
Distribuições k -Modificadas da Família Série de
Potência Uniparamétrica

Sérgio Ozório de Carvalho

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP

SÉRGIO OZÓRIO DE CARVALHO

**DISTRIBUIÇÕES k -MODIFICADAS DA FAMÍLIA SÉRIE DE
POTÊNCIA UNIPARAMÉTRICA**

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

Orientadora: Profa. Dra. Katiane Silva Conceição

São Carlos
Março de 2017

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP

SÉRGIO OZÓRIO DE CARVALHO

**K-MODIFIED DISTRIBUTIONS OF THE FAMILY
UNI-PARAMETRIC POWER SERIES**

Master dissertation submitted to the Departamento de Estatística – DEs-UFSCar and to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master joint Graduate Program in Statistics.

Advisor: Profa. Dra. Katiane Silva Conceição

**São Carlos
March 2017**



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de dissertação de mestrado do candidato Sergio Ozorio de Carvalho realizada em 23/05/2017:

Katiane S. Conceição

Profa. Dra. Katiane Silva Conceição
USP

Marcio Luis Lanfredi Viola

Prof. Dr. Marcio Luis Lanfredi Viola
UFSCar

Profa. Dra. Mariana Rodrigues Motta
UNICAMP

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Profa. Dra. Mariana Rodrigues Motta e, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do(a) aluno(a) Sergio Ozorio de Carvalho.

Katiane S. Conceição

Profa. Dra. Katiane Silva Conceição
Presidente da Comissão Examinadora
USP

RESUMO

CARVALHO, S. O. **Distribuições k -Modificadas da Família Série de Potência Uniparamétrica**. 2017. 98 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, 2017.

Neste trabalho é proposta a família de distribuições Série de Potência k -Modificadas para modelar conjuntos de dados de contagem que apresentam ou não alguma discrepância na frequência da observação k em relação à distribuição Série de Potência associada. É importante ressaltar que o emprego do termo Modificada(s) não possui o mesmo contexto ao empregado por Gupta (1974), o qual introduziu a classe de distribuições Série de Potência Modificadas representada pela sigla MPSD. Neste trabalho, entende-se por modificação, a inclusão de um parâmetro na função massa de probabilidade da distribuição Série de Potência tornando essa nova família de distribuições capaz de modelar adequadamente conjunto de dados para os casos em que há excesso (inflação), falta (deflação), ausência ou até mesmo quando a frequência da observação k estiver de acordo para a suposição de uma distribuição Série de Potência. Para esta nova família de distribuições são apresentadas propriedades como Função de distribuição, Função característica, Função geradora de momentos, Estatística de Ordem dentre outras, além de contextualiza-la como modelo de mistura. As distribuições consideradas para a construção dessa nova família serão as distribuições uniparamétricas pertencentes à família Série de Potência, cuja função massa de probabilidade pode ser escrita em função de sua média.

Palavras-chave: Dados de Contagem; Dados k -Inflacionados; Dados k -Deflacionados; Distribuições Série de Potência; Distribuições Série de Potência k -Modificadas.

ABSTRACT

CARVALHO, S. O. **K-Modified Distributions of the Family Uni-parametric Power Series.** 2017. 98 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, 2017.

This paper proposes a family of distributions Power Series k -Modified from to model sets of count data which have or not any discrepancy in the frequency of observation k in relation to the distribution associated Power Series. Is understood as a modification, inclusion of a parameter in the mass function of probability of the distribution Power Series, making this new family of distributions able to adequately model the data set in cases where there is excess (inflation), poor (deflation) the absence or even when the frequency of the observations k is according to the distribution power series. For this new family of distributions are presented some properties as the distribution functions, Statistics Order among others, besides contextualizes it as mixing model and place it in the context of regression models.

The distributions considered for the construction of this new family will be uni-parametric distributions belonging to the Power Series family, whose probability mass function can be written in terms of their average.

Keywords: Count Data; k -inflated data; k -Deflated data; Distributions Power Series; Distributions Power Series k -Modified.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama dos casos particulares da distribuição k -MPS.	30
Figura 2 – Valores da variância no ponto de modificação $k=0$	42
Figura 3 – Valores da variância no ponto de modificação $k=2$	42
Figura 4 – Valores da variância no ponto de modificação $k=5$	43
Figura 5 – Valores da variância no ponto de modificação $k=7$	43
Figura 6 – Gráficos das distribuições de frequências para as distribuições k -MP, k -MG e k -MB nos pontos de modificação $k = 0$ e $k = 5$, parâmetro $\mu = 0,5$ e parâmetro de modificação $p = 0,7$	45
Figura 7 – Gráficos das distribuições de frequências para as distribuições k -MP, k -MG e k -MB nos pontos de modificação $k = 0$ e $k = 5$, parâmetro $\mu = 2$ e parâmetro de modificação $p = 0,7$	46
Figura 8 – Gráficos das distribuições de frequências para as distribuições k -MP, k -MG e k -MB nos pontos de modificação $k = 0$ e $k = 5$, parâmetro $\mu = 5$ e parâmetro de modificação $p = 0,7$	46
Figura 9 – Mapa do Estado da Bahia. (A) Cidades que apresentaram notificações em 2004. (B) Cidades do grupo de risco em 2004.	72
Figura 10 – Diagrama da obtenção da sub-amostra.	73
Figura 11 – Comparação entre a frequência observada e a esperada segundo a distribuição 0-IP considerando o conjunto de dados das notificações de leptospirose com e sem a observação 415.	74
Figura 12 – Comparação gráfica entre a frequência esperada segundo a distribuição de Poisson, frequência das observações e a frequência esperada segundo a distribuição k -MP ajustada aos dados da sub-amostra considerando os pontos de modificação $k = 0$ e $k = 1$	76
Figura 13 – Gráfico comparativo entre as frequências das distribuições Poisson e k -MP, considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$	78
Figura 14 – Gráfico da Variação Global de Temperatura.	79
Figura 15 – Gráfico comparativo entre as frequências da distribuição Geométrica, observadas e a distribuição 0-IG.	81
Figura 16 – Gráfico comparativo das frequências esperadas das distribuições Binomial tradicional e k -MB com a frequência das observações.	83

Figura 17 – Gráfico comparativo entre as frequências das distribuições Binomial tradicional, dos dados observados e da distribuição k -MB considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$ 86

LISTA DE TABELAS

Tabela 1 – Algumas distribuições da família PS uniparamétrica.	23
Tabela 2 – Variância das distribuições PS e k -MPS.	41
Tabela 3 – Valores dos parâmetros das distribuições k -MPS utilizados no estudo de simulação.	58
Tabela 4 – Probabilidade de cobertura das estimativas dos parâmetros μ e p da distribuição k -MP.	59
Tabela 5 – Média das estimativas dos parâmetros μ e p e dos respectivos intervalos com 95% de confiança da distribuição k -MP.	60
Tabela 6 – Resultados do desempenho dos estimadores $\hat{\mu}$ e \hat{p} para distribuição k -MP.	61
Tabela 7 – Probabilidade de cobertura das estimativas dos parâmetros μ e p da distribuição k -MG.	62
Tabela 8 – Média das estimativas dos parâmetros μ e p e dos respectivos intervalos com 95% de confiança da distribuição k -MG.	63
Tabela 9 – Resultados do desempenho dos estimadores $\hat{\mu}$ e \hat{p} para distribuição k -MG.	64
Tabela 10 – Probabilidade de cobertura das estimativas dos parâmetros μ e p da distribuição k -MB.	65
Tabela 11 – Média das estimativas dos parâmetros μ e p e dos respectivos intervalos com 95% de confiança da distribuição k -MB.	66
Tabela 12 – Resultados do desempenho dos estimadores $\hat{\mu}$ e \hat{p} para distribuição k -MB.	67
Tabela 13 – Teste sobre a seleção de modelos para a distribuição k -MP.	68
Tabela 14 – Teste Sobre a Seleção de Modelos para a distribuição k -MG	68
Tabela 15 – Teste Sobre a Seleção de Modelos para a distribuição k -MB	69
Tabela 16 – Distribuição de frequência de notificações de leptospirose das cidades do Estado da Bahia em 2004.	72
Tabela 17 – Distribuição de frequência de notificações de leptospirose das cidades em risco do Estado da Bahia em 2004.	73
Tabela 18 – Resultados das estimativas dos parâmetros da distribuição 0-MP.	73
Tabela 19 – Teste de aderência Komolgorov Smirnov do ajuste da distribuição 0-IP aos dados de notificações de leptospirose com e sem a observação 415.	74
Tabela 20 – Resultados do ajuste da distribuições k -MP para a sub-amostra.	75
Tabela 21 – Teste de aderência Komolgorov Smirnov para a distribuição k -MP, considerando pontos de modificação $k = 0$ e $k = 1$	76

Tabela 22 – Número de gols marcados pelo Barcelona em todos os confrontos com o Real Madrid no período de 1955 a 2015.	77
Tabela 23 – Resultados das análises para os dados do Número de Gols de Partidas de Futebol considerando a distribuição k -MP nos pontos de modificação $k = 0, k = 1, k = 2$	77
Tabela 24 – Teste de aderência Kolmogorov-Smirnov para a distribuição k -MP, considerando pontos de modificação $k = 0, k = 1$ e $k = 2$	79
Tabela 25 – Distribuição de frequência do número de anos consecutivos com variação negativa da temperatura média anual até a ocorrência de uma variação positiva no período entre 1958 e 2008.	80
Tabela 26 – Estimativas de máxima verossimilhança dos parâmetros.	80
Tabela 27 – Teste de aderência Komolgorov Smirnov para a suposição da distribuição 0-IG.	81
Tabela 28 – Distribuição de frequência do número de acertos por aluno nas traduções das palavras.	82
Tabela 29 – Resultados da análise para a distribuição k -MB.	82
Tabela 30 – Teste Komolgorov-Smirnov de aderência para a distribuição k -MB para os pontos de modificação $k = 0, k = 1$ e $k = 2$	84
Tabela 31 – Distribuição de frequência da ocorrência combinadas dos artigos "the", "a" e "an".	84
Tabela 32 – Estimativas para o conjunto de dados que Caracteriza o Estilo de um Autor.	85
Tabela 33 – Teste de aderência Komolgorov Smirnov para a distribuição k -MB.	86

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Objetivo	19
2	CONCEITOS E NOTAÇÕES PRELIMINARES	21
2.1	Distribuições Séries de Potência Uniparamétricas	21
2.2	Distribuições Série de Potência k -Subtraídas	23
3	DISTRIBUIÇÕES SÉRIE DE POTÊNCIA k -MODIFICADAS	27
3.1	Casos Particulares da distribuição k -MPS(μ, p)	29
3.2	A Família de Distribuições k -MPS(μ, p)	31
3.2.1	<i>Distribuição Poisson k-Modificada (k-MP)</i>	31
3.2.2	<i>Distribuição Geométrica k-Modificada (k-MG)</i>	32
3.2.3	<i>Distribuição Binomial k-Modificada (k-MB)</i>	33
3.2.4	<i>Distribuição Borel k-Modificada (k-MBo)</i>	34
3.2.5	<i>Distribuição Borel-Tanner k-Modificada (k-MBT)</i>	35
3.2.6	<i>Distribuição Haight k-Modificada (k-MH)</i>	36
3.3	Caracterização da distribuição k -MPS	37
3.3.1	<i>Função de Distribuição</i>	37
3.3.2	<i>Estatística de Ordem</i>	37
3.3.3	<i>Função Geradora de Probabilidade</i>	39
3.3.4	<i>Função Característica</i>	39
3.3.5	<i>Função Geradora de Momentos</i>	40
3.3.6	<i>Gerador de Números Pseudo-Aleatórios</i>	44
3.4	Distribuições k -MPS e sua Versão <i>Hurdle</i>	47
4	MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS E SELEÇÃO DE MODELOS	49
4.1	Método dos Momentos	49
4.2	Método de Máxima Verossimilhança	50
4.2.1	<i>Função de Verossimilhança para k-MPS(μ, p)</i>	50
4.2.2	<i>Função de Verossimilhança para k-MPS($\mu; \omega$)</i>	53
4.2.3	<i>Distribuição k-MPS versus Distribuição k-SPS</i>	54
4.3	Critérios de Seleção de Modelos	56

5	ESTUDO DE SIMULAÇÃO	57
5.1	Distribuição Poisson k -Modificada	59
5.2	Distribuição Geométrica k -Modificada	61
5.3	Distribuição Binomial k -Modificada	64
5.4	Trabalhando com Dados Artificiais	67
6	APLICAÇÕES COM DADOS REAIS	71
6.1	Notificações de Leptospirose no Estado da Bahia	71
6.2	Número de Gols em Partidas de Futebol	77
6.3	Variação Global da Temperatura	79
6.4	Significado de Palavras Técnicas de Estatística	81
6.5	Estudo que Caracteriza o Estilo de um Autor.	84
7	CONCLUSÃO	89
7.1	Trabalhos Futuros	90
8	APÊNDICE	91
.1	Derivada de Ordem Primeira da PS	91
.2	Derivada de Segunda Ordem da PS	92
.3	Código	93
	REFERÊNCIAS	97

INTRODUÇÃO

Em estatística, a análise de dados de contagem é objeto de estudo de muitos pesquisadores, que buscam através de formulações matemáticas encontrar modelos estatísticos capazes de proporcionar aos dados uma representação adequada de sua ocorrência. Alguns pesquisadores consideravam que para obter um bom ajuste para dados de contagem bastava considerar a distribuição de Poisson tradicional, como pode ser visto, por exemplo, em [Frome e Checkoway \(1985\)](#) e [Bayarri *et al.* \(2008\)](#). No entanto, em muitos casos verificou-se as seguintes situações:

- i) sobredispersão dos dados, quando a variância da amostra é maior do que a média, ou subdispersão dos dados, quando a variância é menor que a média;
- ii) em conjunto de dados reais a frequência de zeros era maior (ou menor) do que a frequência esperada em relação a distribuição Poisson tradicional.

Em ambos os casos não é adequado considerar a distribuição de Poisson tradicional como um bom ajuste aos dados. Este fato levou pesquisadores como [Patil \(1962\)](#), [Gupta \(1974\)](#), [Jani \(1978\)](#), [Cordeiro, Andrade e Castro \(2009\)](#) entre outros a considerarem a família de distribuições Série de Potência como uma alternativa para modelar dados com essas e outras características. Há, atualmente, na literatura muitas propostas de soluções para modelar conjuntos de dados de contagem com variância amostral diferente da média e com excesso ou falta da observação zero. Tais soluções geralmente são apresentadas no contexto de modelos de mistura.

Em [Conceição \(2013\)](#) é apresentada uma solução interessante que vai além dos modelos de mistura. Utilizando as distribuições Série de Potência Zero-Modificadas, a autora consegue modelar adequadamente a discrepância na frequência de zero, independentemente de se ter conhecimento prévio sobre o tipo de modificação (inflação ou deflação de zero no conjunto de dados).

Também é possível observar distribuições na qual a inflação ocorre em algum ponto k da reta suporte diferente de zero. Por exemplo, em [Rodrigues \(2011\)](#) é descrito um processo no qual há duas filas, sendo uma delas infinita e a outra fila começando com apenas um cliente. Observa-se, então, que o número de clientes atendidos em um período de pico dessas duas filas, sem saber qual deles é servido, é inflacionado em um ponto k diferente de zero. Neste contexto, se considerarmos uma população mista constituída por dois grupos de indivíduos, sendo que no primeiro grupo, a variável de contagem associada aos indivíduos segue uma distribuição da família Série de Potência; e no segundo grupo, a mesma variável associada aos indivíduos assume o valor k com uma frequência significativamente mais alta do que a esperada com base na distribuição Série de Potência. Neste caso teremos, então, que o conjunto de dados formado pelas observações extraídas da população formada por esses dois grupos de indivíduos, sem distinção a qual grupo eles pertencem, é inflacionada no ponto k . Como exemplo de situações reais, podemos citar o experimento descrito por [Pandey \(1965\)](#) no qual se utiliza à distribuição de Poisson Inflacionada para modelar o número de flores produzida por plantas *Primula veris*. Ele mostrou que devido ao número excessivo de plantas com 8 flores, este conjunto de observações resultou em uma distribuição de Poisson inflacionada no ponto $k = 8$. [Murat e Szynal \(1998\)](#) estenderam os resultados de [Gupta, Gupta e Tripathi \(1995\)](#) para distribuições discretas inflacionadas em qualquer ponto k .

No artigo publicado por [Saito e Rodrigues \(2005\)](#), os autores fazem uso de uma mistura de modelos, distribuição ZOIP (*Zeros and Ones Inflated Poisson*), e o aplicam para modelar o número de visitas ao dentista de cidadãos suecos, observando que além do conjunto de dados ter excesso de zero (zero inflacionado), havia também o excesso de um (um inflacionado).

Conjuntos de dados deflacionados de algum valor k são obtidos a partir de experimentos em que este valor é observado com uma frequência significativamente mais baixa do que a esperada com base em uma distribuição Série de Potência. Em [Conceição \(2013\)](#) é usado como exemplo um conjunto de dados de notificações de leptospirose das cidades de risco com IDH menor que 0,66 do estado da Bahia em 2004, o qual após o ajuste considerando as distribuições de Poisson Zero-Modificada e Binomial Negativa Zero-Modificada, mostrou-se ser deflacionado no ponto $k = 0$. Outro exemplo é o processo de contagem do número de filhos em famílias carentes, em que algumas destas famílias participam de algum programa social no qual recebem auxílio pelo número de filhos. Este benefício pode induzir um processo no qual as famílias carentes são incentivadas a terem mais de um filho, causando no processo original uma deflação na observação $k = 1$ (um único filho). Também podemos citar como exemplo de dados deflacionados, o processo de contagem da quantidade de um determinado produto comprado por um cliente em uma rede de supermercados quando é feita uma oferta que propõe a diminuição do preço na unidade deste produto caso o cliente opte por levar uma quantidade acima de k unidades ($k > 1$), também pode resultar na deflação da observação no ponto k de itens vendidos. Além destes, muitos outros exemplos podem ser encontrados na prática, justificando a proposta das distribuições discretas com inflação ou deflação em uma observação k .

A princípio qualquer conjunto de dados de contagem com distribuição de frequência bimodal, que podem ser modelados com uma mistura de distribuições Série de Potência, podem ser modelados por uma distribuição Série de Potência modificada no ponto k , $k \neq 0$, identificado previamente. A vantagem em utilizar esta distribuição está no fato de que o número de parâmetros a ser estimado é menor.

