajustada. As medidas abordadas são: Distância Euclidiana (*DE*), Divergência de Kullback-Leibler (*KL*) e Divergência de Kullback-Leibler Simétrica (*KLS*).

Para tal aplicação, geramos amostras de tamanho $n = 200$ das distribuições $\boldsymbol{k}$-IP, $\boldsymbol{k}$-IG e $\boldsymbol{k}$-IB, em que consideramos os seguintes pontos de modificações $\boldsymbol{k}$=(1,2), $\boldsymbol{k}$=(1,3) e $\boldsymbol{k}$=(2,3). Além disso, tomamos $\mu = 0,5$ e $2,5$, juntamente com os parâmetros θ_1 e θ_2 fixados em $0,60$ e $0,30$, respectivamente.

O estudo consiste na geração de uma amostra de dados artificiais para cada distribuição, a qual usaremos para fazer comparações entre as distribuições k_1-MPS, k_2-MPS e $\boldsymbol{k}$-IPS. Assim, ao final de cada aplicação, verificamos se a escolha foi adequada com base nas distribuições $\boldsymbol{k}$-IPS utilizadas. Além dos resultados obtidos pelas medidas de evidências, apresentamos os estimadores de máxima verossimilhança dos parâmetros e seus intervalos *bootstrap* não-paramétrico com 95% de confiança, seguindo a mesma metodologia utilizada anteriormente no estudo de simulação. Apresentamos estes resultados nas Tabelas 13 - 15.

Tabela 13 – Resultados da aplicação dos dados artificiais de distribuição $\boldsymbol{k}$-IP, com $\theta_1 = 0,60$ e $\theta_2 = 0,30$.

$\boldsymbol{k}$-IPS	Valores reais		Distribuição ajustada	$\hat{\mu}$	IC_μ (95%)	$\hat{\theta}_1$	IC_{θ_1} (95%)	$\hat{\theta}_2$	IC_{θ_2} (95%)	Medidas de evidências		
	k-inflação	μ								*DE*	*KL*	*KLS*
$\boldsymbol{k}$-IP	$\boldsymbol{k}$=(1,2)	0,5	1-MP	1,414	(1,325; 1,500)	0,409	(0,338; 0,483)	-	-	0,250	0,219	0,404
			2-MP	1,130	(1,090; 1,166)	-	-	0,134	(0,076; 0,192)	0,810	1,254	2,061
			1,2-IP	0,808	(0,000; 0,971)	0,558	(0,503; 0,650)	0,290	(0,244; 0,350)	**0,006**	**0,003**	**0,003**
		2,5	1-MP	1,794	(1,700; 1,903)	0,444	(0,376; 0,513)	-	-	0,241	0,211	0,401
			2-MP	1,365	(1,292; 1,443)	-	-	0,124	(0,065; 0,186)	0,806	1,318	2,126
			1,2-IP	2,187	(1,785; 2,770)	0,581	(0,527; 0,636)	0,301	(0,253; 0,353)	**0,004**	**0,008**	**0,003**
	$\boldsymbol{k}$=(1,3)	0,5	1-MP	2,043	(1,872; 2,208)	0,473	(0,405; 0,538)	-	-	0,255	0,340	0,624
			3-MP	1,047	(1,000; 1,098)	-	-	0,258	(0,208; 0,306)	0,689	0,834	1,458
			1,3-IP	0,955	(0,000; 1,112)	0,562	(0,506; 0,628)	0,300	(0,254; 0,347)	**0,026**	**0,013**	**0,026**
		2,5	1-MP	2,526	(2,405; 2,649)	0,511	(0,451; 0,572)	-	-	0,252	0,294	0,544
			3-MP	1,324	(1,234; 1,426)	-	-	0,253	(0,201; 0,304)	0,687	0,871	1,510
			1,3-IP	2,258	(1,855; 2,735)	0,585	(0,533; 0,639)	0,309	(0,261; 0,358)	**0,004**	**0,008**	**0,003**
	$\boldsymbol{k}$=(2,3)	0,5	2-MP	2,237	(2,095; 2,368)	0,441	(0,375; 0,508)	-	-	0,249	0,298	0,393
			3-MP	1,970	(1,899; 2,038)	-	-	0,158	(0,102; 0,213)	0,787	1,304	2,086
			2,3-IP	0,467	(0,195; 0,847)	0,582	(0,531; 0,631)	0,306	(0,260; 0,352)	**0,004**	**0,000**	**0,000**
		2,5	2-MP	2,682	(2,579; 2,787)	0,499	(0,437; 0,564)	-	-	0,242	0,291	0,530
			3-MP	2,213	(2,155; 2,275)	-	-	0,165	(0,109; 0,221)	0,790	1,308	2,078
			2,3-IP	2,374	(1,813; 2,949)	0,599	(0,550; 0,652)	0,311	(0,265; 0,360)	**0,008**	**0,009**	**0,007**

Primeiramente, a Tabela 13 apresenta os resultados dos ajustes das distribuições k_1-MP, k_2-MP e $\boldsymbol{k}$-IP. Observamos que as estimativas dos parâmetros θ_1 e θ_2 foram próximas dos verdadeiros valores, principalmente nas distribuições $\boldsymbol{k}$-IP. Por outro lado, considerando o parâmetro μ notamos que, além das estimativas serem satisfatórias, apenas as distribuições $\boldsymbol{k}$-IP

foram capazes de compreender todos seus verdadeiros valores em seus intervalos de confiança. Por fim, para cada par de pontos de modificação que tomamos como referência, observamos que os menores valores das medidas de evidência calculados foram dadas pelas distribuições $1,2$-IP, $1,3$-IP e $2,3$-IP, reforçando novamente a adequabilidade da distribuição proposta.

Tabela 14 – Resultados da aplicação dos dados artificiais de distribuição $\boldsymbol{k}$-IG, com $\theta_1 = 0,60$ e $\theta_2 = 0,30$.

$\boldsymbol{k}$-IPS	Valores reais		Distribuição ajustada	$\hat{\mu}$	IC_{μ} (95%)	$\hat{\theta}_1$	IC_{θ_1} (95%)	$\hat{\theta}_2$	IC_{θ_2} (95%)	Medidas de evidências		
	k-inflação	μ								*DE*	*KL*	*KLS*
$\boldsymbol{k}$-IG	$\boldsymbol{k}$=(1,2)	0,5	1-MG	1,481	(1,378; 1,577)	0,486	(0,425; 0,549)	-	-	0,213	0,184	0,423
			2-MG	1,050	(1,009; 1,093)	-	-	0,209	(0,156; 0,261)	0,746	0,955	1,640
			$1,2$-IG	0,644	(0,000; 0,903)	0,581	(0,531; 0,640)	0,299	(0,253; 0,348)	**0,017**	**0,003**	**0,006**
		2,5	1-MG	1,853	(1,700; 2,041)	0,508	(0,448; 0,570)	-	-	0,210	0,231	0,472
			2-MG	1,263	(1,161; 1,390)	-	-	0,214	(0,162; 0,267)	0,747	1,040	1,732
			$1,2$-IG	2,265	(1,433; 3,375)	0,601	(0,552; 0,653)	0,309	(0,262; 0,357)	**0,023**	**0,053**	**0,067**
	$\boldsymbol{k}$=(1,3)	0,5	1-MG	2,100	(1,939; 2,248)	0,501	(0,441; 0,562)	-	-	0,252	0,325	0,760
			3-MG	1,054	(1,007; 1,101)	-	-	0,256	(0,207; 0,305)	0,690	0,833	1,465
			$1,3$-IG	0,798	(0,000; 1,059)	0,580	(0,529; 0,633)	0,299	(0,253; 0,346)	**0,027**	**0,008**	**0,017**
		2,5	1-MG	2,537	(2,384; 2,703)	0,526	(0,467; 0,587)	-	-	0242	0,282	0,718
			3-MG	1,307	(1,201; 1,434)	-	-	0,251	(0,202; 0,304)	0,710	0,939	1,605
			$1,3$-IG	2,434	(1,747; 3,244)	0,602	(0,552; 0,651)	0,300	(0,254; 0,347)	**0,021**	**0,029**	**0,051**
	$\boldsymbol{k}$=(2,3)	0,5	2-MG	2,418	(2,266; 2,562)	0,575	(0,520; 0,628)	-	-	0,214	0,294	0,370
			3-MG	1,949	(1,892; 2,003)	-	-	0,219	(0,169; 0,268)	0,782	1,215	1,973
			$2,3$-IG	0,459	(0,132; 0,745)	0,632	(0,585; 0,680)	0,293	(0,249; 0,340)	**0,005**	**0,001**	**0,001**
		2,5	2-MG	2,790	(2,598; 2,995)	0,536	(0,480; 0,593)	-	-	0,230	0,274	0,644
			3-MG	2,168	(2,054; 2,297)	-	-	0,241	(0,190; 0,293)	0,728	1,067	1,782
			$2,3$-IG	2,594	(1,741; 3,530)	0,587	(0,536; 0,638)	0,306	(0,260; 0,354)	**0,032**	**0,058**	**0,085**

A Tabela 14 mostra os resultados dos ajustes das distribuições k_1-MG, k_2-MG e $\boldsymbol{k}$-IG. Notamos que as estimativas dos parâmetros θ_1 e θ_2 das distribuições $\boldsymbol{k}$-IG foram as mais próximas dos verdadeiros valores e, quando nos deparamos com o parâmetro μ, observamos que apenas nestas distribuições seus verdadeiros valores podem ser encontrados nos intervalos de confiança. Além disso, para cada par de pontos de modificação considerado, observamos que os menores valores das medidas de evidência calculados foram dadas pelas distribuições $1,2$-IG, $1,3$-IG e $2,3$-IG, reforçando novamente a adequabilidade da distribuição proposta.

Tabela 15 – Resultados da aplicação dos dados artificiais de distribuição $\boldsymbol{k}$-IB, com $\theta_1 = 0,60$ e $\theta_2 = 0,30$.

$\boldsymbol{k}$-IPS	Valores reais		Distribuição ajustada	$\hat{\mu}$	IC_{μ} (95%)	$\hat{\theta}_1$	IC_{θ_1} (95%)	$\hat{\theta}_2$	IC_{θ_2} (95%)	Medidas de evidências		
	k-inflação	μ								*DE*	*KL*	*KLS*
$\boldsymbol{k}$-IB	$\boldsymbol{k}$=(1,2)	0,5	1-MB	1,402	(1,315; 1,488)	0,398	(0,325; 0,473)	-	-	0,247	0,196	0,375
			2-MB	1,149	(1,109; 1,186)	-	-	0,111	(0,051; 0,171)	0,833	1,362	2,221
			$1,2$-IB	0,865	(0,000; 1,015)	0,552	(0,495; 0,652)	0,283	(0,237; 0,346)	**0,007**	**0,003**	**0,005**
		2,5	1-MB	1,806	(1,723; 1,906)	0,446	(0,378; 0,516)	-	-	0,226	0,177	0,374
			2-MB	1,391	(1,322; 1,462)	-	-	0,095	(0,032; 0,159)	0,832	1,477	2,341
			$1,2$-IB	2,262	(1,892; 2,834)	0,587	(0,534; 0,641)	0,299	(0,250; 0,350)	**0,004**	**0,002**	**0,003**
	$\boldsymbol{k}$=(1,3)	0,5	1-MB	2,053	(1,871; 2,231)	0,480	(0,411; 0,546)	-	-	0,253	0,330	0,716
			3-MB	1,037	(0,986; 1,083)	-	-	0,262	(0,213; 0,310)	0,683	0,807	1,442
			$1,3$-IB	0,948	(0,000; 1,092)	0,562	(0,506; 0,642)	0,300	(0,255; 0,348)	**0,029**	**0,018**	**0,040**
		2,5	1-MB	2,607	(2,515; 2,702)	0,499	(0,441; 0,558)	-	-	0,269	0,318	0,656
			3-MB	1,326	(1,239; 1,423)	-	-	0,289	(0,237; 0,344)	0,633	0,776	1,364
			$1,3$-IB	2,313	(1,965; 2,710)	0,564	(0,515; 0,616)	0,340	(0,291; 0,388)	**0,025**	**0,030**	**0,013**
	$\boldsymbol{k}$=(2,3)	0,5	2-MB	2,226	(2,091; 2,353)	0,416	(0,348; 0,487)	-	-	0,255	0,426	0,380
			3-MB	2,000	(1,926; 2,069)	-	-	0,133	(0,076; 0,190)	0,806	1,370	2,199
			$2,3$-IB	0,448	(0,191; 0,855)	0,583	(0,531; 0,631)	0,307	(0,261; 0,352)	**0,004**	**0,000**	**0,000**
		2,5	2-MB	2,662	(2,568; 2,754)	0,484	(0,420; 0,551)	-	-	0,227	0,225	0,500
			3-MB	2,244	(2,194; 2,297)	-	-	0,131	(0,072; 0,190)	0,815	1,385	2,197
			$2,3$-IB	2,389	(1,953; 2,827)	0,593	(0,543; 0,647)	0,305	(0,258; 0,356)	**0,003**	**0,002**	**0,003**

Por último, a Tabela 15 apresenta os resultados dos ajustes das distribuições k_1-MB, k_2-MB e $\boldsymbol{k}$-IB. Notamos que as estimativas dos parâmetros foram mais próximas dos verdadeiros valores nas distribuições $\boldsymbol{k}$-IB, cujos intervalos de confiança compreendem estes. E ainda, para cada par de pontos de modificação, observamos que os menores valores das medidas de evidência calculados foram apresentadas pelas distribuições $1,2$-IB, $1,3$-IB e $2,3$-IB, reforçando novamente a adequabilidade da distribuição proposta.

CAPÍTULO

6

APLICAÇÕES

Neste capítulo apresentamos algumas aplicações com conjuntos de dados reais considerando as distribuições Poisson, Geométrica e Binomial k_1 e k_2 Inflacionada. Para este estudo, o foco principal está em verificar se as distribuições $\boldsymbol{k}$-IPS são capazes de explicar adequadamente o comportamento dos dados que apresentam altas frequências (inflação) em duas observações (k_1 e k_2). Além disso, iremos realizar um estudo comparativo entre os ajustes das distribuições mais simples (distribuições PS tradicionais e as distribuições k-MPS) e avaliar se, de fato, as distribuições $\boldsymbol{k}$-IPS retornam bons resultados.