1.1 Objetivo

O principal objetivo deste trabalho é estender a ideia da modificação na probabilidade de zero das distribuições Série de Potência Zero-Modificadas (ZMPS) proposta por [Conceição \(2013\)](#), para qualquer observação k , tornando a distribuição ZMPS um caso particular desta nova família de distribuições, denominadas por distribuições Série de Potência k -Modificadas (k -MPS¹). Aqui iremos considerar as distribuições Série de Potência uniparamétricas cuja função massa de probabilidades pode ser escrita em função de sua média para o contexto da k -modificação. Com isso, iremos mostrar que ao flexibilizar o espaço paramétrico do parâmetro de modificação, proporcionamos que as distribuições k -modificadas sejam ajustadas adequadamente à qualquer conjunto de dados de contagem.

Neste trabalho será feita a modificação em um ponto k das distribuições uniparamétricas pertencentes à família Série de Potência: Poisson, Geométrica, Binomial, Borel, Borel-Tanner e Haight.

A pesquisa foi desenvolvida utilizando dos recursos computacionais do Centro de Ciências Matemáticas Aplicadas à Indústria (CeMEAI) financiados pela FAPESP e todos os programas foram implementados no *software* [R Core Team \(2016\)](#).

¹ É utilizada a sigla k -MPS, do inglês *k-Modified Power Series*, referindo-se às distribuições Série de Potência k -Modificadas.

CONCEITOS E NOTAÇÕES PRELIMINARES

2.1 Distribuições Séries de Potência Uniparamétricas

Considere Y uma variável aleatória definida sobre os inteiros não negativos. Se Y tem distribuição Série de Potência (PS¹) uniparamétrica com média $\mu > 0$, então sua função massa de probabilidade é definida por:

$$\pi_{PS}(y; \mu) = \frac{a(y)g(\mu)^y}{f(\mu)}, y \in A_s, \quad (2.1)$$

em que:

- i) $A_s = \{s, s+1, s+2, \dots\}$ tal que $A_s \subset \mathbb{N}$;
- ii) $a(y)$ é uma função positiva;
- iii) $f(\mu)$ e $g(\mu)$ são funções positivas, finitas e duas vezes diferenciáveis; e
- iv) $f(\mu) = \sum_{y \in A_s} a(y)g(\mu)^y$.

Proposição 2.1. Seja Y uma variável aleatória com distribuição PS. Então, sua média μ e variância σ^2 são dadas, respectivamente, por:

$$\mu = E[Y] = \frac{f'(\mu)g(\mu)}{f(\mu)g'(\mu)} \quad \text{e} \quad \sigma^2 = \text{Var}[Y] = \frac{g(\mu)}{g'(\mu)},$$

em que $f'(\mu)$ e $g'(\mu)$ são, respectivamente, as derivadas das funções f e g em relação a μ .

¹ É utilizada a sigla PS, do inglês *Power Series*, referindo-se às distribuições Série de Potência.

Demonstração. Considere $f(\mu) = \sum_{y \in A_s} a(y)g(\mu)^y$. Diferenciando ambos os lados dessa equação em relação à μ , obtemos a seguinte relação de igualdade.

$$\frac{g'(\mu)}{g(\mu)} \sum_{y \in A_s} ya(y)g(\mu)^y = f'(\mu).$$

Multiplicando os dois lados dessa igualdade por $\frac{g(\mu)}{g'(\mu)}$ e dividindo por $f(\mu)$, obtemos que:

$$\sum_{y \in A_s} y \frac{a(y)g(\mu)^y}{f(\mu)} = \frac{f'(\mu)g(\mu)}{f(\mu)g'(\mu)},$$

em que o termo do lado esquerdo da equação é a esperança da variável aleatória Y cuja notação é dada por:

$$\mu = E[Y] = \frac{f'(\mu)g(\mu)}{f(\mu)g'(\mu)}.$$

Ao considerar o resultado para a média, podemos obter a variância usando a seguinte expressão.

$$\sum_{y \in A_s} ya(y)g(\mu)^y = f(\mu)\mu.$$

Diferenciando os dois lados da equação em relação à μ , obtemos a igualdade.

$$\frac{g'(\mu)}{g(\mu)} \sum_{y \in A_s} y^2 a(y)g(\mu)^y = f'(\mu)\mu + f(\mu).$$

Multiplicando a igualdade acima por $\frac{g(\mu)}{g'(\mu)}$ e dividindo por $f(\mu)$, obtemos que:

$$\sum_{y \in A_s} y^2 \frac{a(y)g(\mu)^y}{f(\mu)} = \frac{g(\mu)}{f(\mu)g'(\mu)} (f'(\mu)\mu + f(\mu)),$$

em que o termo do lado esquerdo da equação é a esperança do quadrado da variável aleatória Y , cuja notação é dada por:

$$E[Y^2] = \frac{f'(\mu)g(\mu)\mu}{f(\mu)g'(\mu)} + \frac{f(\mu)g(\mu)}{f(\mu)g'(\mu)},$$

a qual pode ser reescrita da seguinte forma.

$$E[Y^2] = \mu^2 + \frac{g(\mu)}{g'(\mu)}.$$

Uma vez que $\text{Var}[Y] = E[Y^2] - E^2[Y]$, temos que:

$$\sigma^2 = \text{Var}[Y] = \mu^2 + \frac{g(\mu)}{g'(\mu)} - \mu^2 = \frac{g(\mu)}{g'(\mu)}.$$

□

A Tabela ?? exhibe as funções $f(\mu)$, $g(\mu)$ e $a(y)$ das distribuições da família PS.

Tabela 1 – Algumas distribuições da família PS uniparamétrica.

PS	Distribuições	$f(\mu)$	$g(\mu)$	$a(y)$	A_s
P	Poisson	e^μ	μ	$\frac{1}{y!}$	$\{0, 1, 2, \dots\}$
G	Geométrica	$1 + \mu$	$\frac{\mu}{1 + \mu}$	1	$\{0, 1, 2, \dots\}$
B	Binomial	$\left(\frac{m}{m - \mu}\right)^m$	$\frac{\mu}{m - \mu}$	$\binom{m}{y}$	$\{0, 1, 2, \dots, m\}$
Bo	Borel	$1 - \frac{1}{\mu}$	$\left(1 - \frac{1}{\mu}\right) e^{-1 + \frac{1}{\mu}}$	$\frac{y^{y-2}}{(y-1)!}$	$\{1, 2, \dots\}$
BT	Borel-Tanner	$\left(1 - \frac{m}{\mu}\right)^m$	$\left(1 - \frac{m}{\mu}\right) e^{-1 + \frac{m}{\mu}}$	$\frac{m^{y-m-1}}{(y-m)!}$	$\{m, m + 1, \dots\}$
H	Haight	$\frac{\mu - 1}{2\mu - 1}$	$\frac{\mu(\mu - 1)}{(2\mu - 1)^2}$	$\frac{(2y - 2)!}{y!(y - 1)!}$	$\{1, 2, \dots\}$

Para maiores detalhes sobre as distribuições PS, ver artigo do [Gupta \(1974\)](#).

2.2 Distribuições Série de Potência k -Subtraídas

Teorema 1. Seja Y uma variável aleatória com distribuição Série de Potência k -Subtraída (k -SPS²). A sua função massa de probabilidade é dada por:

$$\begin{aligned} \pi_{k\text{-SPS}}(y; \mu) &= \frac{\pi_{PS}(y; \mu)}{1 - \pi_{PS}(k; \mu)} \\ &= \frac{a(y)g(\mu)^y}{f(\mu) - a(k)g(\mu)^k}, \quad y \in A_{\{-k\}}, \end{aligned} \quad (2.2)$$

em que $A_{\{-k\}}$ é o suporte formado pelo subconjunto dos inteiros, tal que $A_{\{-k\}} = A_s - \{k\}$.

Demonstração. Para provar que $\pi_{k\text{-SPS}}(y; \mu)$ é uma função massa de probabilidade, precisamos mostrar que as seguintes propriedades são verdadeiras:

² Usaremos a sigla k -SPS, do inglês *k-Subtracted Power Series*, referindo-se às distribuições Série de Potência k -Subtraídas.

$$\left\{ \begin{array}{l} i) \quad \pi_{k-SPS}(y; \mu) \geq 0, \forall y \in A_{\{-k\}} \\ ii) \quad \sum_{y \in A_{\{-k\}}} \pi_{k-SPS}(y; \mu) = 1 \end{array} \right. .$$

No caso (i), temos que:

$$\pi_{k-SPS}(y; \mu) = \frac{\overbrace{\pi_{PS}(y; \mu)}^{\geq 0}}{\underbrace{1 - \pi_{PS}(k; \mu)}_{\geq 0}} \geq 0, \forall y \in A_{\{-k\}}.$$

Já para o caso (ii), temos que:

$$\begin{aligned} \sum_{y \in A_{\{-k\}}} \pi_{k-SPS}(y; \mu) &= \sum_{y \in A_{\{-k\}}} \frac{a(y)g(\mu)^y}{f(\mu) - a(k)g(\mu)^k} \\ &= \frac{f(\mu)}{f(\mu) - a(k)g(\mu)^k} \sum_{y \in A_{\{-k\}}} \frac{a(y)g(\mu)^y}{f(\mu)} \\ &= \frac{f(\mu)}{f(\mu) \left(1 - \frac{a(k)g(\mu)^k}{f(\mu)}\right)} (1 - \pi_{PS}(k; \mu)) \\ &= \frac{1}{1 - \pi_{PS}(k; \mu)} (1 - \pi_{PS}(k; \mu)) = 1. \end{aligned}$$

□

Proposição 2.2. Seja Y uma variável aleatória tal que $Y \sim k\text{-SPS}(\mu)$. Então, a sua média μ_{k-SPS} e variância σ_{k-SPS}^2 são dadas, respectivamente, por:

$$\begin{aligned} \mu_{k-SPS} &= E[Y] = \frac{\mu - k\pi_{PS}(k; \mu)}{1 - \pi_{PS}(k; \mu)}, \\ \sigma_{k-SPS}^2 &= Var[Y] = \frac{\sigma^2(1 - \pi_{PS}(k; \mu)) - (\mu - k)^2\pi_{PS}(k; \mu)}{(1 - \pi_{PS}(k; \mu))^2}. \end{aligned}$$

Demonstração. Considere a igualdade $\sum_{y \in A_s} a(y)g(\mu)^y = f(\mu)$. Subtraindo os dois lados dessa igualdade por $a(k)g(\mu)^k$, obtemos.

$$\sum_{y \in A_s} a(y)g(\mu)^y - a(k)g(\mu)^k = f(\mu) - a(k)g(\mu)^k.$$

Diferenciando os dois lados dessa equação em relação a μ , obtemos.

$$\frac{g'(\mu)}{g(\mu)} \left(\sum_{y \in A_s} ya(y)g(\mu)^y - ka(k)g(\mu)^k \right) = f'(\mu) - ka(k)g(\mu)^k \frac{g'(\mu)}{g(\mu)}.$$

Multiplicando os dois termos da igualdade acima por $\frac{g(\mu)}{g'(\mu)}$, obtemos.

$$\sum_{y \in A_s} ya(y)g(\mu)^y - ka(k)g(\mu)^k = \frac{g(\mu)}{g'(\mu)} f'(\mu) - ka(k)g(\mu)^k,$$

e finalmente, dividimos os termos dos dois lados da equação por $f(\mu) - a(k)g(\mu)^k$.

$$\sum_{y \in A_s} y \frac{a(y)g(\mu)^y - ka(k)g(\mu)^k}{f(\mu) - a(k)g(\mu)^k} = \frac{\frac{g(\mu)}{g'(\mu)} f'(\mu) - ka(k)g(\mu)^k}{f(\mu) - a(k)g(\mu)^k}.$$

Com isso, o termo do lado esquerdo da equação é a esperança da variável aleatória Y , o qual será substituído por:

$$\begin{aligned} E[Y] &= \frac{g(\mu)f'(\mu)}{g'(\mu)f(\mu)(1 - \pi_{PS}(k; \mu))} - \frac{ka(k)g(\mu)^k}{f(\mu)(1 - \pi_{PS}(k; \mu))} \\ E[Y] &= \frac{\mu}{1 - \pi_{PS}(k; \mu)} - \frac{k\pi_{PS}(k; \mu)}{1 - \pi_{PS}(k; \mu)} \\ \mu_{k-SPS} = E[Y] &= \frac{\mu - k\pi_{PS}(k; \mu)}{1 - \pi_{PS}(k; \mu)}. \end{aligned}$$

Considerando o resultado da média μ_{k-SPS} podemos obter a variância σ_{k-SPS}^2 , usando a expressão.

$$\sum_{y \in A_s} ya(y)g(\mu)^y - ka(k)g(\mu)^k = \mu_{k-SPS} \left(f(\mu) - a(k)g(\mu)^k \right).$$

Substituindo $\mu_{k-SPS} = \frac{f(\mu)\mu - ka(k)g(\mu)^k}{f(\mu) - a(k)g(\mu)^k}$ na equação acima, temos:

$$\begin{aligned} \sum_{y \in A_s} ya(y)g(\mu)^y - ka(k)g(\mu)^k &= \frac{f(\mu)\mu - ka(k)g(\mu)^k}{f(\mu) - a(k)g(\mu)^k} \left(f(\mu) - a(k)g(\mu)^k \right) \\ &= f(\mu)\mu - ka(k)g(\mu)^k. \end{aligned}$$

Diferenciando os dois lados dessa equação em relação a μ e posteriormente dividindo por $f(\mu) - a(k)g(\mu)^k$, obtemos:

$$\begin{aligned}
E[Y^2] &= \frac{\sum_{y \in A_S} y^2 a(y) g(\mu)^y - k^2 a(k) g(\mu)^k}{f(\mu) - a(k) g(\mu)^k} \\
&= \frac{g(\mu)}{g'(\mu)(f(\mu) - a(k)g(\mu)^k)} \left\{ f'(\mu)\mu + f(\mu) - k^2 a(k)g(\mu)^k \frac{g'(\mu)}{g(\mu)} \right\} \\
&= \frac{f'(\mu)g(\mu)\mu}{g'(\mu)f(\mu)(1 - \pi_{PS}(k; \mu))} + \frac{f(\mu)g(\mu)}{g'(\mu)f(\mu)(1 - \pi_{PS}(k; \mu))} - \frac{k^2 a(k)g(\mu)^k}{f(\mu)(1 - \pi_{PS}(k; \mu))} \\
&= \frac{\mu^2}{(1 - \pi_{PS}(k; \mu))} + \frac{\sigma^2}{(1 - \pi_{PS}(k; \mu))} - \frac{k^2 \pi_{PS}(k; \mu)}{(1 - \pi_{PS}(k; \mu))}.
\end{aligned}$$

Como a variância é dada por $Var[Y] = E[Y^2] - E^2[Y]$, temos:

$$\begin{aligned}
Var[Y] &= \frac{\mu^2}{(1 - \pi_{PS}(k; \mu))} + \frac{\sigma^2}{(1 - \pi_{PS}(k; \mu))} - \frac{k^2 \pi_{PS}(k; \mu)}{(1 - \pi_{PS}(k; \mu))} - \left(\frac{\mu - k \pi_{PS}(k; \mu)}{1 - \pi_{PS}(k; \mu)} \right)^2 \\
&= \frac{\sigma^2(1 - \pi_{PS}(k; \mu)) - \mu^2 \pi_{PS}(k; \mu) - k^2 \pi_{PS}(k; \mu) + 2\mu k \pi_{PS}(k; \mu)}{(1 - \pi_{PS}(k; \mu))^2} \\
Var[Y] &= \sigma_{k-SPS}^2 = \frac{\sigma^2(1 - \pi_{PS}(k; \mu)) - (\mu - k)^2 \pi_{PS}(k; \mu)}{(1 - \pi_{PS}(k; \mu))^2}.
\end{aligned}$$

□

O termo k -Subtraída é empregado para representar a remoção ou ausência da observação k do suporte da variável aleatória Y , com isso, a probabilidade do evento $Y = k$ é zero. Este é um conceito diferente daquele empregado às distribuições truncadas, em que a probabilidade zero ocorre para um conjunto de observações menores ou iguais a k , ou seja, $P(Y \leq k) = 0$.

DISTRIBUIÇÕES SÉRIE DE POTÊNCIA k -MODIFICADAS

Neste capítulo, mostraremos que a partir das distribuições PS, é possível construir distribuições discretas com modificação na probabilidade de uma observação k de sua reta suporte.

Teorema 2. Seja a função $\pi_{k-MPS}(y; \mu, p)$ dada por:

$$\pi_{k-MPS}(y; \mu, p) = (1 - p)I_{\{k\}}(y) + p\pi_{PS}(y; \mu), y \in A_s, \quad (3.1)$$

em que A_s é o suporte formado pelo subconjunto dos inteiros $\{s, s + 1, \dots\}$; p é o parâmetro responsável pela modificação das probabilidades em relação as distribuições PS tradicionais, satisfazendo a restrição:

$$0 \leq p \leq \frac{1}{1 - \pi_{PS}(k; \mu)}; \quad (3.2)$$

e $I_{\{k\}}(y)$ é a função indicadora definida por:

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

A equação dada em (3.1) é uma função massa de probabilidade.

Prova. Se $\pi_{k-MPS}(y; \mu, p)$ é uma função massa de probabilidade, as seguintes propriedades devem ser verdadeiras:

$$\left\{ \begin{array}{l} \text{i) } \pi_{k-MPS}(y; \mu, p) \geq 0 \forall y \in A_s \\ \text{ii) } \sum_{y \in A_s} \pi_{k-MPS}(y; \mu, p) = 1 \end{array} \right. .$$

No caso (i), temos que:

Se $y = k$, temos que :

$$\pi_{k-MPS}(k; \mu, p) = 1 - \underbrace{p(1 - \pi_{PS}(k; \mu))}_{\geq 0 \text{ e } \leq 1} \geq 0 \implies \pi_{k-MPS}(k; \mu, p) \geq 0; \text{ para } y = k.$$

Se $y \neq k$, temos que :

$$\pi_{k-MPS}(y; \mu, p) = \underbrace{p\pi_{PS}(y; \mu)}_{\geq 0} \geq 0, \forall y \in A_s \text{ e } y \neq k.$$

No caso (ii), temos que:

$$\begin{aligned} \sum_{y \in A_s} \pi_{k-MPS}(y; \mu, p) &= \sum_{y \in A_s} \{(1-p)I_{\{k\}}(y) + p\pi_{PS}(y; \mu)\} \\ &= (1-p) + p \sum_{y \in A_s} \pi_{PS}(y; \mu) \\ &= 1 - p + p \\ &= 1. \end{aligned}$$

Logo, $\pi_{k-MPS}(y; \mu, p)$ é uma função massa de probabilidade.