Para o procedimento de estimação dos parâmetros, consideramos o Método da Máxima Verossimilhança. Posteriormente, comparamos as distribuições ajustadas via Teste de Aderência Kolmogorov-Smirnov (ver Miller (1956) e Conover (1999)) e pelas seguintes medidas de evidências DE, KL e KLS, as quais verificam a diferença entre a proporção amostral e a probabilidade estimada. Todas essas métricas tem como finalidade verificar se a distribuição ajustada é ou não adequada para explicar o comportamento de cada conjunto de dados.

Apresentamos algumas aplicações em conjuntos de dados reais considerando as distribuições $\boldsymbol{k}$-IP, $\boldsymbol{k}$-IG e $\boldsymbol{k}$-IB. Comparamos cada uma destas distribuições ajustadas com suas respectivas distribuições PS, k_1-MPS e k_2-MPS associadas, que também serão ajustadas aos dados. Além das estimativas pontuais, apresentamos também os respectivos intervalos *bootstrap* com 95% de confiança. Adicionalmente, para o procedimento *bootstrap*, consideramos o *bootstrap* não-paramétrico a fim de obter amostras com as mesmas características da amostra original, com $B = 5.000$ réplicas.

Problema 1: Coelhos brancos que nasceram mortos na Nova Zelândia

Consideramos o conjunto de dados utilizado por Morgan, Palmer e Ridout (2007) que se refere ao número de coelhos brancos que nasceram mortos na Nova Zelândia. A Tabela 16 apresenta a distribuição de frequência do número de coelhos brancos nascidos mortos, assim como a média amostral ($\bar{z}$) e o desvio-padrão amostral (DP).

Tabela 16 – Distribuição de frequência e estatísticas descritivas do número de coelhos brancos que nasceram mortos na Nova Zelândia.

Número de coelhos (z_i)	0	1	2	3	4	5	6	7	8	11	Total	$\bar{z}$	DP
Frequência (f_i)	314	48	20	7	5	2	2	1	2	1	402	0,460	1,229

Fonte: Adaptada de Morgan, Palmer e Ridout (2007).

Observando a Tabela 16, podemos notar que as maiores frequências são das observações zero e um, quando comparadas com as demais. Sendo assim, ajustamos as distribuições Poisson, 0-MP, 1-MP e $0,1$-IP a este conjunto de dados.

A Tabela 17 apresenta as estimativas dos parâmetros μ, θ_1 e θ_2 obtidas pelo procedimento de máxima verossimilhança (e os respectivos intervalos *bootstrap* não paramétrico com 95% de confiança). Além disso, esta Tabela apresenta para cada distribuição ajustada a frequência esperada de cada observação (E_i), o valor da estatística e o valor crítico do teste Kolmogorov-Smirnov (KS_{calc} e KS_{crit}, respectivamente) e ainda, os valores das medidas de evidências (Distância Euclidiana (DE), Divergência de Kullback-Leibler (KL) e Divergência de Kullback-Leibler Simétrica (KLS)).

Ao comparar as distribuições ajustadas observamos que, para distribuição $0,1$-IP, a estatística de teste KS_{calc} e as medidas de evidências DE, KL e KLS apresentaram os menores valores, não rejeitando a hipótese de adequabilidade da distribuição $0,1$-IP aos dados com um nível de 5% de significância. Este fato pode ser facilmente observado, ao compararmos as frequências observadas com as frequências esperadas obtidas com cada distribuição ajustada. Apesar do teste não rejeitar as distribuições 0-MP (ou 0-IP) e 1-MP (ou 1-DP), notamos que as suas medidas de evidências, que são baseadas na máxima distância entre a distribuição acumulada empírica e teórica, são um pouco maiores. Ressaltamos ainda que a distribuição 1-MP resultou em uma estimativa negativa do parâmetro de modificação, o que nos faz levar ao caso particular da distribuição k modificada, isto é, a distribuição 1 deflacionado (1-DP).

Tabela 17 – Estimativas e intervalos de confiança dos parâmetros das distribuições Poisson, 0-MP, 1-MP e $0,1$-IP ajustadas aos dados referentes ao número de coelhos brancos nascidos mortos, juntamente com os resultados de comparação das distribuições ajustadas.

z_i	O_i	P	0-MP	1-MP	$0,1$-IP
		E_i			
0	314	254	314	292	314
1	48	117	33	48	48
2	20	27	28	50	13
3	7	4	16	10	12
4	5	0	7	2	8
5	2	0	3	0	4
≥ 6	6	0	1	0	3
Total	402	402	402	402	402
		Estimativas			
Parâmetros	μ	0,460 (0,351; 0,592)	1,729 (1,230; 2,250)	0,588 (0,480; 0,693)	2,695 (1,594; 3,702)
	θ_1	-	0,734 (0,670; 0,786)	-0,307 (-0,367; -0,237)	0,772 (0,720; 0,812)
	θ_2	-	-	-	0,095 (0,041; 0,133)
Teste KS	KS_{calc}	0,150	0,038	0,055	0,017
	KS_{crit}	0,068	0,068	0,068	0,068
Medidas de evidências	DE	0,229	0,050	0,095	**0,023**
	KL	0,267	0,058	0,165	**0,027**
	KLS	0,401	0,091	0,232	**0,044**

Fonte: Elaborada pelo autor.

Problema 2: Acidentes de trânsito de veículos pesados na Índia

Consideramos o conjunto de dados coletado por Sharma e Landge (2013) referente ao número de acidentes de trânsito que envolvem veículos pesados, ocorridos no ano de 2010 em uma estrada rural na Índia. Apresentamos na Tabela 18 a distribuição de frequência do número de acidentes de trânsito que envolvem veículos pesados na Índia e as estatísticas descritivas dos dados (média e desvio-padrão amostral).

Tabela 18 – Distribuição de frequência e estatísticas descritivas do número de acidentes de trânsito que envolvem veículos pesados no ano de 2010 em uma estrada rural na Índia.

Número de acidentes (z_i)	0	1	2	3	4	5	6	8	Total	$\bar{z}$	DP
Frequência (f_i)	55	26	4	3	3	1	3	1	96	0,896	1,579

Fonte: Adaptada de Sharma e Landge (2013).

Ao observar a Tabela 18, notamos altas frequências das observações zero e um, quando comparadas com as frequências das demais observações. Dessa forma, ajustamos as distribuições Poisson, 0-MP, 1-MP e $0,1$-IP a este conjunto de dados.

A Tabela 19 apresenta as estimativas dos parâmetros μ, θ_1 e θ_2 obtidas pelo procedimento de máxima verossimilhança e os respectivos intervalos *bootstrap* com 95% de confiança. Além disso, esta Tabela apresenta também para cada distribuição ajustada a frequência esperada para cada observação (E_i), o valor da estatística e o valor crítico do teste Kolmogorov-Smirnov (KS_{calc} e KS_{crit}, respectivamente) e os valores das medidas de evidências (Distância Euclidiana (DE), Divergência de Kullback-Leibler (KL) e Divergência de Kullback-Leibler Simétrica (KLS)).

Analisando os resultados apresentados, podemos notar que, para a distribuição $0,1$-IP, a estatística de teste KS_{calc} e as medidas de evidências DE, KL e KLS apresentaram os menores valores quando comparadas com as demais distribuições ajustadas. Assim, a hipótese de adequabilidade da distribuição $0,1$-IP não é rejeitada, com um nível de 5% de significância. Este fato pode também ser facilmente observado ao compararmos as frequências observadas com as frequências esperadas obtidas com cada distribuição ajustada. Apesar do teste não rejeitar as distribuições 0-MP (ou 0-IP) e 1-MP (1-DP), observamos que suas medidas de evidências são maiores. Por fim, observamos que a distribuição 1-MP ajustada resultou em estimativa negativa do parâmetro de modificação, indicando ser o caso particular da distribuição k modificada, isto é, a distribuição 1 deflacionado (1-DP).

Tabela 19 – Estimativas e intervalos de confiança dos parâmetros das distribuições Poisson, 0-MP, 1-MP e $0,1$-IP ajustadas aos dados referentes ao número de acidentes de trânsito que envolvem veículos pesados, juntamente com os resultados de comparação das distribuições ajustadas.

z_i	O_i	P	0-MP	1-MP	$0,1$-IP
		E_i			
0	55	39	55	45	55
1	26	35	15	26	26
2	4	16	13	18	3
3	3	5	8	6	4
4	3	1	3	1	3
≥ 5	5	0	2	0	5
Total	96	96	96	96	96
		Estimativas			
Parâmetros	μ	0,896 (0,615; 1,219)	1,720 (1,000; 2,432)	0,910 (0,639; 1,196)	3,538 (2,170; 4,699)
	θ_1	-	0,480 (0,283; 0,606)	-0,151 (-0,284; 0,002)	0,568 (0,462; 0,664)
	θ_2	-	-	-	0,252 (0,154; 0,345)
Teste KS	KS_{calc}	0,165	0,111	0,110	0,015
	KS_{crit}	0,139	0,139	0,139	0,139
Medidas de evidências	DE	0,230	0,157	0,192	**0,026**
	KL	0,319	0,172	0,299	**0,029**
	KLS	0,521	0,322	0,513	**0,037**

Fonte: Elaborada pelo autor.

Problema 3: Atos criminosos vistos pela sociologia

Consideramos o conjunto de dados divulgados por Carr-Hill e Macdonald (1973) que diz a respeito a um estudo da área de sociologia que observou, após um período de nove anos, o número de atos criminosos de pessoas que apresentavam comportamentos agressivos. A Tabela 20 apresenta a distribuição de frequência do número de atos criminosos, assim como a média amostral ($\bar{z}$) e o desvio-padrão amostral (DP).

Tabela 20 – Distribuição de frequência e estatísticas descritivas do número de atos criminosos em pacientes com comportamentos agressivos.

Número de atos (z_i)	0	1	2	3	4	5	Total	$\bar{z}$	DP
Frequência (f_i)	4.037	219	29	9	5	2	4.301	0,078	0,348

Fonte: Adaptada de Carr-Hill e Macdonald (1973).

A partir da Tabela 20, podemos notar que as maiores frequências são das observações zero e um, quando comparadas com as demais. Por isso, vamos ajustar as distribuições Poisson, 0-MP, 1-MP e $0,1$-IP.

A Tabela 21 apresenta as estimativas dos parâmetros μ, θ_1 e θ_2 obtidas pelo procedimento de máxima verossimilhança e os respectivos intervalos *bootstrap* com 95% de confiança. Esta Tabela também apresenta para cada distribuição ajustada a frequência esperada para cada observação (E_i), o valor da estatística e o valor crítico do teste Kolmogorov-Smirnov (KS_{calc} e KS_{crit}, respectivamente) e os valores das medidas de evidências (DE, KL e KLS).

A partir dos resultados apresentados na Tabela, notamos que a estatística de teste KS_{calc} apresentou o menor valor para a distribuição $0,1$-IP, não rejeitando a hipótese de adequabilidade a nível de 5% de significância. Observamos esse fato ainda ao compararmos as frequências observadas com as frequências esperadas obtidas com cada distribuição ajustada. Apesar do teste não rejeitar as distribuições Poisson, 0-MP (ou 0-IP) e 1-MP (ou 1-DP) temos valores maiores para as medidas de evidências. Vale ressaltar que a distribuição 1-MP ajustada resultou em estimativa negativa do parâmetro de modificação, o que nos leva ao caso particular da distribuição k modificada, isto é, a distribuição 1 deflacionado (1-DP).

Tabela 21 – Estimativas e intervalos de confiança dos parâmetros das distribuições Poisson, 0-MP, 1-MP e 0,1-IP ajustadas aos dados referentes ao número de atos criminosos, juntamente com os resultados de comparação das distribuições ajustadas.

z_i	O_i	P	0-MP	1-MP	0,1-IP
		E_i			
0	4.037	3.980	4.037	4.026	4.037
1	219	309	205	219	219
2	29	12	50	53	29
3	9	0	8	3	11
4	5	0	1	0	4
5	2	0	0	0	1
Total	4.301	4.301	4.301	4.301	4.301
		Estimativas			
Parâmetros	μ	0,078 (0,068; 0,088)	0,490 (0,348; 0,631)	0,162 (0,137; 0,185)	1,175 (0,935; 0,610)
	θ_1	-	0,842 (0,788; 0,874)	-0,101 (-0,123; -0,077)	0,929 (0,917; 0,939)
	θ_2	-	-	-	0,039 (0,030; 0,048)
Teste KS	KS_{calc}	0,013	0,003	0,003	0,000
	KS_{crit}	0,021	0,021	0,021	0,021
Medidas de evidências	DE	0,026	0,006	0,006	**0,001**
	KL	0,021	0,003	0,008	**0,000**
	KLS	0,030	0,006	0,012	**0,001**

Fonte: Elaborada pelo autor.

Problema 4: Variação da cotação do euro

Inicialmente, consideramos os conjuntos de dados referentes aos valores semanais (abertura e fechamento) da cotação do euro em relação ao real no período entre Janeiro de 2000 e Outubro de 2019 (Investing (2019)). A partir destes dados, calculamos a variação da cotação semanal, conhecida como variação cambial (CALCBANK, 2019), dada por:

$$\begin{aligned} \text{Variação cambial} &= \left(\frac{\text{Valor atual}}{\text{Valor inicial}} - 1\right) \cdot 100\% \\ &= \left(\frac{\text{Valor de fechamento}}{\text{Valor de abertura}} - 1\right) \cdot 100\%. \end{aligned}$$

A Figura 5 apresenta as variações cambiais semanais, as quais podem ser positivas ou negativas (indicando um aumento ou queda do valor do euro, respectivamente). Podemos observar que há muitas oscilações no decorrer das semanas, com variação máxima de 13,97% (semana de setembro de 2002) e variação mínima de -9,25% (semana de outubro de 2008).

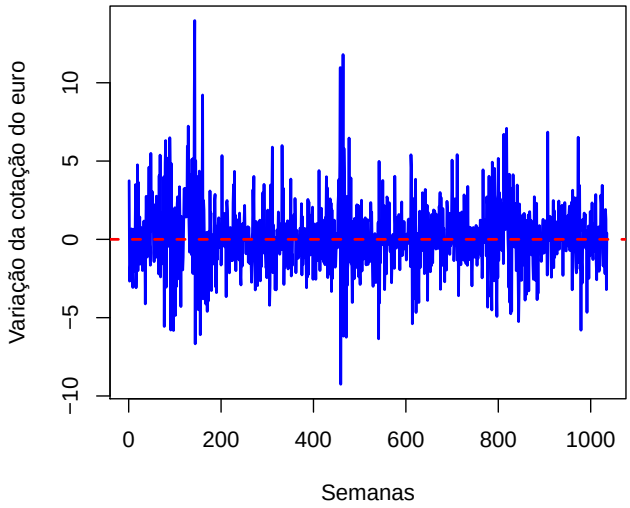


Figura 5 – Variação cambial semanal do euro no período entre Janeiro de 2000 e Outubro 2019.