□

Definição 1. Definimos a função $\pi_{k-MPS}(y; \mu, p)$ dada em (3.1) por distribuição Série de Potência k -Modificada (k -MPS¹).

Notação: k -MPS(μ, p).

Ressaltamos que o parâmetro p da distribuição k -MPS pode assumir valor maior que 1, este fato torna distribuição diferente da distribuição de mistura tradicional.

¹ Usaremos a sigla k -MPS, do inglês k -Modified Power Series, referindo-se às distribuições Série de Potência k -Modificadas.

3.1 Casos Particulares da distribuição k -MPS(μ, p)

Para conjunto de dados de contagem cuja variável aleatória tem distribuição k -MPS, diferentes proporções das observações k na amostra levam a diferentes valores do parâmetro de modificação p e, como consequência, a diferentes distribuições ao avaliarmos a diferença entre as probabilidades da observação k :

$$\begin{aligned}\pi_{k-MPS}(k; \mu, p) - \pi_{PS}(k; \mu) &= 1 - p + p\pi_{PS}(k; \mu) - \pi_{PS}(k; \mu) \\ &= (1 - p) - \pi_{PS}(k; \mu)(1 - p) \\ &= (1 - p)(1 - \pi_{PS}(k; \mu)).\end{aligned}\tag{3.3}$$

Uma vez que a restrição dada em (3.2) é satisfeita, temos os seguintes corolários:

Corolário 3.1. Se $p = 0$, então $\pi_{k-MPS}(y; \mu, p)$ é a distribuição degenerada em que toda sua massa está no ponto k .

Prova. Fazendo $p = 0$ em (3.3), temos que:

$$\pi_{k-MPS}(k; \mu, p) - \pi_{PS}(k; \mu) = 1 - \pi_{PS}(k; \mu) \implies \pi_{k-MPS}(k; \mu, p) = 1.$$

□

Corolário 3.2. Para todo $0 < p < 1$, então $\pi_{k-MPS}(y; \mu, p)$ é a distribuição Série de Potência k -Inflacionada (k -IPS²).

Prova. Para todo $0 < p < 1$ em (3.3), temos que:

$$\pi_{k-MPS}(k; \mu, p) - \pi_{PS}(k; \mu) = \underbrace{(1 - p)}_{>0} \underbrace{(1 - \pi_{PS}(k; \mu))}_{>0} > 0 \implies \pi_{k-MPS}(k; \mu, p) > \pi_{PS}(k; \mu).$$

□

Corolário 3.3. Se $p = 1$, então $\pi_{k-MPS}(y; \mu, p)$ é a distribuição Série de Potência usual.

Prova. Fazendo $p = 1$ em (3.3), temos que:

$$\pi_{k-MPS}(k; \mu, p) - \pi_{PS}(k; \mu) = 0 \implies \pi_{k-MPS}(k; \mu, p) = \pi_{PS}(k; \mu).$$

□

² Usaremos a sigla k -IPS, do inglês k -Inflated Power Series, referindo-se às distribuições Série de Potência k -Inflacionadas.

Corolário 3.4. Para todo $1 < p < \frac{1}{1-\pi_{PS}(k;\mu)}$, $\pi_{k-MPS}(y; \mu, p)$ é a distribuição Série de Potência k -Deflacionada (k -DPS³).

Prova. Para todo $1 < p < \frac{1}{1-\pi_{PS}(k;\mu)}$ em (3.3), temos que:

$$\pi_{k-MPS}(k; \mu, p) - \pi_{PS}(k; \mu) = \underbrace{(1-p)}_{<0} \underbrace{(1-\pi_{PS}(k; \mu))}_{>0} < 0 \implies \pi_{k-MPS}(k; \mu, p) < \pi_{PS}(k; \mu).$$

□

Corolário 3.5. Se $p = \frac{1}{1-\pi_{PS}(k;\mu)}$, então $\pi_{k-MPS}(y; \mu, p)$ é a distribuição Série de Potência k -Subtraída, a qual tem uma probabilidade nula de ocorrência da observação k .

Prova. Se $p = \frac{1}{1-\pi_{PS}(k;\mu)}$ em (3.3), temos que:

$$\begin{aligned} \pi_{k-MPS}(k; \mu, p) - \pi_{PS}(k; \mu) &= \left(1 - \frac{1}{1-\pi_{PS}(k;\mu)}\right)(1-\pi_{PS}(k;\mu)) \\ \pi_{k-MPS}(k; \mu, p) &= 1 - \frac{1}{1-\pi_{PS}(k;\mu)} + \frac{\pi_{PS}(k;\mu)}{1-\pi_{PS}(k;\mu)} \\ \pi_{k-MPS}(k; \mu, p) &= \frac{1-\pi_{PS}(k;\mu) - 1 + \pi_{PS}(k;\mu)}{1-\pi_{PS}(k;\mu)} = 0. \end{aligned}$$

□

Na Figura 1 temos um diagrama representando os casos particulares da distribuição k -MPS apresentados nos Corolários 3.1 - 3.5.

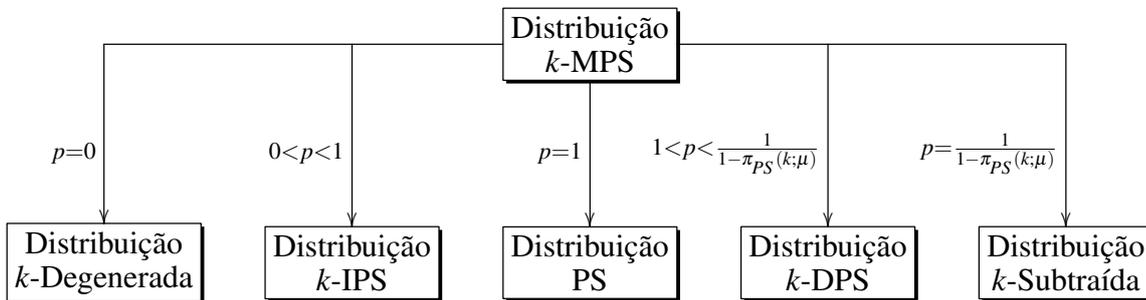


Figura 1 – Diagrama dos casos particulares da distribuição k -MPS.

Na Figura 1 verificamos que o parâmetro de modificação p ao assumir valores no intervalo $[0, \frac{1}{1-\pi_{PS}(k;\mu)}]$, vão se configurando os casos particulares da distribuição k -MPS. Este

³ Usaremos a sigla k -DPS, do inglês k -Deflated Power Series, referindo-se às distribuições Série de Potência k -Deflacionadas.

fato proporciona grande vantagem no uso da distribuição k -MPS, pois, uma vez realizado o ajuste do modelo, para determinada observação k , basta verificar a estimativa de p para então estabelecer em qual desses casos particulares pertence o conjunto de dados em análise e, assim, caracterizá-la.

3.2 A Família de Distribuições k -MPS(μ, p)

Nesta seção apresentaremos as distribuições que compõem a família k -MPS, além de descrever suas particularidades em relação as modificações.

3.2.1 Distribuição Poisson k -Modificada (k -MP)

Considere Y uma variável aleatória tendo distribuição de Poisson com parâmetro de média $\mu > 0$, cuja função massa de probabilidade é dada por:

$$\pi_p(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots \quad (3.4)$$

A distribuição k -modificada de Y , dita distribuição Poisson k -Modificada (k -MP⁴) é definida a partir de (3.4) e sua função massa de probabilidade é dada por:

$$\pi_{k-MP}(y; \mu, p) = (1 - p)I_{\{k\}}(y) + p\pi_p(y; \mu), \quad y = 0, 1, 2, \dots,$$

sob a restrição de $0 \leq p \leq \frac{1}{1 - \pi_p(k; \mu)}$, em que $\pi_p(k; \mu)$ é a distribuição de Poisson tradicional avaliada no ponto k .

Casos particulares da distribuição k -MP:

- (i) Para $p = 0$, a distribuição k -MP configura-se como distribuição de Poisson k -Degenerada.
- (ii) Para $0 < p < 1$, a distribuição k -MP configura-se como distribuição de Poisson k -Inflacionada (k -IP).
- (iii) Para $p = 1$, a distribuição k -MP configura-se como distribuição de Poisson tradicional.
- (iv) Para $1 < p < \frac{1}{1 - \pi_p(k; \mu)}$, a distribuição k -MP configura-se como distribuição de Poisson k -Deflacionada (k -DP).

⁴ Usaremos a sigla k -MP, do inglês k -Modified Poisson, referindo-se à distribuição Poisson k -Modificada.

- (v) Para $p = \frac{1}{1 - \pi_p(k; \mu)}$, a distribuição k -MP configura-se como distribuição de Poisson k -Subtraída (k -SP).

3.2.2 Distribuição Geométrica k -Modificada (k -MG)

Considere Y uma variável aleatória tendo distribuição Geométrica com parâmetro de média $\mu > 0$, cuja função massa de probabilidade é dada por:

$$\pi_G(y; \mu) = \left(1 - \frac{1}{1 + \mu}\right)^y \frac{1}{1 + \mu} = \left(\frac{\mu}{1 + \mu}\right)^y \frac{1}{1 + \mu}, \quad y = 0, 1, 2, \dots \quad (3.5)$$

A distribuição k -modificada de Y , é dita distribuição Geométrica k -Modificada (k -MG⁵) é definida a partir de (3.5) e sua função massa de probabilidade é dada por:

$$\pi_{k-MG}(y; \mu, p) = (1 - p)I_{\{k\}}(y) + p\pi_G(y; \mu), \quad y = 0, 1, 2, \dots,$$

sob a restrição de $0 \leq p \leq \frac{1}{1 - \pi_G(k; \mu)}$, em que $\pi_G(k; \mu)$ é a distribuição Geométrica avaliada no ponto k .

Casos particulares da distribuição k -MG:

- (i) Para $p = 0$, a distribuição k -MG configura-se como distribuição Geométrica k -Degenerada.
- (ii) Para $0 < p < 1$, a distribuição k -MG configura-se como distribuição Geométrica k -Inflacionada (k -IG).
- (iii) Para $p = 1$, a distribuição k -MG configura-se como distribuição Geométrica tradicional.
- (iv) Para $1 < p < \frac{1}{1 - \pi_G(k; \mu)}$, a distribuição k -MG configura-se como distribuição Geométrica k -Deflacionada (k -DG).
- (v) Para $p = \frac{1}{1 - \pi_G(k; \mu)}$, a distribuição k -MG configura-se como distribuição Geométrica k -Subtraída (k -SG).

⁵ Usaremos a sigla k -MG, do inglês k -Modified Geometric, referindo-se à distribuição Geométrica k -Modificada.

3.2.3 Distribuição Binomial k -Modificada (k -MB)

Considere Y uma variável aleatória tendo distribuição Binomial com parâmetro de média $0 < \mu < m$, cuja função massa de probabilidade é dada por:

$$\pi_B(y; \mu) = \binom{m}{y} \left(\frac{\mu}{m-\mu} \right)^y \left(\frac{m}{m-\mu} \right)^m, \quad y = 0, 1, 2, \dots, m, \quad (3.6)$$

em que m é um número inteiro positivo, conhecido e maior que 1, correspondente ao número de ensaios de Bernoulli.

A distribuição k -modificada de Y , é dita distribuição Binomial k -Modificada (k -MB⁶) é definida a partir de (3.6) e sua função massa de probabilidade é dada por:

$$\pi_{k-MB}(y; \mu, p) = (1-p)I_{\{k\}}(y) + p\pi_B(y; \mu), \quad y = \{0, 1, 2, \dots, m\},$$

sob a restrição de $0 \leq p \leq \frac{1}{1 - \pi_B(k; \mu)}$, em que $\pi_B(k; \mu)$ é a distribuição Binomial avaliada no ponto k .

Casos particulares da distribuição k -MB:

- (i) Para $p = 0$, a distribuição k -MB configura-se como distribuição Binomial k -Degenerada.
- (ii) Para $0 < p < 1$, a distribuição k -MB configura-se como distribuição Binomial k -Inflacionada (k -IB).
- (iii) Para $p = 1$, a distribuição k -MB configura-se como distribuição de Binomial tradicional.
- (iv) Para $1 < p < \frac{1}{1 - \pi_B(k; \mu)}$, a distribuição k -MB configura-se como distribuição Binomial k -Deflacionada (k -DB).
- (v) Para $p = \frac{1}{1 - \pi_B(k; \mu)}$, a distribuição k -MB configura-se como distribuição Binomial k -Subtraída (k -SB).

⁶ Usaremos a sigla k -MB, do inglês k -Modified Binomial, referindo-se à distribuição Binomial k -Modificada.

3.2.4 Distribuição Borel k -Modificada (k -MBo)

Considere Y uma variável aleatória tendo distribuição Borel com parâmetro de média $\mu > 0$, cuja função massa de probabilidade é dada por:

$$\pi_{Bo}(y; \mu) = \frac{y^{y-2}}{(y-1)!} \frac{\left[e^{-(1-\frac{1}{\mu})} (1 - \frac{1}{\mu}) \right]^y}{1 - \frac{1}{\mu}}, \quad y = 0, 1, 2, \dots \quad (3.7)$$

A distribuição k -modificada de Y , é dita distribuição Borel k -Modificada (k -MBo⁷) é definida a partir de (3.7) e sua função massa de probabilidade é dada por:

$$\pi_{k-MBo}(y; \mu, p) = (1-p)I_{\{k\}}(y) + p\pi_{Bo}(y; \mu), \quad y = 0, 1, 2, \dots,$$

sob a restrição de $0 \leq p \leq \frac{1}{1 - \pi_{Bo}(k; \mu)}$, em que $\pi_{Bo}(k; \mu)$ é a distribuição Borel avaliada no ponto k .

Casos particulares da distribuição k -MBo:

- (i) Para $p = 0$, a distribuição k -MBo configura-se como distribuição Borel k -Degenerada.
- (ii) Para $0 < p < 1$, a distribuição k -MBo configura-se como distribuição Borel k -Inflacionada (k -IBo).
- (iii) Para $p = 1$, a distribuição k -MBo configura-se como distribuição Borel tradicional.
- (iv) Para $1 < p < \frac{1}{1 - \pi_{Bo}(k; \mu)}$, a distribuição k -MBo configura-se como distribuição Borel k -Deflacionada (k -DBo).
- (v) Para $p = \frac{1}{1 - \pi_{Bo}(k; \mu)}$, a distribuição k -MBo configura-se como distribuição Borel k -Subtraída (k -SBo).

⁷ Usaremos a sigla k -MBo, do inglês k -Modified Borel, referindo-se à distribuição Borel k -Modificada.

3.2.5 Distribuição Borel-Tanner k -Modificada (k -MBT)

Considere Y uma variável aleatória tendo distribuição Borel-Tanner com parâmetro de média $\mu > 0$, cuja função massa de probabilidade é dada por:

$$\pi_{BT}(y; \mu) = \frac{my^{y-m-1}}{(y-m)!} \frac{\left[e^{-(1-\frac{1}{\mu})} (1 - \frac{1}{\mu}) \right]^y}{1 - \frac{1}{\mu}}, \quad y = m, m+1, m+2, \dots, \quad (3.8)$$

em que m é um número inteiro conhecido. Se $m = 1$, temos a distribuição Borel.

A distribuição k -modificada de Y , é dita distribuição Borel-Tanner k -Modificada (k -MBT⁸) é definida a partir de (3.8) e sua função massa de probabilidade é dada por:

$$\pi_{k-MBT}(y; \mu, p) = (1-p)I_{\{k\}}(y) + p\pi_{BT}(y; \mu), \quad y = m, m+1, m+2, \dots,$$

sob a restrição de $0 \leq p \leq \frac{1}{1 - \pi_{BT}(k; \mu)}$, em que $\pi_{BT}(k; \mu)$ é a distribuição Borel-Tanner avaliada no ponto k .

Casos particulares da distribuição k -MBT:

- (i) Para $p = 0$, a distribuição k -MBT configura-se como distribuição Borel-Tanner k -Degenerada.
- (ii) Para $0 < p < 1$, a distribuição k -MBT configura-se como distribuição Borel-Tanner k -Inflacionada (k -IBT).
- (iii) Para $p = 1$, a distribuição k -MBT configura-se como distribuição Borel-Tanner tradicional.
- (iv) Para $1 < p < \frac{1}{1 - \pi_{BT}(k; \mu)}$, a distribuição k -MBT configura-se como distribuição Borel-Tanner k -Deflacionada (k -DBT).
- (v) Para $p = \frac{1}{1 - \pi_{BT}(k; \mu)}$, a distribuição k -MBT configura-se como distribuição Borel-Tanner k -Subtraída (k -SBT).

⁸ Usaremos a sigla k -MBT, do inglês k -Modified Borel-Tanner, referindo-se à distribuição Borel-Tanner k -Modificada.

3.2.6 Distribuição Haight k -Modificada (k -MH)

Considere Y uma variável aleatória tendo distribuição Haight com parâmetro de média $\mu > 0$, cuja função massa de probabilidade é dada por:

$$\pi_H(y; \mu) = \frac{(2y-2)!}{y!(y-1)!} \left[\frac{\mu(\mu-1)}{(2\mu-1)^2} \right]^y \frac{(2\mu-1)}{(\mu-1)}, \quad y = 1, 2, 3, \dots \quad (3.9)$$

A distribuição k -modificada de Y , é dita distribuição Haight k -Modificada (k -MH⁹) é definida a partir de (3.9) e sua função massa de probabilidade é dada por:

$$\pi_{k-MH}(y; \mu, p) = (1-p)I_{\{k\}}(y) + p\pi_H(y; \mu), \quad y = 1, 2, 3, \dots,$$

sob a restrição de $0 \leq p \leq \frac{1}{1 - \pi_H(k; \mu)}$, em que $\pi_H(k; \mu)$ é a distribuição Haight avaliada no ponto k .

Casos particulares da distribuição k -MH:

- (i) Para $p = 0$, a distribuição k -MH configura-se como distribuição Haight k -Degenerada.
- (ii) Para $0 < p < 1$, a distribuição k -MH configura-se como distribuição Haight k -Inflacionada (k -IH).
- (iii) Para $p = 1$, a distribuição k -MH configura-se como distribuição de Haight tradicional.
- (iv) Para $1 < p < \frac{1}{1 - \pi_H(k; \mu)}$, a distribuição k -MH configura-se como distribuição Haight k -Deflacionada (k -DH).
- (v) Para $p = \frac{1}{1 - \pi_H(k; \mu)}$, a distribuição k -MH configura-se como distribuição Haight k -Subtraída (k -SH).

⁹ Usaremos a sigla k -MH, do inglês k -Modified Haight, referindo-se à distribuição Haight k -Modificada.

3.3 Caracterização da distribuição k -MPS

Nesta seção mostraremos algumas propriedades da família de distribuições k -MPS, apresentando suas características em função das mesmas características da distribuição PS associada. Pretendemos com isso colaborar com a vasta literatura das distribuições discretas, como feito em Consul (1990); Consul e Famoye (2006); Consul e Shenton (1973); Gupta (1974); Khatri (1959).

Considere Y uma variável aleatória com distribuição de probabilidade k -MPS(μ, p), cuja função massa de probabilidade foi definida em (3.1). A seguir são apresentadas algumas características de Y de acordo com a teoria de probabilidades.

3.3.1 Função de Distribuição

Define-se a função de distribuição da variável aleatória Y por:

$$\begin{aligned} F_Y(b) &= P(Y \leq b) = \sum_{y:y \leq b} \{(1-p)I_{\{k\}}(y) + p\pi_{PS}(y; \mu)\} \\ &= \sum_{y:y \leq b} (1-p)I_{\{k\}}(y) + \sum_{y:y \leq b} p\pi_{PS}(y; \mu) \\ &= (1-p) \sum_{y:y \leq b} I_{\{k\}}(y) + pF(b), \quad b \in \mathbb{R}. \end{aligned}$$

Logo,

$$F_Y(b) = \begin{cases} 0, & \text{se } b < s \\ pF(b), & \text{se } s \leq b < k \\ 1 - p(1 - F(b)), & \text{se } b \geq k \end{cases} .$$

em que $F(b)$ é a função de distribuição da PS associada a Y no ponto b .

3.3.2 Estatística de Ordem

Considere que $(Y_{(1)}, Y_{(2)}, \dots, Y_{(n)})$ seja uma amostra aleatória ordenada de n realizações da variável aleatória Y , tal que $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$. Os elementos desta amostra são denominados estatísticas de ordem da amostra aleatória de Y e a sua i -ésima estatística de ordem é representada por $Y_{(i)}$, para $i = 1, 2, \dots, n$. As funções de distribuições das estatísticas de ordem do mínimo ($Y_{(1)}$) e do máximo ($Y_{(n)}$) podem ser obtidas da seguinte forma.

A função de distribuição do mínimo da amostra, $Y_{(1)}$, é dada por:

$$\begin{aligned}
 F_{Y_{(1)}}(b) &= P(Y_{(1)} \leq b) = 1 - P(Y_{(1)} > b) \\
 &= 1 - P(Y_1 > b, Y_2 > b, \dots, Y_n > b) \\
 &= 1 - \prod_{i=1}^n P(Y_i > b) \\
 &= 1 - \prod_{i=1}^n (1 - P(Y_i \leq b)) \\
 &= 1 - (1 - P(Y \leq b))^n \\
 &= 1 - (1 - F_Y(b))^n.
 \end{aligned}$$

Logo,

$$F_{Y_{(1)}}(b) = \begin{cases} 0, & \text{se } b < s \\ 1 - (1 - pF(b))^n, & \text{se } s \leq b < k \\ 1 - (p(1 - F(b)))^n, & \text{se } b \geq k \end{cases}.$$

A função de distribuição do máximo da amostra, $Y_{(n)}$, é dada por:

$$\begin{aligned}
 F_{Y_{(n)}}(b) &= P(Y_{(n)} \leq b) \\
 &= P(Y_1 \leq b, Y_2 \leq b, \dots, Y_n \leq b) \\
 &= P(Y_1 \leq b)P(Y_2 \leq b), \dots, P(Y_n \leq b) \\
 &= \prod_{i=1}^n P(Y_i \leq b) \\
 &= (F_Y(b))^n.
 \end{aligned}$$

Logo,

$$F_{Y_{(n)}}(b) = \begin{cases} 0, & \text{se } b < s \\ (pF(b))^n, & \text{se } s \leq b < k \\ (1 - p(1 - F(b)))^n, & \text{se } b \geq k \end{cases}.$$

3.3.3 Função Geradora de Probabilidade

A função geradora de probabilidade de Y , $G_Y(z)$, com $|z| \leq 1$, é dada por:

$$\begin{aligned}
 G_Y(z) &= E[z^Y] \\
 &= \sum_{y=s}^{\infty} z^y \pi_{k-MPS}(y; \mu, p) \\
 &= \sum_{y=s}^{\infty} z^y \{ (1-p)I_{\{k\}}(y) + p\pi_{PS}(y; \mu) \} \\
 &= \sum_{y=s}^{\infty} z^y (1-p)I_{\{k\}}(y) + \sum_{y=s}^{\infty} z^y p\pi_{PS}(y; \mu) \\
 &= (1-p)z^k + p \sum_{y=s}^{\infty} z^y p\pi_{PS}(y; \mu) \\
 &= (1-p)z^k + pG(z),
 \end{aligned}$$

para todo z tal que $G(z)$ exista, em que $G(z)$ é a função geradora de probabilidade da distribuição PS associada a Y .

3.3.4 Função Característica

A função característica de Y , $\varphi_Y(t)$, é dada por:

$$\begin{aligned}
 \varphi_Y(t) &= E[e^{itY}] \\
 &= \sum_{y=s}^{\infty} e^{ity} \pi_{k-MPS}(y; \mu, p) \\
 &= \sum_{y=s}^{\infty} e^{ity} \{ (1-p)I_{\{k\}}(y) + p\pi_{PS}(y; \mu) \} \\
 &= \sum_{y=s}^{\infty} e^{ity} (1-p)I_{\{k\}}(y) + p \sum_{y=s}^{\infty} e^{ity} \pi_{PS}(y; \mu) \\
 &= e^{itk} (1-p) + p\varphi(t),
 \end{aligned}$$

em que $\varphi(t)$ é a função característica da distribuição PS associada a Y .

3.3.5 Função Geradora de Momentos

A função geradora de momentos de Y , $\mathcal{M}_Y(t)$, é dada por:

$$\begin{aligned}
 \mathcal{M}_Y(t) &= E[e^{tY}] \\
 &= \sum_{y=s}^{\infty} e^{ty} \pi_{k-MPS}(y; \mu, p) \\
 &= \sum_{y=s}^{\infty} e^{ty} \{ (1-p)I_{\{k\}}(y) + p\pi_{PS}(y; \mu) \} \\
 &= \sum_{y=s}^{\infty} e^{ty} (1-p)I_{\{k\}}(y) + p \sum_{y=s}^{\infty} e^{ty} \pi_{PS}(y; \mu) \\
 &= e^{tk}(1-p) + p\mathcal{M}(t),
 \end{aligned}$$

para todo t tal que $\mathcal{M}(t)$ exista, em que $\mathcal{M}(t)$ é a função geradora de momentos da distribuição PS associada a Y .

Ao derivar $\mathcal{M}_Y(t)$ sucessivamente em relação à t , obtemos

$$\begin{aligned}
 \mathcal{M}'_Y(t) &= (1-p)ke^{tk} + p\mathcal{M}'(t) \\
 \mathcal{M}''_Y(t) &= (1-p)k^2e^{tk} + p\mathcal{M}''(t) \\
 \mathcal{M}'''_Y(t) &= (1-p)k^3e^{tk} + p\mathcal{M}'''(t) \\
 &\quad \vdots \quad \quad \quad \vdots \\
 \mathcal{M}^{(r)}_Y(t) &= (1-p)k^r e^{tk} + p\mathcal{M}^{(r)}(t),
 \end{aligned}$$

que, de forma geral, temos:

$$\mathcal{M}_Y^{(r)}(t) = \frac{d^r \mathcal{M}_Y(t)}{dt^r} = (1-p)k^r e^{tk} + p\mathcal{M}^{(r)}(t), \quad (3.10)$$

em que (r) indica a derivada de r -ésima ordem de $\mathcal{M}_Y(t)$ e $\mathcal{M}(t)$ em relação a t .

Uma vez que a equação (3.10) depende de t , conseguimos obter o r -ésimo momento populacional, $\mu_{k-MPS}^{(r)} = E[Y^r]$, $r \geq 1$, fazendo $t = 0$:

$$\mu_{k-MPS}^r = (1-p)k^r + p\mu^r, \quad (3.11)$$

em que μ^r é o momento de ordem r da distribuição PS associada a Y . Desta forma, podemos obter a média e a variância da variável aleatória Y a partir do primeiro e segundo momento populacional.

Proposição 3.1. Seja Y uma variável aleatória com distribuição k -MPS, então a sua média μ_{k-MPS} e a sua variância σ_{k-MPS}^2 são dadas respectivamente por:

$$\mu_{k-MPS} = (1-p)k + p\mu \quad \text{e} \quad \sigma_{k-MPS}^2 = p \{ (1-p)(k-\mu)^2 + \sigma^2 \}.$$

Prova. As demonstrações destas medidas são obtidas diretamente utilizando a equação dada em (3.11). \square

Nota: Utilizando a Desigualdade de Jensen é possível mostrar que, para todo valor de p , a variância $\sigma_{k-MPS}^2 \geq 0$.

Na Tabela 2 são apresentadas as variâncias das distribuições da família k -MPS e das distribuições PS associadas. Nota-se que a variância σ_{k-MPS}^2 tem ordem quadrática para valores de k e μ e, além disso, nos casos em que $k = \mu$, a variância σ_{k-MPS}^2 atinge seu menor valor.

Tabela 2 – Variância das distribuições PS e k -MPS.

Distribuições	σ^2	σ_{k-MPS}^2
Poisson	μ	$p \{ (1-p)(k-\mu)^2 + \mu \}$
Geométrica	$\mu(1+\mu)$	$p \{ (1-p)(k-\mu)^2 + \mu(1+\mu) \}$
Binomial	$\mu(1-\frac{\mu}{m})$	$p \{ (1-p)(k-\mu)^2 + \mu(1-\frac{\mu}{m}) \}$
Borel	$\mu^2(\mu-1)$	$p \{ (1-p)(k-\mu)^2 + \mu^2(\mu-1) \}$
Borel-Tanner	$(\frac{\mu}{m})^2(\mu-m)$	$p \{ (1-p)(k-\mu)^2 + (\frac{\mu}{m})^2(\mu-m) \}$
Haight	$\frac{2\mu}{(2\mu-1)^3}$	$p \left\{ (1-p)(k-\mu)^2 + \frac{2\mu}{(2\mu-1)^3} \right\}$

A função geradora de momentos está relacionada com a função geradora de probabilidade por:

$$\mathcal{M}_Y(t) = E[e^{tY}] = G_Y[e^t].$$

Uma ilustração do comportamento da variância σ_{k-MPS}^2 para as distribuições k -MP, k -MG e k -MB em função do parâmetro p , para os seguintes valores de $\mu = 0,5; 1, 1,5; 2$ e $2,5$ e para os respectivos pontos de modificação $k = 0, 2, 5$ e 7 , são apresentadas nas Figuras 2 - 5, em que no caso da distribuição k -MB o parâmetro m , referente ao número de ensaio de Bernoulli, foi fixado em 10. Nestes gráficos, a linha tracejada em $p = 1$ corresponde aos valores das variâncias das distribuições PS. Para valores de p à esquerda desta linha, temos valores das variâncias σ_{k-MPS}^2 de variáveis aleatórias com distribuições k -IPS. Já para valores de p à direita desta linha, temos valores das variâncias σ_{k-MPS}^2 de variáveis aleatórias com distribuições k -DPS.

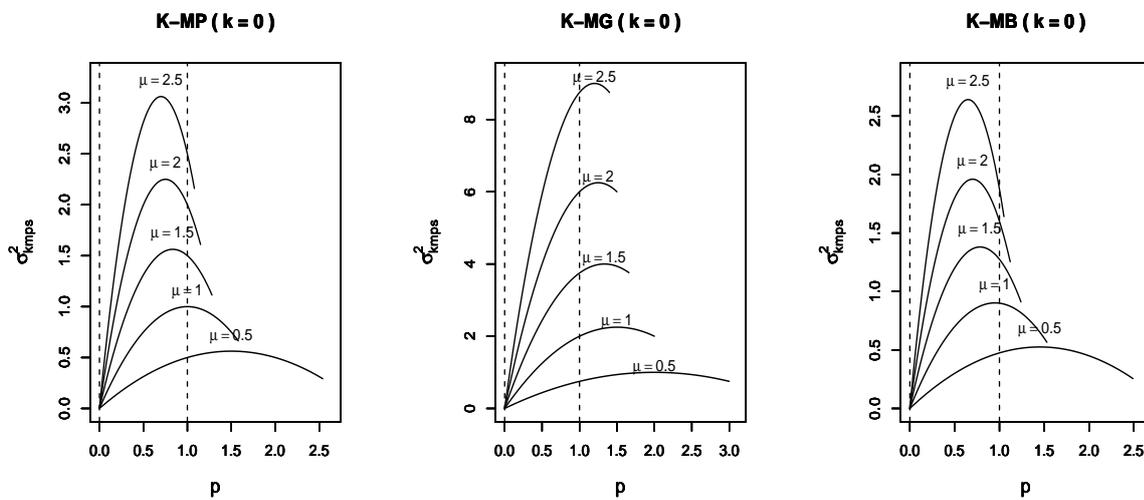


Figura 2 – Valores da variância no ponto de modificação $k=0$.

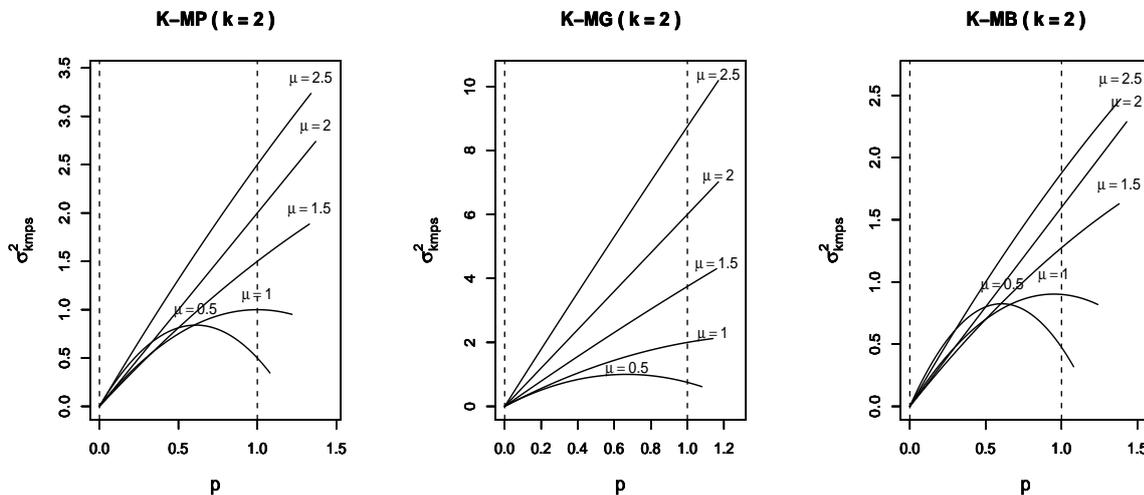


Figura 3 – Valores da variância no ponto de modificação $k=2$.

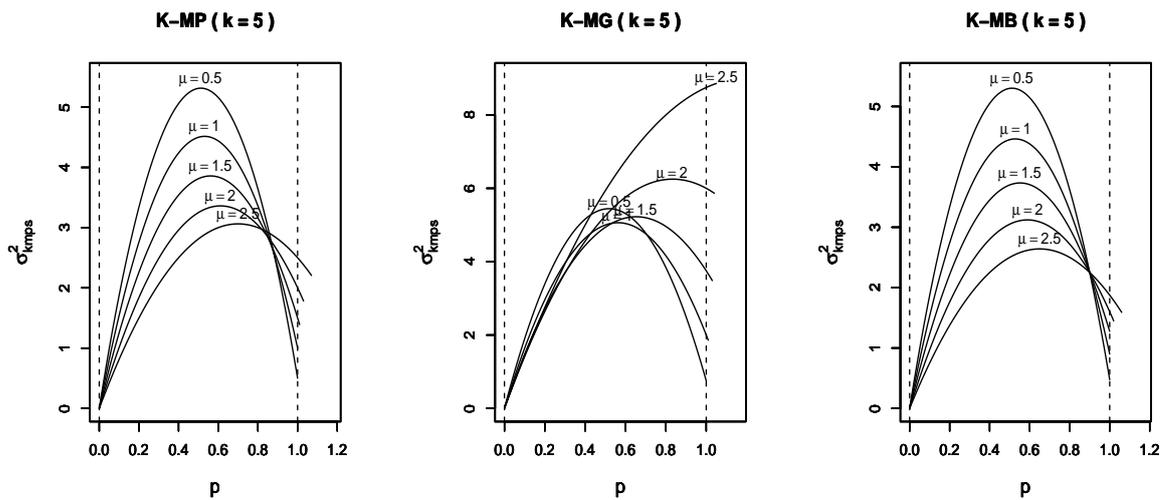


Figura 4 – Valores da variância no ponto de modificação $k=5$.

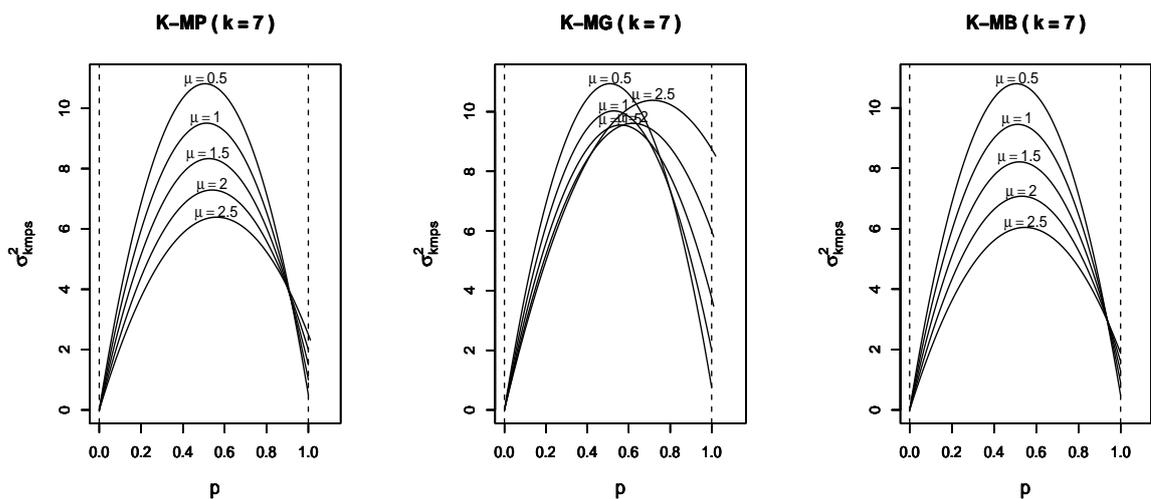


Figura 5 – Valores da variância no ponto de modificação $k=7$.