Fonte: Elaborada pelo autor.

Temos interesse em verificar se em uma determinada semana ocorreu uma variação negativa (sucesso, indicando que o preço de fechamento do euro foi menor do que o preço de abertura de determinada semana) ou positiva (fracasso, indicando que o preço de fechamento do euro foi maior do que o preço de abertura de determinada semana). Denotaremos por η a probabilidade de sucesso (variação negativa) no processo de Bernoulli ($\eta = \frac{1}{1+\mu}$, com $\eta \in (0,1)$) e, consequentemente, $1-\eta$ a probabilidade de fracasso (variação positiva). Portanto, interpretamos η como a probabilidade de ocorrer uma queda na cotação do euro no final de cada semana.

A partir do processo de Bernoulli, definimos uma variável aleatória Z como sendo o número de semanas consecutivas que houve variação positiva da cotação do euro (fracasso) até a ocorrência de uma variação negativa (sucesso). Sendo assim, consideramos que a variável aleatória Z corresponde a um processo Geométrico. A Tabela 22 apresenta a distribuição de frequência deste conjunto, bem como a média amostral ($\bar{z}$) e o desvio-padrão amostral (DP).

Tabela 22 – Distribuição de frequência e estatísticas descritivas do número de semanas consecutivas com variação positiva da cotação do euro até a ocorrência de uma variação negativa, no período entre Janeiro de 2000 e Outubro de 2019.

Número de variações positivas (z_i)	0	1	2	3	4	5	6	7	9	13	Total	$\bar{z}$	DP
Frequência (f_i)	240	123	65	35	16	10	6	1	1	1	498	1,078	1,514

Fonte: Elaborada pelo autor.

Das frequências apresentas na Tabela 22, notamos valores altos para as observações zero e um, nos dando evidências de que um ajuste com as distribuições Geométrica, 0-MG, 1-MG e $0,1$-IG poderia ser adequado.

Apresentamos na Tabela 23 as estimativas dos parâmetros μ, θ_1 e θ_2 obtidas pelo procedimento de máxima verossimilhança e os respectivos intervalos *bootstrap* com 95% de confiança. Adicionalmente, apresentamos também nesta Tabela a frequência esperada para cada observação (E_i), o valor da estatística de teste Kolmogorov-Smirnov (KS_{calc}) e seu o valor crítico (KS_{crit}), juntamente com as medidas de evidências DE, KL e KLS, para cada distribuição ajustada.

Analisando os resultados desta Tabela, observamos que, ao comparar as distribuições ajustadas, os valores da estatística de teste KS na distribuição Geométrica, 0-MG (ou 0-IG) e 1-MG (ou 1-DG, caso particular da distribuição k modificada) foram iguais ($KS_{calc} = 0,004$). Observamos ainda que os valores das medidas de evidências DE, KL e KLS também são próximas ao considerarmos as mesmas distribuições. A justificativa destes valores serem aproximadamente iguais é que as estimativas de θ_1 e θ_2 são próximas de zero e que o mesmo está contido nos intervalos de confiança, indicando que os parâmetros não são significantes e, consequentemente, ausência de k-inflação nos dados. Portanto, os dados podem ser explicados por uma distribuição Geométrica tradicional. Vemos claramente estes fatos ao comparar as frequências observadas com as frequências esperadas segundo as distribuições ajustadas. Por outro lado, vale ressaltar que, apesar da complexidade da distribuição $0,1$-IG, esta também mostrou ser adequada para explicar o comportamento dos dados, uma vez que apresentou valores baixos para os parâmetros de $\boldsymbol{k}$-modificação das probabilidades.

Tabela 23 – Estimativas e intervalos de confiança dos parâmetros das distribuições Geométrica, 0-MG, 1-MG e 0,1-IG ajustadas aos dados referentes ao número de semanas consecutivas com variação positiva da cotação do euro até a ocorrência de uma variação negativa, juntamente com os resultados de comparação das distribuições ajustadas.

z_i	O_i	G	0-MG	1-MG	0,1-IG
		E_i			
0	240	240	240	240	240
1	123	124	124	123	123
2	65	65	65	65	58
3	35	33	33	33	34
4	16	17	17	17	19
5	10	9	9	9	11
6	6	5	5	5	6
≥ 7	3	5	5	5	8
Total	498	498	498	498	498
		Estimativas			
Parâmetros	μ	1,078 (0,948; 1,217)	1,081 (0,900; 1,273)	1,078 (0,951; 1,217)	1,344 (1,177; 1,554)
	θ_1	-	0,002 (-0,128; 0,113)	-0,004 (-0,052; 0,048)	0,130 (0,016; 0,235)
	θ_2	-	-	-	0,045 (0,005; 0,106)
Teste KS	KS_{calc}	0,004	0,004	0,004	0,019
	KS_{crit}	0,061	0,061	0,061	0,061
Medidas de evidências	*DE*	**0,008**	**0,007**	**0,007**	0,014
	KL	**0,006**	**0,006**	**0,006**	0,009
	KLS	**0,010**	**0,010**	**0,010**	0,013

Fonte: Elaborada pelo autor.

Problema 5: Variação da temperatura máxima média da cidade do Rio de Janeiro

Apresentamos agora a análise de dados referentes à variação de temperatura máxima média mensal da cidade do Rio de Janeiro (em graus Celsius, ^{o}C) no período entre 1961 e 2017.

Com o conjunto de dados das temperaturas máximas médias mensais, obtidas em INMET

(2019), foi calculada a média global destas temperaturas. A variação de temperatura foi obtida a partir da diferença entre as temperaturas máximas médias mensais e a média global. Assim, variações positivas indicam que as temperaturas máximas médias mensais são maiores do que a média global. Similarmente, variações negativas indicam que as temperaturas máximas médias mensais são menores do que a média global. A Figura 6 apresenta as variações de temperatura máxima média mensal. Podemos observar que as temperaturas oscilam no decorrer dos meses, com variação máxima de 8,97^{o}C (fevereiro de 2003) e variação mínima de -6,34 (julho de 1964).

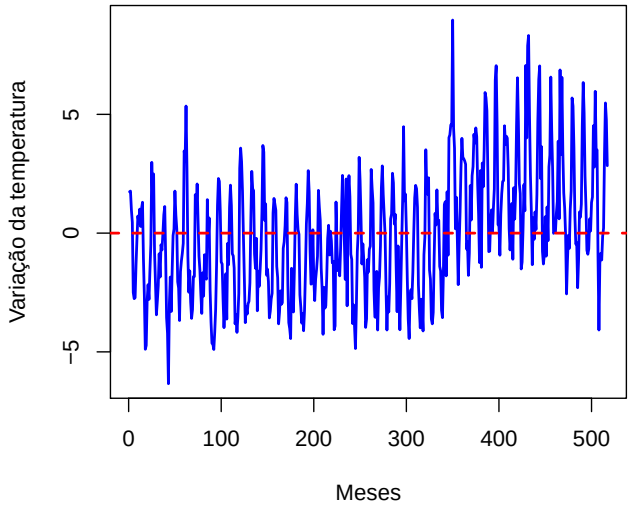


Figura 6 – Variação da temperatura máxima média mensal da cidade do Rio de Janeiro no período entre 1961 e 2017.