Das Figuras 2 - 5 é possível observar que a variância da distribuição k -MPS tem ordem quadrática. Além disso, como os valores para o parâmetro μ são baixos, a medida em que aumenta-se os valores do ponto de modificação k , o valor do parâmetro de modificação p limita-se ao valor máximo 1. Isto ocorre pelo fato de que para valores grandes do ponto de modificação k , dado esses valores de μ , a função de massa da distribuição PS tende a valores muito pequenos próximos de zero, $\pi_{PS}(k; \mu) \rightarrow 0$, fazendo com que o valor de p se aproxime do valor 1, $p = \frac{1}{1 - \pi_{PS}(k; \mu)} \rightarrow 1$, fazendo com que as distribuições k -DPS se aproximem das distribuições PS. Na Figura 5 este fato é ilustrado com clareza.

3.3.6 Gerador de Números Pseudo-Aleatórios

Em trabalhos que fazem uso de teorias estatísticas com aplicações em casos reais, recomenda-se um estudo de simulação para validar o modelo proposto utilizado na análise de conjunto de dados.

Considere Y como uma variável aleatória com distribuição k -MPS. Apresentaremos aqui um algoritmo que gera valores para a variável aleatória Y usando o método da inversa, levando em consideração as restrições pertinentes aos parâmetros μ e p e que $y \in A_s$.

O método da inversa:

- (i) Seja U uma variável aleatória tal que $U \sim \text{Uniforme}(0, 1)$;
- (ii) Então a variável aleatória $V = \min \{y \geq s : U \leq F_Y(y)\}$ tem função de distribuição $F_Y(y)$, com suporte $\{s, s + 1, s + 2, \dots\}$.

O algoritmo para gerar números aleatórios de uma distribuição $k - MPS(\mu, p)$ é descrito nas etapas a seguir.

Algoritmo 1 – Gerador de número aleatório (rkmps).

- 1: **procedimento** RKMPS
 - 2: **Entrada:** s, k, μ, p
 - 3: **Saída:** Valor y gerado da distribuição k -MPS.
 - 4: **Início:**
 - 5: $y = s$
 - 6: **calcular:**
 - 7: $\pi_{k-MPS}(y; \mu, p) = (1 - p)I_{\{k\}}(y) + p\pi_{PS}(y; \mu)$.
 - 8: $F_Y(y) = \pi_{k-MPS}(y; \mu, p)$
 - 9: **gerar:** $u \sim U(0, 1)$
 - 10: **enquanto** ($u > F_Y(y)$) **faça**
 - 11: $y = y + 1$
 - 12: $F_Y(y) = F_Y(y - 1) + \pi_{k-MPS}(y; \mu, p)$
 - 13: **fim enquanto**
 - 14: **retorna** y .
 - 15: **fim procedimento**
-

Nas Figuras 6 - 8 são apresentados os gráficos das distribuições de frequência dos valores simulados da variável aleatória Y com distribuição k -MPS. Em cada gráfico, o título k -MPS($k; \mu, p$) refere-se, respectivamente, ao nome da distribuição, ao ponto de modificação k , ao parâmetro de média μ da distribuição PS associada e ao parâmetro de modificação p . Foram simulados valores para amostras de tamanho $n = 200$ das distribuições k -MP, k -MG e k -MB, sendo que para cada distribuição consideramos dois valores para o ponto de modificação $k = 0$ e 5 , três valores para o parâmetro de média $\mu = 0,5; 2$ e 5 , parâmetro de modificação $p = 0,7$ (caso inflacionado) e para o caso deflacionado o valor para o parâmetro p foi determinado pela média dos valores no intervalo $(1, \frac{1}{1-\pi_{PS}(k;\mu)})$ e o parâmetro m , para uso da k -MB, foi fixado em $m = 10$.

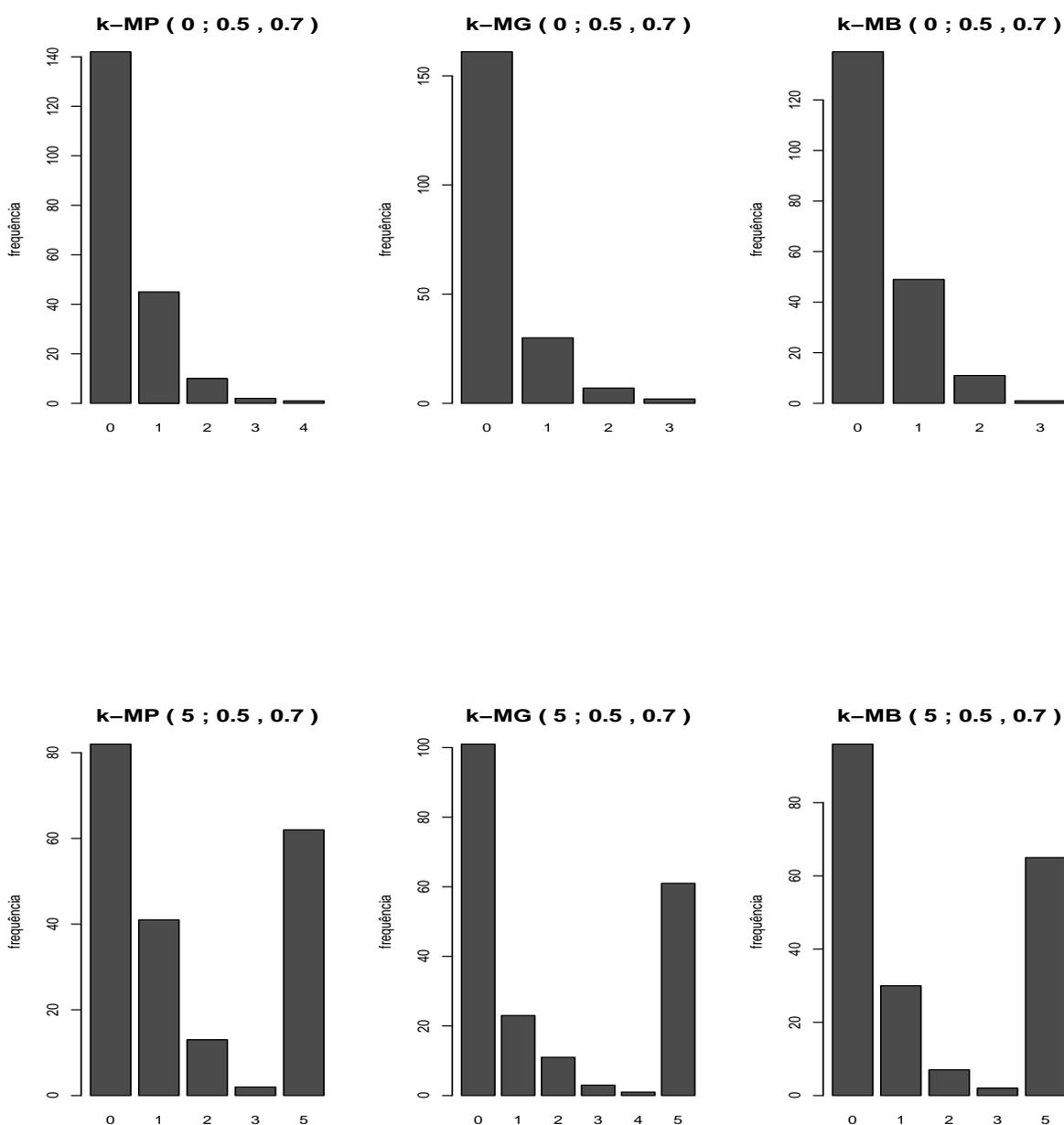


Figura 6 – Gráficos das distribuições de frequências para as distribuições k -MP, k -MG e k -MB nos pontos de modificação $k = 0$ e $k = 5$, parâmetro $\mu = 0,5$ e parâmetro de modificação $p = 0,7$.

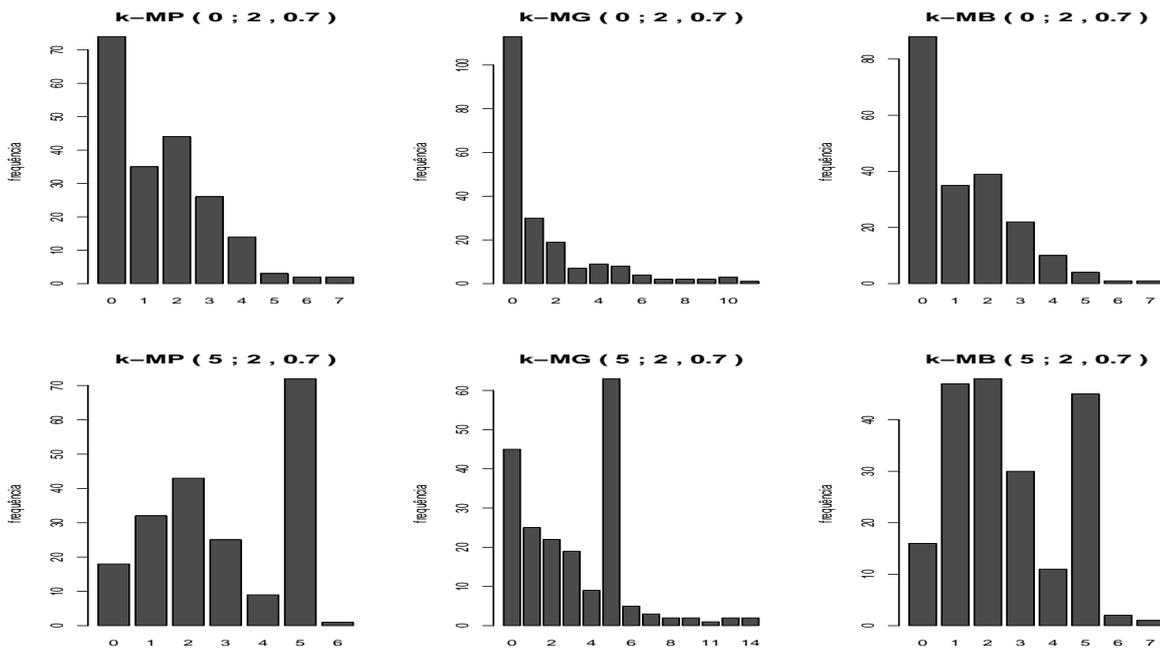


Figura 7 – Gráficos das distribuições de frequências para as distribuições k -MP, k -MG e k -MB nos pontos de modificação $k = 0$ e $k = 5$, parâmetro $\mu = 2$ e parâmetro de modificação $p = 0,7$.

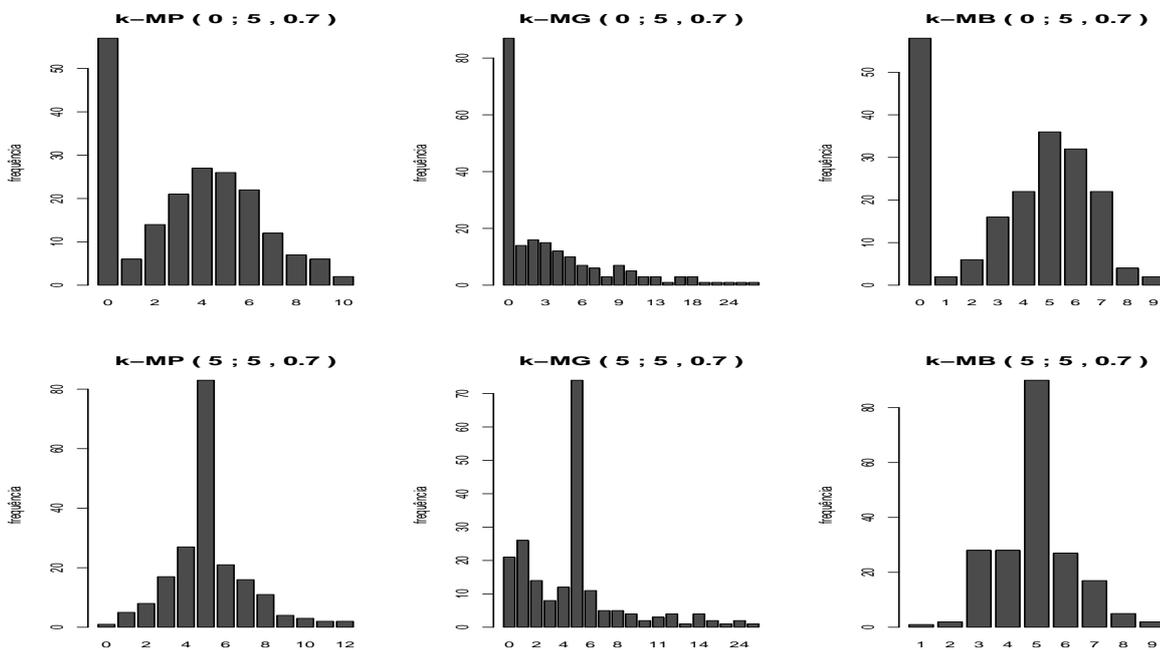


Figura 8 – Gráficos das distribuições de frequências para as distribuições k -MP, k -MG e k -MB nos pontos de modificação $k = 0$ e $k = 5$, parâmetro $\mu = 5$ e parâmetro de modificação $p = 0,7$.

Observando as Figuras 6 - 8, notamos o efeito que o valor do parâmetro de modificação p e o parâmetro μ causam na frequência do ponto de modificação k . Nestes casos, como o valor de p foi fixado em 0,7, há uma inflação no ponto de modificação k , ou seja, a probabilidade de ocorrência da observação k para essas distribuições são maiores do que as probabilidades das suas distribuições PS associadas. Além disso, a inflação é mais contundente quando o ponto de modificação k estiver próximo do valor do parâmetro μ , como nos gráficos em que $k = 0$ e $\mu = 0,5$ ou $k = 5$ e $\mu = 5$. Nos casos em que o ponto de modificação k estiver distante do valor do parâmetro μ , como nos gráficos em que $k = 0$ e $\mu = 5$ ou $k = 5$ e $\mu = 0,5$, a probabilidade de ocorrência da observação k será necessariamente baixa, consequentemente sua frequência também será baixa. No entanto, distribuições de frequências que apresentam frequências elevadas em algum ponto k distante da média μ , possivelmente serão casos de conjunto de dados inflacionados.

3.4 Distribuições k -MPS e sua Versão Hurdle

Frequentemente os modelos propostos para analisar conjuntos de dados de contagem com inflação de observações em algum ponto $k \in A_s$ são classificados como modelos de mistura, ver artigo [Murat e Szydal \(1998\)](#)). Nesta seção mostraremos uma forma de escrever e interpretar as distribuições k -MPS também como distribuições de mistura, a versão *hurdle*, porém com a vantagem de modelar os casos de inflação ou deflação de uma observação k em relação às distribuições PS.

Teorema 3. Se Y é uma variável aleatória com distribuição k -MPS, então podemos reescrevê-la como uma distribuição *hurdle*, cuja função massa de probabilidade é dada por:

$$\pi_{k-MPS}(y; \mu, \omega) = (1 - \omega)I_{\{k\}}(y) + \omega\pi_{k-SPS}(y; \mu), y \in A_s, \quad (3.12)$$

sendo ω o parâmetro de modificação da probabilidade da distribuição k -SPS associada, sob a seguinte restrição:

$$0 \leq \omega \leq 1;$$

e $I_{\{k\}}(y)$ é uma função indicadora definida por:

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

Prova. A partir da função $\pi_{k-MPS}(y; \mu, p)$, dada em (3.1), podemos escrevê-la da seguinte forma:

$$\begin{aligned} \pi_{k-MPS}(y; \mu, p) &= (1 - p + p\pi_{PS}(k; \mu))I_{\{k\}}(y) + p\pi_{PS}(y; \mu)(1 - I_{\{k\}}(y)) \\ &= (1 - p(1 - \pi_{PS}(k; \mu)))I_{\{k\}}(y) + p\pi_{PS}(y; \mu)(1 - I_{\{k\}}(y)). \end{aligned}$$

Multiplicando e dividindo o segundo termo por $(1 - \pi_{PS}(k; \mu))$, temos:

$$\pi_{k-MPS}(y; \mu, p) = (1 - p(1 - \pi_{PS}(k; \mu)))I_{\{k\}}(y) + p \frac{(1 - \pi_{PS}(k; \mu))}{(1 - \pi_{PS}(k; \mu))} \pi_{PS}(y; \mu) (1 - I_{\{k\}}(y)).$$

Fazendo $\omega = p(1 - \pi_{PS}(k; \mu))$, temos:

$$\begin{aligned} \pi_{k-MPS}(y; \mu, \omega) &= (1 - \omega)I_{\{k\}}(y) + \omega \frac{\pi_{PS}(y; \mu)}{(1 - \pi_{PS}(k; \mu))} (1 - I_{\{k\}}(y)) \\ &= (1 - \omega)I_{\{k\}}(y) + \omega \pi_{k-SPS}(y; \mu). \end{aligned}$$

□

Notamos que a distribuição k -MPS escrita como distribuição *hurdle* tem como distribuição associada a distribuição k -SPS, o que a torna diferente das distribuições de misturas tradicionais. Além disso, é fácil verificar que a probabilidade de ocorrência do evento $y = k$ é dado por $1 - \omega$, e a probabilidade do evento $y \neq k$, é dada por $\omega \pi_{k-SPS}(y; \mu)$. Como consequência, temos que as distribuições *hurdle* podem ser interpretadas como a superposição de dois processos:

- i) o que irá produzir as observações diferentes de k , baseadas nas distribuições k -SPS;
- ii) e um segundo processo em que somente serão produzidas observações k .