Fonte: Elaborada pelo autor.

O interesse deste estudo é verificar se em um determinado mês ocorreu uma variação positiva (sucesso) ou negativa (fracasso). Mais uma vez, ao considerarmos η como a probabilidade de sucesso no processo de Bernoulli ($\eta = \frac{1}{1+\mu}$, com $\eta \in (0,1)$), podemos interpretá-lo como sendo a probabilidade de um aumento da temperatura máxima média em um determinado mês.

A partir do processo de Bernoulli, definimos uma variável aleatória Z como sendo o número de meses consecutivos que houve variação negativa de temperatura máxima média mensal (fracasso) até a ocorrência de uma variação positiva (sucesso). Dessa forma, para esta variável aleatória, é natural a suposição de uma distribuição Geométrica. A Tabela 24 apresenta a distribuição de frequência deste conjunto de dados, a média amostral ($\bar{z}$) e o desvio-padrão amostral (DP). Observamos altas frequências das observações zero, um e oito, quando comparadas com as frequências das demais observações. Dessa forma, decidimos ajustar para este conjunto de dados a distribuição Geométrica, as distribuições 0-MG, 1-MG, 8-MG e as distribuições $0,1$-IG e $0,8$-IG.

Tabela 24 – Distribuição de frequência e estatísticas descritivas do número de meses consecutivos com variação negativa da temperatura máxima média até a ocorrência de uma variação positiva, no período entre 1961 e 2017.

Número de variações negativas (z_i)	0	1	2	3	4	5	6	7	8	9	10	Total	$\bar{z}$	DP
Frequência (f_i)	190	14	4	4	2	6	3	4	11	5	2	245	1,106	2,530

Fonte: Elaborada pelo autor.

A Tabela 25 apresenta as estimativas dos parâmetros μ, θ_1 e θ_2 obtidas pelo procedimento de máxima verossimilhança e os respectivos intervalos *bootstrap* com 95% de confiança. Além disso, esta Tabela apresenta, para cada distribuição ajustada, a frequência esperada para cada observação (E_i), o valor da estatística e o valor crítico do teste Kolmogorov-Smirnov (KS_{calc} e KS_{crit}, respectivamente) e as medidas de evidências DE, KL e KLS.

Observando seus resultados notamos que, para as distribuições 0-MG, $0,1$-IG e $0,8$-IG, tanto a estatística de teste KS_{calc} quanto as medidas de evidências DE, KL e KLS, apresentaram os menores valores quando comparadas com as demais distribuições ajustadas. Dessa forma, a hipótese de adequabilidade das distribuições 0-MG, $0,1$-IG e $0,8$-IG não é rejeitada a um nível de 5% de significância. Ao compararmos as frequências observadas com as frequências esperadas, também observamos este fato facilmente.

Por outro lado, ao compararmos essas três distribuições, notamos que a estimativa do parâmetro θ_2 da distribuição $0,1$-IG não é estatisticamente significante (já que o IC contém o valor zero). Assim, podemos afirmar que este conjunto de dados apresenta uma modificação significativa na observação zero e seu comportamento pode ser explicado adequadamente pela distribuição 0-MG (ou 0-IG). E, em contrapartida, analisando os valores dados pelas medidas de evidências, concluímos que a distribuição $0,8$-IG se ajusta ainda melhor aos dados do que a própria distribuição 0-IG.

Tabela 25 – Estimativas e intervalos de confiança dos parâmetros das distribuições Geométrica, 0-MG, 1-MG, 8-MG, 0,1-IG e 0,8-IG ajustadas aos dados referentes ao número de meses consecutivos com variação negativa da temperatura máxima média até a ocorrência de uma variação positiva, juntamente com os resultados de comparação das distribuições ajustadas.

z_i	O_i	G	0-MG	1-MG	8-MG	0,1-IG	0,8-IG
		E_i					
0	190	116	190	148	129	190	190
1	14	61	11	14	58	14	9
2	4	32	9	40	26	8	7
3	4	17	7	21	12	6	6
4	2	9	6	11	5	5	5
5	6	5	4	5	2	4	4
6	3	2	4	3	1	4	3
7	4	1	3	1	1	3	3
8	11	1	2	1	11	2	11
≥ 9	7	1	9	1	0	9	7
Total	245	245	245	245	245	245	245
		Estimativas					
Parâmetros	μ	1,106 (0,796; 1,437)	3,871 (3,091; 4,716)	1,084 (0,847; 1,347)	0,813 (0,563; 1,093)	4,295 (3,574; 5,020)	4,123 (3,471; 4,867)
	θ_1	-	0,718 (0,650; 0,783)	-0,257 (-0,292; -0,214)	0,044 (0,020; 0,069)	0,728 (0,657; 0,790)	0,730 (0,665; 0,791)
	θ_2	-	-	-	-	0,018 (0,001; 0,052)	0,037 (0,012; 0,065)
Teste KS	KS_{calc}	0,301	0,037	0,173	0,248	0,037	0,021
	KS_{crit}	0,087	0,087	0,087	0,087	0,087	0,087
Medidas de evidências	DE	0,383	0,049	1,451	0,322	0,045	**0,033**
	KL	0,453	0,086	1,645	0,345	0,084	**0,049**
	KLS	0,987	0,116	5,490	0,728	0,103	**0,054**

Fonte: Elaborada pelo autor.

Problema 6: Quantidade da vogal A em palavras com treze letras

Para esta aplicação, consideramos o conjunto de dados obtidos a partir da contagem de letras da vogal “A” em palavras da língua portuguesa que são compostas por treze letras e que a última letra corresponde a "r", as quais incluíam verbos e substantivos. As palavras foram coletadas de Dicio (2019). A Tabela 26 apresenta a distribuição de frequência deste conjunto de dados, bem como a média e o desvio-padrão amostral.

Com base nos valores apresentados na Tabela 26, observamos altas frequências das observações um e dois, quando comparadas com as frequências das demais observações. Sendo

assim, para este conjunto de dados, optamos em ajustar as distribuições Binomial, 1-MB, 2-MB e $1,2$-IB, considerando $m = 13$.

Tabela 26 – Distribuição de frequência e estatísticas descritivas do número de ocorrências da vogal “A” em palavras terminadas em "r"com treze letras.

Número de vogais A (z_i)	0	1	2	3	4	5	Total	$\bar{z}$	DP
Frequência (f_i)	35	213	228	88	12	2	578	1,715	0,893

Fonte: Elaborada pelo autor.

A Tabela 27 apresenta as estimativas dos parâmetros μ, θ_1 e θ_2 obtidas pelo procedimento de máxima verossimilhança e os respectivos intervalos *bootstrap* com 95% de confiança. Além disso, esta Tabela apresenta para cada distribuição ajustada o número esperado para cada observação, o valor da estatística e o valor crítico do teste Kolmogorov-Smirnov (KS_{calc} e KS_{crit}, respectivamente) e ainda, os valores das medidas de evidências DE, KL e KLS.