Corolário 3.6. Seja Y uma variável aleatória cuja função massa de probabilidades é dada por (3.12). Então a sua média e variância são dadas, respectivamente, por:

$$\mu_{k-MPS} = (1 - \omega)k + \omega \mu_{k-SPS}$$

e

$$\sigma_{k-MPS}^2 = k^2 + \frac{\omega(\mu^2 - k^2)}{1 - \pi_{PS}(k; \mu)} - ((1 - \omega)k + \omega \mu_{k-SPS})^2.$$

Prova. Ao substituir o valor de p por $\frac{\omega}{1 - \pi_{PS}(k; \mu)}$ em μ_{k-MPS} e σ_{k-MPS}^2 dadas pela proposição 3.1 a demonstração deste Corolário é obtida de forma imediata. □

MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS E SELEÇÃO DE MODELOS

Considere $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ uma amostra aleatória da variável aleatória Y com distribuição k -MPS(μ, p) e $y = (y_1, y_2, \dots, y_n)$ o vetor de observações associado a \mathbf{Y} .

4.1 Método dos Momentos

Podemos obter os estimadores dos parâmetros μ e p da distribuição k -MPS através do método dos momentos, relacionando os momentos populacionais e amostrais correspondentes. Para uma amostra aleatória \mathbf{Y} , o r -ésimo momento amostral, denotado por $M^{(r)}$, é definido por:

$$M^{(r)} = \frac{1}{n} \sum_{i=1}^n Y_i^r, \quad r = 1, 2, \dots$$

Ao igualar cada momento populacional $\mu_{k-MPS}^{(r)}$ ao respectivo r -ésimo momento amostral $M^{(r)}$, $r = 1, 2$, obtemos um sistema de duas equações envolvendo os parâmetros do modelo:

$$\mu_{k-MPS}^{(r)}(\mu, p) = M^{(r)}, \quad r = 1, 2.$$

De forma geral, temos as seguintes expressões:

$$\mu_{k-MPS}^{(1)} = E(Y) = \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{e} \quad \mu_{k-MPS}^{(2)} = E(Y^2) = \frac{1}{n} \sum_{i=1}^n Y_i^2.$$

Utilizando a relação $E(Y^2) = \text{Var}(Y) + E^2(Y)$, podemos obter os estimadores de momentos de μ e p a partir da solução do sistema de equações:

$$\left\{ \begin{array}{l} (1-p)k + p\mu = \frac{1}{n} \sum_{i=1}^n Y_i \\ p(\mu^2 + \sigma^2) + (1-p)k^2 = \frac{1}{n} \sum_{i=1}^n Y_i^2 \end{array} \right. ,$$

sendo a variância σ^2 uma função de μ como pode ser visto na Tabela 2.

Considerando a parametrização dada em (3.12), também podemos obter os estimadores de momentos de μ e ω a partir da solução do sistema de equações.

$$\left\{ \begin{array}{l} (1-\omega)k + \omega\mu_{k-SPS} = \frac{1}{n} \sum_{i=1}^n Y_i \\ k^2 + \frac{\omega(\mu^2 - k^2)}{1 - \pi_{PS}(k; \mu)} - ((1-\omega)k + \omega\mu_{k-SPS})^2 = \frac{1}{n} \sum_{i=1}^n Y_i^2 \end{array} \right. ,$$

em que k é o ponto de modificação conhecido e μ_{k-SPS} é uma função de μ .

Se em cada uma destes sistemas de equações existir uma única solução, sendo estas soluções denotadas, respectivamente, por $\hat{\mu}, \hat{p}$ ou $\hat{\omega}$, dizemos que estes são os estimadores de μ , p e ω obtidos pelo método dos momentos, sendo que $\hat{\mu} = h_{\mu}(M^{(1)}, M^{(2)})$ estima μ , $\hat{p} = h_p(M^{(1)}, M^{(2)})$ estima p e $\hat{\omega} = h_{\omega}^*(M^{(1)}, M^{(2)})$ estima ω .

4.2 Método de Máxima Verossimilhança

Nesta seção abordaremos o método de máxima verossimilhança para estimar os parâmetros μ e p da distribuição k -MPS.

Reunimos as informações do vetor de observações \mathbf{y} da seguinte forma.

$$\mathcal{D} = (\mathbf{y}, n, n_j),$$

em que n_j é o número de observações j no vetor \mathbf{y} . Construiremos a função de verossimilhança levando em consideração a distribuição k -MPS dada em (3.1) e a sua forma alternativa dada em (3.12).

4.2.1 Função de Verossimilhança para k -MPS(μ, p)

A função de verossimilhança é dada por:

$$\begin{aligned}
L(\mu, p; \mathcal{D}) &= \prod_{i=1}^n \{ (1-p) I_{\{k\}}(y_i) + p \pi_{PS}(y_i; \mu) \} \\
&= \prod_{i=1}^n \left\{ (1-p + p \pi_{PS}(k; \mu))^{I_{\{k\}}(y_i)} (p \pi_{PS}(y_i; \mu))^{1-I_{\{k\}}(y_i)} \right\} \\
&= (1-p + p \pi_{PS}(k; \mu))^{n_k} \prod_{j \in A_S; j \neq k}^{\infty} (p \pi_{PS}(j; \mu))^{n_j},
\end{aligned}$$

em que $n = n_k + \sum_{j \in A_S; j \neq k}^{\infty} n_j$.

A função log-verossimilhança é dado por:

$$\ell(\mu, p; \mathcal{D}) = n_k \log(1-p + p \pi_{PS}(k; \mu)) + \sum_{j \in A_S; j \neq k}^{\infty} n_j \log(p \pi_{PS}(j; \mu)).$$

O vetor escore U , de dimensão (2×1) , é obtido a partir da primeira derivada parcial da função log-verossimilhança em relação aos parâmetros p e μ tal que $U = (U_p, U_\mu)^T$. Seus elementos são dados por:

$$\begin{aligned}
U_p &= \frac{\partial \ell(\mu, p; \mathcal{D})}{\partial p} = \frac{-n_k(1 - \pi_{PS}(k; \mu))}{1 - p + p \pi_{PS}(k; \mu)} + \sum_{j \in A_S; j \neq k}^{\infty} n_j \frac{\pi_{PS}(j; \mu)}{p \pi_{PS}(j; \mu)} \\
&= \frac{-n_k(1 - \pi_{PS}(k; \mu))}{1 - p + p \pi_{PS}(k; \mu)} + \frac{n - n_k}{p}. \\
U_\mu &= \frac{\partial \ell(\mu, p; \mathcal{D})}{\partial \mu} = \frac{n_k p \frac{\partial}{\partial \mu} (\pi_{PS}(k; \mu))}{1 - p + p \pi_{PS}(k; \mu)} + \sum_{j \in A_S; j \neq k}^{\infty} \frac{n_j \frac{\partial}{\partial \mu} (\pi_{PS}(j; \mu))}{\pi_{PS}(j; \mu)}. \tag{4.1}
\end{aligned}$$

Seja a matriz J , denominada matriz de informação observada, obtida a partir da primeira derivada de U e, definida por:

$$J = - \begin{bmatrix} U_{pp} & U_{p\mu} \\ U_{\mu p} & U_{\mu\mu} \end{bmatrix},$$

cujos elementos são:

$$U_{pp} = \frac{\partial}{\partial p} \left(\frac{-n_k(1 - \pi_{PS}(k; \mu))}{1 - p + p \pi_{PS}(k; \mu)} + \frac{n - n_k}{p} \right) = \frac{-n_k(1 - \pi_{PS}(k; \mu))^2}{(1 - p + p \pi_{PS}(k; \mu))^2} - \frac{(n - n_k)}{p^2}.$$

$$U_{\mu\mu} = n_k p \left(\frac{\frac{\partial^2}{\partial \mu^2} (\pi_{PS}(k; \mu)) (1 - p + p\pi_{PS}(k; \mu)) - p \left(\frac{\partial}{\partial \mu} (\pi_{PS}(k; \mu)) \right)^2}{(1 - p + p\pi_{PS}(k; \mu))^2} \right) + \sum_{j \in A_s; j \neq k}^{\infty} n_j \left(\frac{\frac{\partial^2}{\partial \mu^2} (\pi_{PS}(j; \mu)) \pi_{PS}(j; \mu) - \left(\frac{\partial}{\partial \mu} (\pi_{PS}(j; \mu)) \right)^2}{(\pi_{PS}(j; \mu))^2} \right).$$

$$U_{\mu p} = U_{p\mu} = \frac{\partial}{\partial \mu} \left(\frac{-n_k(1 - \pi_{PS}(k; \mu))}{1 - p + p\pi_{PS}(k; \mu)} + \frac{n - n_k}{p} \right) = \frac{n_k \frac{\partial}{\partial \mu} (\pi_{PS}(k; \mu))}{(1 - p + p\pi_{PS}(k; \mu))^2}.$$

Igualando a função score U de cada parâmetro a zero obtemos os estimadores de máxima verossimilhança (EMVs). Somente para o parâmetro p o estimador de máxima verossimilhança pode ser obtido explicitamente:

$$\hat{p} = \frac{n - n_k}{n(1 - \pi_{PS}(k; \hat{\mu}))}.$$

Nota: Note que para $n_k = 0$, o estimador de máxima verossimilhança de p , \hat{p} , assume o valor $\frac{1}{(1 - \pi_{PS}(k; \hat{\mu}))}$, que é o limite superior de p . Este fato implica que o vetor \mathbf{y} tem distribuição k -SPS. Logo, a probabilidade do evento $y = k$ é nula.

No caso do parâmetro μ , sua estimativa de máxima verossimilhança pode ser obtida através de algum método iterativo para estimação de parâmetros, como por exemplo, o método Newton-Raphson.

Substituindo \hat{p} na equação (4.2.1) obtemos:

$$\tilde{U}_{\mu} = \frac{(n - n_k) \frac{\partial}{\partial \mu} (\pi_{PS}(k; \mu))}{1 - \pi_{PS}(k; \mu)} + \sum_{j \in A_s; j \neq k}^{\infty} \frac{n_j \frac{\partial}{\partial \mu} (\pi_{PS}(j; \mu))}{\pi_{PS}(j; \mu)}.$$

A função $\tilde{U}_{\mu\mu}$ é obtida através da primeira derivada de \tilde{U}_{μ} em relação ao parâmetro μ e, é dada por:

$$\begin{aligned} \tilde{U}_{\mu\mu} = & (n - n_k) \left(\frac{\frac{\partial^2}{\partial \mu^2} (\pi_{PS}(k; \mu)) (1 - \pi_{PS}(k; \mu)) + \left(\frac{\partial}{\partial \mu} (\pi_{PS}(k; \mu)) \right)^2}{(1 - \pi_{PS}(k; \mu))^2} \right) + \\ & \sum_{j \in A_s; j \neq k}^{\infty} n_j \left(\frac{\frac{\partial^2}{\partial \mu^2} (\pi_{PS}(j; \mu)) \pi_{PS}(j; \mu) - \left(\frac{\partial}{\partial \mu} (\pi_{PS}(j; \mu)) \right)^2}{(\pi_{PS}(j; \mu))^2} \right). \end{aligned}$$

Como as equações acima são não lineares, para obtermos a estimativa de máxima verossimilhança de μ , devemos recorrer à algum método iterativo para estimação de parâmetros. Além disso, é importante observar que este procedimento não inclui as covariâncias referentes ao parâmetro p , sendo assim, não podemos utiliza-lo para realizar inferência sobre este parâmetro.

4.2.2 Função de Verossimilhança para k -MPS($\mu; \omega$)

Para as distribuições escritas na versão *hurdle*, a função de verossimilhança associada a y é dada por:

$$\begin{aligned} L(\mu, \omega; \mathcal{D}) &= \prod_{i=1}^n \{ (1 - \omega) I_{\{k\}}(y_i) + \omega \pi_{k-SPS}(y_i; \mu) \} \\ &= \prod_{i=1}^n \left\{ (1 - \omega + \omega \pi_{k-SPS}(k; \mu))^{I_{\{k\}}(y_i)} (\omega \pi_{k-SPS}(y_i; \mu))^{1 - I_{\{k\}}(y_i)} \right\} \\ &= \prod_{i=1}^n \left\{ (1 - \omega)^{I_{\{k\}}(y_i)} (\omega \pi_{k-SPS}(y_i; \mu))^{1 - I_{\{k\}}(y_i)} \right\} \\ &= (1 - \omega)^{n_k} \omega^{n - n_k} \prod_{j \in A_s; j \neq k}^{\infty} (\pi_{k-SPS}(j; \mu))^{n_j} \\ &= (1 - \omega)^{n_k} \omega^{n - n_k} \prod_{j \in A_s; j \neq k}^{\infty} \left(\frac{\pi_{PS}(j; \mu)}{1 - \pi_{PS}(k; \mu)} \right)^{n_j}. \end{aligned}$$

A sua função log-verossimilhança é dada por:

$$\begin{aligned} \ell(\mu, \omega; \mathcal{D}) &= n_k \log(1 - \omega) + (n - n_k) \log(\omega) + \sum_{j \in A_s; j \neq k}^{\infty} n_j \log \left(\frac{\pi_{PS}(j; \mu)}{1 - \pi_{PS}(k; \mu)} \right) \\ &= \ell_1(\mu; \mathcal{D}) + \ell_2(\omega; \mathcal{D}), \end{aligned}$$

em que,

$$\begin{aligned} \ell_1(\mu; \mathcal{D}) &= \sum_{j \in A_s; j \neq k}^{\infty} n_j \log \left(\frac{\pi_{PS}(j; \mu)}{1 - \pi_{PS}(k; \mu)} \right), \\ \ell_2(\omega; \mathcal{D}) &= n_k \log(1 - \omega) + (n - n_k) \log(\omega). \end{aligned}$$

Observando a equação acima verificamos que o seu segundo termo, $\ell_2(\omega; \mathcal{D})$, é independente de μ , portanto, ω e μ são ortogonais. Os elementos do vetor escore, $U = (U_\omega, U_\mu)^T$, são obtidos a partir da primeira derivada da função log-verossimilhança e seus elementos são dados por:

$$\begin{aligned} U_\omega &= \frac{\partial \ell_2(\omega; \mathcal{D})}{\partial \omega} = \frac{-n_k}{1-\omega} + \frac{n-n_k}{\omega}, \\ U_\mu &= \frac{\partial \ell_1(\mu; \mathcal{D})}{\partial \mu} = \frac{\partial}{\partial \mu} \left[\sum_{j \in A_s; j \neq k}^{\infty} n_j \log(\pi_{PS}(j; \mu)) - (n-n_k) \log(1 - \pi_{PS}(k; \mu)) \right] \\ &= \sum_{j \in A_s; j \neq k}^{\infty} n_j \frac{\partial}{\partial \mu} (\log(\pi_{PS}(j; \mu))) - (n-n_k) \frac{\partial}{\partial \mu} (\log(1 - \pi_{PS}(k; \mu))) \\ &= \sum_{j \in A_s; j \neq k}^{\infty} \frac{n_j \frac{\partial}{\partial \mu} (\pi_{PS}(j; \mu))}{\pi_{PS}(j; \mu)} + \frac{(n-n_k) \frac{\partial}{\partial \mu} (\pi_{PS}(k; \mu))}{(1 - \pi_{PS}(k; \mu))}. \end{aligned}$$

A matriz de informação observada J é dada por:

$$J = - \begin{bmatrix} U_{\omega\omega} & 0 \\ 0 & U_{\mu\mu} \end{bmatrix},$$

cujos elementos são:

$$\begin{aligned} U_{\omega\omega} &= \frac{-n_k}{(1-\omega)^2} - \frac{(n-n_k)}{\omega^2}, \\ U_{\mu\mu} &= \tilde{U}_{\mu\mu}. \end{aligned}$$

Igualando a função score de cada parâmetro a zero, encontramos seus respectivos estimadores de máxima verossimilhança (EMVs). Entretanto, somente para o parâmetro ω conseguimos obter o EMV explicitamente, o qual é dado por:

$$\hat{\omega} = \frac{n-n_k}{n},$$

que corresponde a proporção de observações diferentes de k em \mathbf{y} .

4.2.3 Distribuição k -MPS versus Distribuição k -SPS

Teorema 4. Seja Y uma variável aleatória com distribuição k -MPS e $Y^{(-k)}$ uma outra variável aleatória com distribuição k -SPS associada à Y . Os seus EMVs para o parâmetro μ são iguais.

Prova. Considere \mathbf{y} um vetor de tamanho n das observações da variável aleatória Y com distribuição k -MPS. Considere $\mathbf{y}^{(-k)}$ um vetor formado pelas observações de \mathbf{y} diferentes de k .

Considere ainda que $n_{(-k)}$ seja o número de elementos de $\mathbf{y}^{(-k)}$ tal que $n_{(-k)} = \sum_{j \in A_s; j \neq k}^{\infty} n_j$, e que n_j é o número de observações j no vetor $\mathbf{y}^{(-k)}$, portanto, $n_{(-k)} = n - n_k$, em que n_k é o número de observações k presentes em \mathbf{y} .

Suponha que $Y^{(-k)}$ seja uma variável aleatória com distribuição k -SPS associada à distribuição de Y e que $\mathbf{y}^{(-k)}$ é o vetor de suas observações, então, $Y^{(-k)}$ têm função massa de probabilidade dada por:

$$\pi_{k\text{-SPS}}(\mathbf{y}^{(-k)}; \boldsymbol{\mu}) = \frac{\pi_{PS}(\mathbf{y}^{(-k)}; \boldsymbol{\mu})}{1 - \pi_{PS}(k; \boldsymbol{\mu})}, \quad \forall \mathbf{y}^{(-k)} \in A_{\{-k\}}.$$

Com isso sua função de verossimilhança é dada por:

$$\begin{aligned} L(\boldsymbol{\mu}; \mathbf{y}^{(-k)}) &= \prod_{i=1}^{n_{(-k)}} \left\{ \frac{\pi_{PS}(y_i^{(-k)}; \boldsymbol{\mu})}{1 - \pi_{PS}(k; \boldsymbol{\mu})} \right\} \\ &= \left(\frac{1}{1 - \pi_{PS}(k; \boldsymbol{\mu})} \right)^{n_{(-k)}} \prod_{i=1}^{n_{(-k)}} \left\{ \pi_{PS}(y_i^{(-k)}; \boldsymbol{\mu}) \right\} \\ &= \left(\frac{1}{1 - \pi_{PS}(k; \boldsymbol{\mu})} \right)^{n_{(-k)}} \prod_{j \in A_s; j \neq k}^{\infty} \left\{ \pi_{PS}(j; \boldsymbol{\mu}) \right\}^{n_j}. \end{aligned}$$

A função log-verossimilhança é dada por:

$$\ell(\boldsymbol{\mu}; \mathbf{y}^{(-k)}) = \sum_{j \in A_s; j \neq k}^{\infty} n_j \log(\pi_{PS}(j; \boldsymbol{\mu})) - n_{(-k)} \log(1 - \pi_{PS}(k; \boldsymbol{\mu})).$$

O vetor score $U^{(-k)}$ é obtido a partir da primeira de $\ell(\boldsymbol{\mu}; \mathbf{y}^{(-k)})$ em relação a $\boldsymbol{\mu}$.

$$U_{\boldsymbol{\mu}}^{(-k)} = \sum_{j \in A_s; j \neq k}^{\infty} \frac{n_j \frac{\partial}{\partial \boldsymbol{\mu}}(\pi_{PS}(j; \boldsymbol{\mu}))}{\pi_{PS}(j; \boldsymbol{\mu})} + \frac{(n_{(-k)}) \frac{\partial}{\partial \boldsymbol{\mu}}(\pi_{PS}(k; \boldsymbol{\mu}))}{(1 - \pi_{PS}(k; \boldsymbol{\mu}))},$$

como $n_{(-k)} = n - n_k$, então $U_{\boldsymbol{\mu}}^{(-k)} = \tilde{U}_{\boldsymbol{\mu}}$. Desta forma conclui-se que os EMVs de $\boldsymbol{\mu}$ de uma distribuição k -MPS podem ser obtidos a partir da sua distribuição k -SPS associada, baseada apenas nas observações diferentes de k presentes na amostra observada da variável aleatória Y . □

4.3 Critérios de Seleção de Modelos

Há na literatura inúmeros métodos para selecionar modelos ajustados a um certo conjunto de dados. Neste trabalho faremos uso dos conhecidos critérios para seleção de modelos AIC (*Akaike information criterion*), proposto por Akaike (1974) e BIC (*Bayesian Information Criterion*), proposto por (SCHWARZ *et al.*, 1978).