Analisando os resultados desta Tabela, observamos que apenas para a distribuição $1,2$-IB a estatística de teste KS_{calc} é inferior ao valor crítico KS_{crit} e que as menores medidas de evidências são dadas pela mesma, indicando a não rejeição da hipótese de adequabilidade da distribuição $1,2$-IB, a um nível de 5% de significância. Este fato pode ser facilmente observado ao compararmos as frequências observadas com as frequências esperadas obtidas com cada distribuição ajustada.

Tabela 27 – Estimativas e intervalos de confiança dos parâmetros das distribuições Binomial, 1-MB, 2-MB e 1,2-IB ajustadas aos dados referentes ao número de vogais A nas palavras terminadas em "r"com treze letras, juntamente com os resultados de comparação das distribuições ajustadas.

z_i	O_i	B	1-MB	2-MB	1,2-IB
		E_i			
0	35	92	76	82	45
1	213	182	213	157	213
2	228	165	152	228	228
3	88	92	89	75	59
4	12	35	35	27	24
≥ 5	2	12	13	9	9
Total	578	578	578	578	578
		Estimativas			
Parâmetros	μ	1,715 (1,642; 1,786)	1,788 (1,714; 1,868)	1,663 (1,574; 1,748)	1,853 (1,748; 1,966)
	θ_1	-	0,094 (0,038; 0,154)	0,155 (0,098; 0,212)	0,202 (0,146; 0,259)
	θ_2	-	-	-	0,228 (0,173; 0,284)
Teste KS	KS_{calc}	0,098	0,072	0,082	0,034
	KS_{crit}	0,057	0,057	0,057	0,057
Medidas de evidências	DE	0,162	0,156	0,132	**0,058**
	KL	0,092	0,084	0,064	**0,028**
	KLS	0,202	0,177	0,137	**0,052**

Fonte: Elaborada pelo autor.

Problema 7: Número de sintomas em pacientes que morreram por COVID-19

Para a análise a seguir, consideramos os dados referentes ao número de sintomas sentidos por indivíduos contaminados com o vírus do Covid-19 que vieram à morte por esta doença em Alagoas, no período entre Março e Julho de 2020. Estes foram coletados pelo governo do Estado de Alagoas e disponibilizados pelo Portal Brasileiro de Dados Abertos em Dados.gov (2020). Assim, para o conjunto de dados, consideramos os sintomas mais comuns e mais relatados pelos pacientes: febre, tosse, dor de cabeça, dor muscular, falta de ar e dificuldade respiratória.

Sendo assim, nosso interesse está em analisar o número de sintomas mais relatados e que

foram sentidos por pacientes alagoanos que vieram à óbito por este vírus. A Tabela 28 apresenta a distribuição de frequência deste conjunto de dados, bem como a média e o desvio-padrão amostral.

Tabela 28 – Distribuição de frequência e estatísticas descritivas do número de sintomas de Covid-19 sentidos por pacientes que vieram à óbito em Alagoas, no período entre Março e Julho de 2020.

Número de sintomas (z_i)	0	1	2	3	4	5	6	Total	$\bar{z}$	DP
Frequência (f_i)	462	279	410	116	22	1	1	1.291	1,198	1,087

Fonte: Elaborada pelo autor.

Observando as frequências apresentadas na Tabela 28, podemos notar altos valores para as observações zero e dois, quando comparadas com as demais. Dessa forma, ajustamos as distribuições Binomial, 0-MB, 2-MB e $0,2$-IB a este conjunto de dados, considerando $m = 6$. Ressaltamos que 462 indivíduos vieram à óbito sem apresentar os sintomas listados, indicando que esses indivíduos vieram à óbito apresentando outros sintomas como diarreia, dor de garganta, fadiga, perda de paladar ou olfato, pressão no peito, dentre outros. Possíveis comorbidades pode ter agravado o estado de saúde do indivíduo levando ao óbito mas, infelizmente, não tivemos acesso a esse tipo de informação.

A Tabela 29 apresenta as estimativas dos parâmetros μ, θ_1 e θ_2 obtidas pelo procedimento de máxima verossimilhança e os respectivos intervalos *bootstrap* com 95% de confiança. Esta Tabela também apresenta para cada distribuição ajustada o número esperado para cada observação, o valor da estatística e o valor crítico do teste Kolmogorov-Smirnov (KS_{calc} e KS_{crit}, respectivamente) e ainda, os valores das medidas de evidências DE, KL e KLS.

Com base nos resultados apresentados na Tabela 29, observamos que apenas para a distribuição $0,2$-IB a estatística de teste KS_{calc} e as medidas DE, KL e KLS apresentaram os menores valores quando comparadas com as demais distribuições ajustadas. Dessa maneira, a hipótese de adequabilidade da distribuição $0,2$-IB não é rejeitada, a um nível de 5% de significância. Ao compararmos as frequências observadas com as frequências esperadas, também podemos observar este fato facilmente.

Tabela 29 – Estimativas e intervalos de confiança dos parâmetros das distribuições Binomial, 0-MB, 2-MB e $0,2$-IB ajustadas aos dados referentes ao número de sintomas de Covid-19 sentidos por pacientes que vieram à óbito em Alagoas, juntamente com os resultados de comparação das distribuições ajustadas.

z_i	O_i	B	0-MB	2-MB	$0,2$-IB
		E_i			
0	462	339	462	342	462
1	279	508	344	453	306
2	410	317	302	410	410
3	116	105	141	73	90
4	22	20	37	12	20
5	1	2	5	1	3
6	1	0	0	0	0
Total	1.291	1.291	1.291	1.291	1.291
		Estimativas			
Parâmetros	μ	1,198 (1,138; 1,259)	1,557 (1,485; 1,632)	1,084 (1,000; 1,164)	1,375 (1,319; 1,435)
	θ_1	-	0,231 (0,197; 0,266)	0,124 (0,086; 0,164)	0,225 (0,190; 0,259)
	θ_2	-	-	-	0,141 (0,107; 0,176)
Teste KS	KS_{calc}	0,095	0,050	0,093	0,021
	KS_{crit}	0,038	0,038	0,038	0,038
Medidas de evidências	DE	0,214	0,101	0,167	**0,029**
	KL	0,075	0,026	0,057	**0,005**
	KLS	0,157	0,052	0,115	**0,010**

Fonte: Elaborada pelo autor.

CAPÍTULO

7

CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Existem situações práticas em que duas observações quaisquer podem ter uma frequência observada significativamente mais alta do que a esperada, ao considerarmos uma distribuição discreta, tornando-a inadequada. Buscando contornar este problema, neste trabalho propomos a distribuição Série de Potência k_1 e k_2 Inflacionada ($\boldsymbol{k}$-IPS), que é capaz de modelar conjuntos de dados que apresentam ou não uma alta frequência das observações k_1 e k_2, simultaneamente.

Para a estimação dos parâmetros de interesse da distribuição $\boldsymbol{k}$-IPS, consideramos uma abordagem clássica, via método da máxima verossimilhança. Para inferir sobre estes parâmetros, consideramos os intervalos de confiança *bootstrap*, utilizando o procedimento não-paramétrico.