Para um número finito de modelos concorrentes ao ajuste de determinado conjunto de dados, a seleção do modelo é determinada com base no menor valor de AIC e BIC. Considera-se nesta tomada de decisão um certo equilíbrio entre a qualidade do ajuste do modelo estatístico, a função de verossimilhança e o número de parâmetros envolvidos. Neste caso, recomenda-se selecionar o modelo mais parcimonioso, pois para modelos com uma quantidade razoável de parâmetros, há a possibilidade do ajuste resultar em *overfitting*.

O critério AIC usa o conceito de teoria da informação, fornecendo uma medida relativa à informação perdida quando se é utilizado o modelo estatístico para descrever o conjunto de dados. Seu valor é obtido pela seguinte fórmula:

$$AIC = -2\ell(\hat{\mu}, \hat{p}; \mathcal{D}) + 2\lambda,$$

em que $\ell(\hat{\mu}, \hat{p}; \mathcal{D})$ é o valor do logaritmo da função de verossimilhança avaliada nas estimativas dos parâmetros, λ é o número de parâmetros desconhecidos e estimados do modelo.

Já o critério BIC, é mais rigoroso ao penalizar o modelo pela quantidade de parâmetros. Seu valor é obtido pela seguinte fórmula.

$$BIC = -2\ell(\hat{\mu}, \hat{p}; \mathcal{D}) + \lambda \ln(n),$$

em que n é o tamanho do vetor de observações.

ESTUDO DE SIMULAÇÃO

Neste Capítulo, é apresentado um estudo de simulação com dados gerados considerando o procedimento de máxima verossimilhança para obter as estimativas dos parâmetros μ e p ou ω da distribuição k -MPS.

Para o estudo de simulação consideramos três distribuições: k -MP, k -MG e k -MB. Para cada uma dessas distribuições geramos N amostras de tamanho n com diferentes pontos de modificação e diferentes valores do parâmetro μ e apenas dois valores para o parâmetro de modificação p . Feito isto, realizamos o ajuste das respectivas distribuições para cada uma das suas N amostras correspondentes. Com isso, obtivemos N estimativas para cada um dos parâmetros μ e p , bem como os seus respectivos intervalos assintóticos e *bootstrap* com 95% de confiança. Utilizando cada uma das N estimativas dos parâmetros μ e p de cada distribuição, realizamos o procedimento de *bootstrap* paramétrico considerando B réplicas, do qual obtivemos o intervalo de *bootstrap* com 95% de confiança. De posse desses resultados, calculamos as estimativas de Monte Carlo para os parâmetros.

O desempenho dos métodos assintótico e de *bootstrap* para cada parâmetro foi avaliado através de medidas como a probabilidade de cobertura (PC). Já a eficiência de cada estimador foi avaliada através do vício (\mathcal{B}) e da razão entre o EQM e Var. Tais medidas, foram obtidas considerando o método de Monte Carlo. Pela teoria assintótica, à medida que aumenta-se o tamanho do conjunto de dados, espera-se que os estimadores sejam assintoticamente não viciados e, conseqüentemente, a razão entre o EQM e Var se aproxime de um e o vício (\mathcal{B}) se aproxime de zero.

A seguir, apresentaremos as equações utilizadas para a obtenção das estimativas de Monte Carlo e das medidas de eficiência dos estimadores:

$$EQM(\hat{\mu}) = \frac{1}{N} \sum_{i=1}^N (\hat{\mu}_i - \mu)^2, \quad EQM(\hat{p}) = \frac{1}{N} \sum_{i=1}^N (\hat{p}_i - p)^2,$$

$$Var(\hat{\mu}) = \frac{1}{N-1} \sum_{i=1}^N (\hat{\mu}_i - \bar{\mu})^2, \quad Var(\hat{p}) = \frac{1}{N-1} \sum_{i=1}^N (\hat{p}_i - \bar{p})^2.$$

Em que

$$\bar{\mu} = \frac{1}{N} \sum_{i=1}^N \hat{\mu}_i, \quad \bar{p} = \frac{1}{N} \sum_{i=1}^N \hat{p}_i,$$

$$\mathcal{B}(\hat{\mu}) = \frac{1}{N} \sum_{i=1}^N |\hat{\mu}_i - \mu|, \quad \mathcal{B}(\hat{p}) = \frac{1}{N} \sum_{i=1}^N |\hat{p}_i - p|.$$

Neste estudo de simulação geramos $N = 500$ amostras com tamanhos $n = 200$ e $n = 500$. Para o procedimento de *bootstrap* consideramos $B = 5000$ réplicas.

Na Tabela 3 apresentamos os valores para o parâmetro de modificação p , o ponto de modificação k e os valores do parâmetro μ para as distribuições consideradas neste estudo, além do valor de m utilizado na distribuição k -MB.

Tabela 3 – Valores dos parâmetros das distribuições k -MPS utilizados no estudo de simulação.

Distribuição	p	k	μ	m
k -MP	0,7	0	0,5; 2 e 5	
		2		
	1,2	5		
		0	0,5; 1 e 1,5	
1				
k -MG	0,7	0	0,5; 2 e 5	
		2		
	1,2	5		
		0	0,5; 1 e 1,5	
1				
k -MB	0,7	0	0,5; 2 e 5	10
		2		
	1,2	5		
		0	0,5; 1 e 1,5	
1				

5.1 Distribuição Poisson k -Modificada

Na Tabela 4 temos as probabilidades de cobertura dos intervalos com 95% de confiança. De forma geral, os resultados obtidos das probabilidades de cobertura para os parâmetros μ e p pelos métodos assintótico e de *bootstrap* ficaram próximos ao nível nominal com 95% de confiança, tanto para as amostras de tamanho 200 quanto para as amostras de tamanho 500.

Tabela 4 – Probabilidade de cobertura das estimativas dos parâmetros μ e p da distribuição k -MP.

n	k	μ	PC μ		p	PC p			
			Assint.	Boot		Assint.	Boot		
200	0	0,50	93,60	93,80	0,7	93,80	92,20		
		2,00	96,40	96,20		93,40	94,80		
		5,00	93,40	93,80		93,60	94,60		
	2	0,50	93,40	94,40		93,60	96,20		
		2,00	95,40	95,00		95,60	95,40		
		5,00	94,60	95,00		94,40	95,00		
	5	0,50	95,20	95,20		95,00	95,00		
		2,00	94,80	93,60		95,00	95,40		
		5,00	95,00	94,20		96,00	96,60		
	0	0,50	99,00	99,20		1,2	97,40	99,40	
			1,00	97,20			97,40	94,60	98,80
			1,50	98,60			98,60	95,20	99,20
1		0,50	97,40	97,40	96,80		98,40		
		1,00	96,60	96,20	95,40		94,80		
		1,50	95,80	95,60	97,00		98,00		
500	0	0,50	96,80	97,00	0,7	95,40	97,00		
		2,00	96,00	96,00		93,20	95,00		
		5,00	94,60	94,60		95,80	95,60		
	2	0,50	96,40	96,20		93,80	95,20		
		2,00	95,40	95,40		94,20	93,60		
		5,00	94,40	94,60		95,60	96,00		
	5	0,50	96,20	95,60		94,20	94,20		
		2,00	94,60	94,60		95,40	95,60		
		5,00	93,60	93,80		93,20	94,20		
	0	0,50	98,80	99,00		1,2	93,20	99,20	
		1,00	98,60	98,40			95,40	99,80	
		1,50	97,80	97,80			93,80	98,80	
1	0,50	96,80	96,40	97,20	98,40				
	1,00	94,80	94,80	94,20	94,00				
	1,50	96,00	96,20	97,00	97,80				

Na Tabela 5 estão dispostos os resultados das médias das estimativas pontuais para os parâmetros μ e p obtidas por máxima verossimilhança, além das médias das estimativas para os limites inferior e superior dos intervalos de confiança obtidos pelos métodos assintótico e de *bootstrap*. Notamos que em todos os casos, o verdadeiro valor do parâmetro está contido no intervalo médio.

Tabela 5 – Média das estimativas dos parâmetros μ e p e dos respectivos intervalos com 95% de confiança da distribuição k -MP.

n	k	μ	$\bar{\mu}$	IC μ (95%)		p	\bar{p}	IC p (95%)		
				Assint.	Boot.			Assint.	Boot.	
200	0	0,5	0,50	(0,26 ; 0,74)	(0,26 ; 0,75)	0,70	0,72	(0,58 ; 0,92)	(0,49 ; 1,42)	
		2,0	2,00	(1,72 ; 2,28)	(1,7 ; 2,24)		0,70	(0,62 ; 0,78)	(0,62 ; 0,79)	
		5,0	4,99	(4,62 ; 5,37)	(4,62 ; 5,37)		0,70	(0,63 ; 0,76)	(0,63 ; 0,76)	
	2	0,5	0,50	(0,36 ; 0,65)	(0,37 ; 0,66)		0,70	(0,63 ; 0,77)	(0,63 ; 0,78)	
		2,0	2,01	(1,77 ; 2,24)	(1,76 ; 2,23)		0,70	(0,60 ; 0,79)	(0,60 ; 0,79)	
		5,0	4,99	(4,58 ; 5,39)	(4,59 ; 5,40)		0,70	(0,63 ; 0,77)	(0,63 ; 0,78)	
	5	0,5	0,50	(0,38 ; 0,61)	(0,39 ; 0,62)		0,70	(0,64 ; 0,76)	(0,64 ; 0,76)	
		2,0	2,01	(1,75 ; 2,27)	(1,76 ; 2,27)		0,70	(0,64 ; 0,77)	(0,63 ; 0,77)	
		5,0	5,01	(4,63 ; 5,38)	(4,64 ; 5,38)		0,70	(0,62 ; 0,78)	(0,62 ; 0,78)	
	0	0,50	0,50	(0,31 ; 0,68)	(0,32 ; 0,69)		1,2	1,22	(1,04 ; 1,39)	(0,90 ; 1,80)
		1,00	1,00	(0,80 ; 1,19)	(0,80 ; 1,19)			1,20	(1,11 ; 1,30)	(1,06 ; 1,40)
		1,50	1,51	(1,30 ; 1,71)	(1,30 ; 1,71)			1,20	(1,15 ; 1,24)	(1,12 ; 1,29)
0,50		0,50	(0,40 ; 0,60)	(0,39 ; 0,60)	1,20	(1,13 ; 1,27)		(1,10 ; 1,29)		
1,00		1,00	(0,87 ; 1,13)	(0,87 ; 1,13)	1,20	(1,10 ; 1,29)		(1,10 ; 1,29)		
1,50		1,50	(1,34 ; 1,66)	(1,34 ; 1,67)	1,20	(1,12 ; 1,28)		(1,11 ; 1,29)		
500	0	0,50	0,51	(0,35 ; 0,67)	(0,39 ; 0,67)	0,70	0,69	(0,60 ; 0,79)	(0,53 ; 0,85)	
		2,00	2,00	(1,82 ; 2,18)	(1,82 ; 2,18)		0,70	(0,65 ; 0,75)	(0,65 ; 0,75)	
		5,00	4,99	(4,75 ; 5,23)	(4,76 ; 5,23)		0,70	(0,66 ; 0,74)	(0,66 ; 0,74)	
	2	0,50	0,50	(0,41 ; 0,59)	(0,41 ; 0,59)		0,70	(0,66 ; 0,75)	(0,65 ; 0,75)	
		2,00	2,00	(1,85 ; 2,14)	(1,85 ; 2,15)		0,70	(0,64 ; 0,76)	(0,64 ; 0,76)	
		5,00	5,00	(4,74 ; 5,26)	(4,75 ; 5,26)		0,70	(0,65 ; 0,74)	(0,65 ; 0,74)	
	5	0,50	0,50	(0,43 ; 0,58)	(0,43 ; 0,58)		0,70	(0,66 ; 0,74)	(0,66 ; 0,74)	
		2,00	2,00	(1,84 ; 2,16)	(1,84 ; 2,16)		0,70	(0,66 ; 0,74)	(0,66 ; 0,74)	
		5,00	4,99	(4,76 ; 5,23)	(4,76 ; 5,23)		0,70	(0,65 ; 0,75)	(0,65 ; 0,75)	
	0	0,50	0,50	(0,38 ; 0,62)	(0,38 ; 0,62)		1,2	1,21	(1,09 ; 1,32)	(1,00 ; 1,52)
		1,00	1,00	(0,88 ; 1,13)	(0,88 ; 1,13)			1,20	(1,14 ; 1,26)	(1,10 ; 1,31)
		1,50	1,50	(1,37 ; 1,63)	(1,37 ; 1,63)			1,20	(1,17 ; 1,23)	(1,15 ; 1,26)
0,50	0,50	(0,44 ; 0,56)	(0,43 ; 0,56)	1,20	(1,15 ; 1,25)	(1,14 ; 1,25)				
1,00	1,00	(0,92 ; 1,08)	(0,92 ; 1,08)	1,20	(1,14 ; 1,26)	(1,14 ; 1,26)				
1,50	1,50	(1,40 ; 1,60)	(1,40 ; 1,60)	1,20	(1,15 ; 1,25)	(1,14 ; 1,26)				

Na Tabela 6 são apresentados os resultados das medidas de eficiência calculadas para os estimadores dos parâmetros μ e p . Observando os resultados da razão $\frac{EOM}{VAR}$ e do vício (\mathcal{B}), notamos respectivamente, uma coluna com 1's e outra coluna com valores próximos de zero, indicando que os estimadores $\hat{\mu}$ e \hat{p} são assintoticamente não viciados.

Tabela 6 – Resultados do desempenho dos estimadores $\hat{\mu}$ e \hat{p} para distribuição k -MP.

n	k	μ	$\hat{\mu}$		p	\hat{p}			
			$\frac{EQM(\hat{\mu})}{Var(\hat{\mu})}$	$\mathcal{B}(\hat{\mu})$		$\frac{EQM(\hat{p})}{Var(\hat{p})}$	$\mathcal{B}(\hat{p})$		
200	0	0,50	1,00	$1,0 \times 10^{-1}$	0,70	1,05	$1,5 \times 10^{-1}$		
		2,00	1,00	$1,1 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
		5,00	1,00	$1,6 \times 10^{-1}$		1,02	$3,0 \times 10^{-2}$		
	2	0,50	1,00	$6,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$		
		2,00	1,00	$1,0 \times 10^{-1}$		1,00	$4,0 \times 10^{-2}$		
		5,00	1,00	$1,7 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
	5	0,50	1,00	$5,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$		
		2,00	1,00	$1,1 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
		5,00	1,00	$1,4 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
	0	0,50	1,00	1,00		$5,0 \times 10^{-2}$	1,20	1,02	$9,0 \times 10^{-2}$
			1,00	1,00		$6,0 \times 10^{-2}$		1,00	$5,0 \times 10^{-2}$
			1,50	1,00		$7,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$
1		0,50	1,00	$4,0 \times 10^{-2}$	1,00	$3,0 \times 10^{-2}$			
		1,00	1,00	$5,0 \times 10^{-2}$	1,00	$4,0 \times 10^{-2}$			
		1,50	1,00	$6,0 \times 10^{-2}$	1,00	$3,0 \times 10^{-2}$			
500	0	0,50	1,02	$6,0 \times 10^{-2}$	0,70	1,00	$7,0 \times 10^{-2}$		
		2,00	1,00	$7,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$		
		5,00	1,00	$1,0 \times 10^{-1}$		1,01	$2,0 \times 10^{-2}$		
	2	0,50	1,00	$4,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$		
		2,00	1,00	$3,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$		
		5,00	1,00	$1,1 \times 10^{-1}$		1,01	$2,0 \times 10^{-2}$		
	5	0,50	1,00	$3,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$		
		2,00	1,00	$6,0 \times 10^{-2}$		1,01	$2,0 \times 10^{-2}$		
		5,00	1,00	$9,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$		
	0	0,50	1,00	1,00		$3,0 \times 10^{-2}$	1,20	1,01	$6,0 \times 10^{-2}$
			1,00	1,00		$4,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$
			1,50	1,00		$4,0 \times 10^{-2}$		1,00	$1,0 \times 10^{-2}$
1		0,50	1,00	$2,0 \times 10^{-2}$	1,00	$2,0 \times 10^{-2}$			
		1,00	1,00	$3,0 \times 10^{-2}$	1,00	$2,0 \times 10^{-2}$			
		1,50	1,00	$4,0 \times 10^{-2}$	1,00	$2,0 \times 10^{-2}$			

5.2 Distribuição Geométrica k -Modificada

No caso da distribuição k -MG geramos conjuntos de dados considerando os valores dos parâmetros apresentados na Tabela 3, dos quais para cada parâmetro μ e p obtivemos as estimativas de máxima verossimilhança, seus respectivos intervalos de confiança assintótico e de *bootstrap*, ambos com 95% de confiança, as probabilidades de coberturas e as medidas de eficiência. Os resultados deste estudo de simulação são apresentados nas Tabelas 7 - 9.