Realizamos um estudo de simulação com o intuito de verificar o desempenho do método clássico e avaliar as propriedades dos estimadores obtidos. De fato, os resultados obtidos foram satisfatórios, indicando a eficiência do procedimento. Além disso, as aplicações envolvendo dados artificiais também nos mostraram que a distribuição proposta foi a mais adequada para explicar o comportamento destes dados, quando comparadas aos ajustes das demais distribuições correspondentes.

A metodologia proposta neste trabalho foi aplicado em conjuntos de dados reais. Consideramos e analisamos sete conjuntos, os quais ajustamos as distribuições $\boldsymbol{k}$-IPS, além das distribuições PS e k_i-MPS, com $i = 1,2$, associadas para fins de comparações. As distribuições PS associadas consideradas nas análises dos dados foram: Poisson, Geométrica e Binomial. Os resultados obtidos foram bastante satisfatórios, uma vez que foi possível estabelecermos o tipo de modificação existente em cada conjunto de dados. A adequação das distribuições ajustadas foi observada quando comparamos as frequências observadas com as esperadas e confirmada pelo teste de Aderência Kolmogorov-Sminorv, juntamente com as medidas de evidência Distância Euclidiana, Divergência de Kullback-Leibler e Divergência de Kullback-Leibler Simétrica.

Em suma, é recomendada a utilização das distribuições $\boldsymbol{k}$-IPS como alternativas às distribuições discretas tradicionais, uma vez que muitos conjuntos de dados reais podem apresentar uma discrepância nas frequências, tornando inadequadas a suposição destas distribuições tradicionais. Além disso, recomendamos a utilização do método da máxima verossimilhança para estimação dos parâmetros, devido aos resultados satisfatórios nas aplicações com dados reais.

Como trabalhos futuros, podemos estender a metodologia proposta para outras distribuições da família PS, além de nos aprofundar na ideia da distribuição para o contexto de Modelos de regressão, além de considerar uma abordagem bayesiana para a estimação dos parâmetros das distribuições e dos modelos.

REFERÊNCIAS

ALSHKAKI, R. S. A. On the zero-one inflated poisson distribution. **International Jounal of Statistical Distribution and Applications**, v. 2, n. 4, p. 42–48, 2016. Citado nas páginas 26 e 39.

CALCBANK. **Como calcular a variação cambial de operações de crédito.** 2019. Disponível em: <https://www.calcbank.com.br>. Acesso em: 19/11/2019. Citado na página 77.

CARR-HILL, R.; MACDONALD, K. Problems in the analysis of life histories. **Stochastic Processes in Sociology**, p. 57–95, 1973. Citado nas páginas 75 e 76.

CARVALHO, S. O. **Distribuições $k-$Modificadas da Família Série de Potência Uniparamétrica**. Monografia (Dissertação de Mestrado) — PIPGES – Programa Interinstitucional de Pós Graduação em Estatística, Universidade de São Paulo e Universidade Federal de São Carlos, São Carlos - SP, 2017. Citado nas páginas 26, 30, 32, 33, 34 e 45.

CONCEIÇÃO, K. S.; ANDRADE, M. G.; LOUZADA, F.; HELOU, E. S. Zero-modified power series distribution and its hurdle distribution version. **Journal of Statistical Computation and Simulation**, v. 9, p. 1842–1862, 2017. Citado na página 25.

CONOVER, W. J. **Practical Nonparametric Statistics**. [S.l.]: Wiley Series in Probability and Mathematical Statistics, 1999. Citado na página 71.

CORDEIRO, G. M.; ANDRADE, M. G.; CASTRO, M. de. Power series generalized nonlinear models. **Computational Statistics & Data Analysis**, v. 53, p. 1155–1166, 2009. Citado na página 25.

DADOS.GOV. **Painel Covid-19 em Alagoas.** 2020. Disponível em: <https://dados.gov.br/>. Acesso em: 12/07/2020. Citado na página 85.

DALRYMPLE, M. L.; HUDSON, I. L.; FORD, R. P. K. Finite mixture, zero-inflated poisson and hurdle models with application te sids. **Computational Statistics Data Analysis**, v. 41, p. 491–504, 2003. Citado na página 37.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algotithm. **Journal of the Royal Statistical Society**, v. 39, n. 1, p. 1–38, 1977. Citado na página 56.

DICIO. **Palavras terminadas em R, com 13 letras.** 2019. Disponível em: <https://www.dicio.com.br/>. Acesso em: 03/12/2019. Citado na página 83.

GUPTA, P. L. Probability generating functions of a mpsd with applications. **Mathematics Operationsforsch. Statistics**, v. 13, n. 1, p. 99–103, 1982. Citado na página 31.

GUPTA, P. L.; GUPTA, R. C.; TRIPATHI, R. C. Inflated modified power series distributions with applications. **Communications in Statistics-Theory and Methods**, v. 24, n. 9, p. 2355–2374, 1995. Citado na página 26.

GUPTA, R. C. Modified power series distribution and some of its applications. **The Indian Journal of Statistics**, v. 36, n. 3, p. 288–298, 1974. Citado nas páginas 25, 30 e 31.

INMET. **Banco de Dados Meteorológicos para Ensino e Pesquisa.** 2019. Disponível em: <http://www.inmet.gov.br>. Acesso em: 13/12/2019. Citado na página 81.

INVESTING. **EUR/BRL - Euro Real Brasileiro.** 2019. Disponível em: <https://br.investing.com>. Acesso em: 19/11/2019. Citado na página 77.

JANI, P. N. On modified power series distribution. **Metron**, v. 36, p. 173–185, 1978. Citado na página 25.

KHATRI, C. G. On certain properties of power series distributions. **Biometrika**, v. 46, n. 3/4, p. 486–490, 1959. Citado na página 25.

KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. **Ann. Math. Statistics**, v. 22, p. 79–86, 1951. Citado na página 58.

MELKERSSON, M.; OLSSON, C. **Is visiting the dentist a good habit? Analyzing count data with excess zeros and access ones.** 1999. Disponível em: <http://www.econ.umu.se>. Acesso em: 10/07/2019. Citado nas páginas 26 e 39.

MILLER, L. Table of percentage points of kolmogorov statistic. **Journal of the American Statistical Association**, v. 51, p. 111–121, 1956. Citado na página 71.

MORGAN, B. J. T.; PALMER, K. J.; RIDOUT, M. S. Negative score test statistics. **American Statistical**, v. 61, p. 285–288, 2007. Citado na página 72.

MURAT, M.; SZYNAL, D. Non–zero inflated modified power series distributions. **Communications in Statistics-Theory and Methods**, v. 27, n. 12, p. 3047–3064, 1998. Citado nas páginas 26 e 34.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <https://www.R-project.org/>. Citado na página 27.

SAITO, M. Y.; RODRIGUES, J. Análise bayesiana de dados de contagem com excesso de zeros e uns. **Revista de Matemática e Estatística**, v. 23, n. 1, p. 47–57, 2005. Citado nas páginas 26 e 39.

SHARMA, A. K.; LANDGE, V. S. Zero inflated negative binomial for modeling heavy vehicle crash rate on indian rural highway. **International Journal of Advanced Engineering and Technology**, v. 5, n. 1, p. 292–301, 2013. Citado nas páginas 73 e 74.


ICMC
USP
SÃO CARLOS