Na Tabela 7 temos as probabilidades de cobertura dos intervalos com 95% de confiança. De forma geral, os resultados obtidos das probabilidades de cobertura para os parâmetros μ e p pelos métodos assintótico e de *bootstrap* ficaram próximos ao nível nominal com 95% de

confiança, tanto para as amostras de tamanho 200 quanto para as amostras de tamanho 500, indicando o bom desempenho deste método de estimação.

Tabela 7 – Probabilidade de cobertura das estimativas dos parâmetros μ e p da distribuição k -MG.

n	k	μ	PC μ		p	PC p			
			Assint.	Boot.		Assint.	Boot.		
200	0	0,50	99,40	99,00	0,7	92,20	97,20		
		2,00	95,40	95,00		92,80	94,00		
		5,00	93,80	94,80		93,20	94,40		
	2	0,50	94,00	94,00		92,20	93,20		
		2,00	95,00	94,00		97,60	97,40		
		5,00	93,80	94,20		95,40	95,60		
	5	0,50	95,60	96,80		92,40	92,00		
		2,00	93,00	93,20		95,80	95,60		
		5,00	96,80	97,00		93,40	92,80		
	0	0,50	97,80	98,20		1,2	95,00	99,60	
			1,00	96,60			97,00	98,00	99,20
			1,50	96,20			96,60	94,00	98,20
1		0,50	94,60	95,00	97,20		98,20		
		1,00	95,00	94,60	94,80		94,80		
		1,50	93,40	93,60	96,20		96,40		
500	0	0,50	95,80	96,40	0,7	95,60	97,40		
		2,00	94,20	94,00		93,20	94,20		
		5,00	95,40	95,40		95,00	95,80		
	2	0,50	95,00	95,80		95,40	96,40		
		2,00	93,60	94,20		94,40	94,40		
		5,00	94,60	94,20		94,60	94,40		
	5	0,50	96,20	96,00		96,00	96,20		
		2,00	96,40	96,60		95,40	96,00		
		5,00	94,00	94,00		94,00	94,00		
	0	0,50	98,00	98,00		1,2	95,80	99,00	
			1,00	97,00			96,80	96,60	98,60
			1,50	96,00			96,20	93,20	98,20
1		0,50	96,80	96,40	97,40		98,80		
		1,00	94,80	94,60	93,20		94,60		
		1,50	95,80	95,60	97,20		97,40		

Na Tabela 8 são apresentados os resultados das médias das estimativas pontuais obtidas por máxima verossimilhança para os parâmetros μ e p , além das médias dos limites inferior e superior dos intervalos de confiança obtidos pelos métodos assintótico e de *bootstrap*. Analisando esta Tabela, podemos notar que, em todos os casos, os valores das médias das estimativas pontuais estão próximos dos verdadeiros valores dos parâmetros μ e p , além das médias das estimativas dos intervalos com 95% confiança incluem os verdadeiros valores dos parâmetros.

Tabela 8 – Média das estimativas dos parâmetros μ e p e dos respectivos intervalos com 95% de confiança da distribuição k -MG.

n	k	μ	$\bar{\mu}$	IC μ (95%)		p	\bar{p}	IC p (95%)			
				Assint.	Boot.			Assint.	Boot.		
200	0	0,50	0,53	(0,27 ; 0,79)	(0,38 ; 0,82)	0,70	0,69	(0,52 ; 0,87)	(0,46 ; 0,95)		
		2,00	2,02	(1,52 ; 2,52)	(1,54 ; 2,52)		0,70	(0,60 ; 0,80)	(0,59 ; 0,83)		
		5,00	4,97	(3,98 ; 5,96)	(4,02 ; 6,00)		0,70	(0,62 ; 0,78)	(0,62 ; 0,79)		
	2	0,50	0,50	(0,33 ; 0,66)	(0,34 ; 0,66)		0,70	(0,63 ; 0,77)	(0,62 ; 0,77)		
		2,00	2,00	(1,59 ; 2,41)	(1,59 ; 2,34)		0,70	(0,62 ; 0,78)	(0,62 ; 0,78)		
		5,00	5,01	(4,08 ; 5,93)	(4,12 ; 5,98)		0,70	(0,63 ; 0,78)	(0,63 ; 0,78)		
	5	0,50	0,51	(0,36 ; 0,66)	(0,40 ; 0,67)		0,70	(0,64 ; 0,76)	(0,64 ; 0,76)		
		2,00	1,98	(1,57 ; 2,40)	(1,58 ; 2,41)		0,70	(0,63 ; 0,77)	(0,63 ; 0,77)		
		5,00	4,99	(4,09 ; 5,90)	(4,12 ; 5,94)		0,70	(0,63 ; 0,77)	(0,63 ; 0,77)		
	0	0,50	0,50	0,50	(0,31 ; 0,69)		(0,32 ; 0,70)	1,20	1,22	(1,01 ; 1,42)	(0,91 ; 1,73)
			1,00	1,01	(0,75 ; 1,26)		(0,76 ; 1,27)		1,20	(1,06 ; 1,33)	(1,01 ; 1,44)
			1,50	1,51	(1,19 ; 1,83)		(1,21 ; 1,84)		1,20	(1,10 ; 1,31)	(1,07 ; 1,36)
1		0,50	0,50	(0,38 ; 0,61)	(0,38 ; 0,61)	1,20	(1,15 ; 1,24)		(1,14 ; 1,25)		
		1,00	1,00	(0,82 ; 1,18)	(0,83 ; 1,18)	1,20	(1,14 ; 1,25)		(1,14 ; 1,25)		
		1,50	1,51	(1,26 ; 1,76)	(1,27 ; 1,77)	1,20	(1,15 ; 1,25)		(1,14 ; 1,25)		
500	0	0,50	0,51	(0,35 ; 0,67)	(0,38 ; 0,68)	0,70	0,71	(0,60 ; 0,82)	(0,54 ; 0,85)		
		2,00	2,01	(1,69 ; 2,32)	(1,70 ; 2,33)		0,70	(0,63 ; 0,77)	(0,63 ; 0,78)		
		5,00	5,01	(4,38 ; 5,64)	(4,40 ; 5,66)		0,70	(0,65 ; 0,75)	(0,65 ; 0,75)		
	2	0,50	0,50	(0,39 ; 0,60)	(0,39 ; 0,60)		0,70	(0,65 ; 0,74)	(0,65 ; 0,74)		
		2,00	1,99	(1,74 ; 2,25)	(1,74 ; 2,26)		0,70	(0,65 ; 0,75)	(0,65 ; 0,75)		
		5,00	5,01	(4,42 ; 5,59)	(4,44 ; 5,61)		0,70	(0,65 ; 0,75)	(0,65 ; 0,75)		
	5	0,50	0,50	(0,41 ; 0,60)	(0,41 ; 0,60)		0,70	(0,66 ; 0,74)	(0,66 ; 0,74)		
		2,00	2,00	(1,73 ; 2,27)	(1,74 ; 2,27)		0,70	(0,66 ; 0,74)	(0,66 ; 0,74)		
		5,00	4,98	(4,41 ; 5,56)	(4,42 ; 5,57)		0,70	(0,65 ; 0,74)	(0,65 ; 0,74)		
	0	0,50	0,50	0,50	(0,38 ; 0,62)		(0,39 ; 0,63)	1,20	1,20	(1,08 ; 1,33)	(1,00 ; 1,48)
			1,00	1,00	(0,84 ; 1,16)		(0,84 ; 1,16)		1,20	(1,12 ; 1,29)	(1,08 ; 1,34)
			1,50	1,51	(1,30 ; 1,71)		(1,31 ; 1,71)		1,20	(1,13 ; 1,26)	(1,11 ; 1,30)
1		0,50	0,50	(0,43 ; 0,57)	(0,43 ; 0,57)	1,20	(1,17 ; 1,23)		(1,17 ; 1,23)		
		1,00	1,00	(0,89 ; 1,11)	(0,89 ; 1,11)	1,20	(1,16 ; 1,23)		(1,16 ; 1,23)		
		1,50	1,50	(1,35 ; 1,66)	(1,35 ; 1,66)	1,20	(1,17 ; 1,23)		(1,16 ; 1,23)		

Na Tabela 9 são apresentados os resultados das medidas de eficiência calculadas para os estimadores dos parâmetros μ e p . Observando os resultados da razão $\frac{EOM}{VAR}$ e do vício (\mathcal{B}), notamos respectivamente, uma coluna com 1's e outra coluna com valores próximos de zero, indicando que os estimadores $\hat{\mu}$ e \hat{p} são assintoticamente não viciados.

Tabela 9 – Resultados do desempenho dos estimadores $\hat{\mu}$ e \hat{p} para distribuição k -MG.

n	k	μ	$\hat{\mu}$		p	\hat{p}			
			$\frac{EQM(\hat{\mu})}{Var(\hat{\mu})}$	$\mathcal{B}(\hat{\mu})$		$\frac{EQM(\hat{p})}{Var(\hat{p})}$	$\mathcal{B}(\hat{p})$		
200	0	0,50	1,07	$9,0 \times 10^{-2}$	0,70	1,00	$1,1 \times 10^{-1}$		
		2,00	1,00	$2,1 \times 10^{-1}$		1,00	$5,0 \times 10^{-2}$		
		5,00	1,00	$4,1 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
	2	0,50	1,00	$7,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$		
		2,00	1,00	$1,6 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
		5,00	1,00	$3,7 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
	5	0,50	1,01	$6,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$		
		2,00	1,00	$1,8 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
		5,00	1,00	$3,5 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
	1,20	0	0,50	1,00		$6,0 \times 10^{-2}$	1,20	1,01	$1,0 \times 10^{-1}$
			1,00	1,00		$9,0 \times 10^{-2}$		1,00	$6,0 \times 10^{-2}$
			1,50	1,00		$1,2 \times 10^{-1}$		1,00	$5,0 \times 10^{-2}$
1		0,50	1,00	$4,0 \times 10^{-2}$	1,00	$2,0 \times 10^{-2}$			
		1,00	1,00	$7,0 \times 10^{-2}$	1,00	$2,0 \times 10^{-2}$			
		1,50	1,00	$1,0 \times 10^{-1}$	1,00	$2,0 \times 10^{-2}$			
500	0	0,50	1,00	$7,0 \times 10^{-2}$	0,70	1,01	$8,0 \times 10^{-2}$		
		2,00	1,00	$1,3 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$		
		5,00	1,00	$2,5 \times 10^{-1}$		1,00	$2,0 \times 10^{-2}$		
	2	0,50	1,00	$4,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$		
		2,00	1,00	$1,0 \times 10^{-1}$		1,01	$2,0 \times 10^{-2}$		
		5,00	1,00	$2,3 \times 10^{-1}$		1,00	$2,0 \times 10^{-2}$		
	5	0,50	1,00	$4,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$		
		2,00	1,00	$1,1 \times 10^{-1}$		1,00	$2,0 \times 10^{-2}$		
		5,00	1,00	$2,4 \times 10^{-1}$		1,00	$2,0 \times 10^{-2}$		
	1,20	0	0,50	1,00		$4,0 \times 10^{-2}$	1,20	1,00	$6,0 \times 10^{-2}$
			1,00	1,00		$6,0 \times 10^{-2}$		1,00	$4,0 \times 10^{-2}$
			1,50	1,00		$8,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$
1		0,50	1,00	$3,0 \times 10^{-2}$	1,00	$1,0 \times 10^{-2}$			
		1,00	1,00	$5,0 \times 10^{-2}$	1,00	$1,0 \times 10^{-2}$			
		1,50	1,00	$6,0 \times 10^{-2}$	1,00	$1,0 \times 10^{-2}$			

5.3 Distribuição Binomial k -Modificada

No caso da distribuição k -MB, geramos conjunto de dados considerando os valores presentes na Tabela 3, dos quais para cada parâmetro μ e p obtivemos as estimativas de máxima verossimilhança, seus respectivos intervalos de confiança assintótico e de bootstrap, ambos com 95% de confiança, as probabilidades de coberturas e as medidas de eficiência. Os resultados deste estudo de simulação são apresentados nas Tabelas 10 - 12.

Na Tabela 10 temos as probabilidades de cobertura dos intervalos com 95% de confiança. De forma geral, os resultados obtidos das probabilidades de cobertura para os parâmetros μ e p pelos métodos assintótico e de *bootstrap* ficaram próximos ao nível nominal com 95% de

confiança, tanto para as amostras de tamanho 200 quanto para as amostras de tamanho 500. Estes resultados indicam o bom desempenho que estes métodos tiveram nas estimativas de seus intervalos de confiança.

Tabela 10 – Probabilidade de cobertura das estimativas dos parâmetros μ e p da distribuição k -MB.

n	k	μ	PC_{μ}		p	PC_p		
			Assint.	Boot		Assint.	Boot	
200	0	0,50	93,40	93,60	0,7	92,20	92,00	
		2,00	96,80	96,80		93,00	94,00	
		5,00	95,40	94,60		94,20	94,20	
	2	0,50	93,20	94,40		93,60	95,00	
		2,00	93,40	92,80		94,80	95,20	
		5,00	96,80	96,80		95,40	95,60	
	5	0,50	95,00	96,00		94,80	95,00	
		2,00	96,20	96,40		94,20	94,40	
		5,00	95,80	95,80		96,20	95,80	
	0	0,50	99,00	99,00	1,2	96,60	99,60	
			1,00	99,40		99,20	94,20	98,80
			1,50	99,00		99,40	98,00	98,60
		1	0,50	96,80		96,20	96,80	99,20
			1,00	93,00		92,60	96,00	96,00
			1,50	95,40		95,40	98,00	98,80
500	0	0,50	98,00	98,00	0,7	95,80	97,80	
		2,00	95,80	96,00		94,80	96,00	
		5,00	97,00	97,00		95,00	95,60	
	2	0,50	95,20	95,80		92,20	94,00	
		2,00	94,20	93,80		94,60	94,40	
		5,00	96,60	96,80		93,60	94,40	
	5	0,50	94,20	94,00		96,60	96,00	
		2,00	96,40	96,20		94,40	95,00	
		5,00	95,80	95,40		93,60	93,00	
	0	0,50	98,60	98,80	1,2	93,80	99,20	
			1,00	97,60		97,60	93,20	98,20
			1,50	99,60		99,60	97,40	99,80
		1	0,50	96,60		96,60	96,60	98,60
			1,00	94,80		93,80	94,40	94,60
			1,50	94,80		95,40	97,20	98,00

Na Tabela 11 são apresentados os resultados das médias das estimativas pontuais obtidas por máxima verossimilhança para os parâmetros μ e p , além das médias dos limites inferior e superior dos intervalos de confiança obtidos pelos métodos assintótico e de *bootstrap*. Analisando esta Tabela, podemos notar que, em todos os casos, os valores das médias das estimativas pontuais estão próximos dos verdadeiros valores dos parâmetros μ e p , além das médias das estimativas dos intervalos com 95% confiança incluírem os verdadeiros valores dos parâmetros.

Tabela 11 – Média das estimativas dos parâmetros μ e p e dos respectivos intervalos com 95% de confiança da distribuição k -MB.

n	k	μ	$\bar{\mu}$	IC μ (95%)		p	\bar{p}	IC p (95%)		
				Assint.	Boot.			Assint.	Boot.	
200	0	0,50	0,50	(0,25 ; 0,74)	(0,26 ; 0,75)	0,70	0,72	(0,59 ; 0,92)	(0,49 ; 1,45)	
		2,00	2,01	(1,75 ; 2,26)	(1,73 ; 2,22)		0,70	(0,62 ; 0,77)	(0,62 ; 0,79)	
		5,00	4,99	(4,73 ; 5,26)	(4,73 ; 5,25)		0,70	(0,64 ; 0,76)	(0,64 ; 0,76)	
	2	0,50	0,50	(0,36 ; 0,65)	(0,37 ; 0,66)		0,70	0,70	(0,63 ; 0,77)	(0,63 ; 0,78)
		2,00	2,00	(1,79 ; 2,22)	(1,79 ; 2,21)		0,70	0,70	(0,60 ; 0,80)	(0,60 ; 0,80)
		5,00	5,00	(4,72 ; 5,29)	(4,71 ; 5,29)		0,70	0,70	(0,63 ; 0,77)	(0,63 ; 0,77)
	5	0,50	0,50	(0,38 ; 0,61)	(0,39 ; 0,62)		0,70	0,70	(0,64 ; 0,76)	(0,63 ; 0,76)
		2,00	2,00	(1,77 ; 2,23)	(1,78 ; 2,23)		0,70	0,70	(0,63 ; 0,77)	(0,63 ; 0,77)
		5,00	5,00	(4,74 ; 5,26)	(4,74 ; 5,26)		0,70	0,70	(0,61 ; 0,80)	(0,61 ; 0,79)
	0	0,50	0,51	(0,32 ; 0,70)	(0,32 ; 0,70)		1,20	1,20	(1,02 ; 1,37)	(0,89 ; 1,78)
		1,00	1,00	(0,81 ; 1,19)	(0,81 ; 1,19)			1,20	(1,11 ; 1,29)	(1,06 ; 1,39)
		1,50	1,50	(1,31 ; 1,69)	(1,31 ; 1,69)			1,20	(1,17 ; 1,23)	(1,13 ; 1,28)
0,50		0,50	(0,40 ; 0,60)	(0,39 ; 0,60)	1,20	(1,12 ; 1,28)		(1,10 ; 1,29)		
1,00		1,00	(0,88 ; 1,12)	(0,88 ; 1,12)	1,20	(1,10 ; 1,30)		(1,10 ; 1,30)		
1,50		1,50	(1,35 ; 1,66)	(1,36 ; 1,66)	1,19	(1,11 ; 1,28)		(1,10 ; 1,29)		
500	0	0,50	0,52	(0,36 ; 0,67)	(0,39 ; 0,68)	0,70	0,69	(0,59 ; 0,78)	(0,52 ; 0,85)	
		2,00	2,00	(1,84 ; 2,15)	(1,84 ; 2,15)		0,70	(0,65 ; 0,75)	(0,65 ; 0,75)	
		5,00	5,00	(4,83 ; 5,16)	(4,83 ; 5,16)		0,70	(0,66 ; 0,74)	(0,66 ; 0,74)	
	2	0,50	0,50	(0,41 ; 0,59)	(0,41 ; 0,59)		0,70	0,70	(0,65 ; 0,74)	(0,65 ; 0,75)
		2,00	2,00	(1,87 ; 2,13)	(1,87 ; 2,13)		0,70	0,70	(0,64 ; 0,76)	(0,64 ; 0,76)
		5,00	4,99	(4,81 ; 5,18)	(4,81 ; 5,17)		0,70	0,70	(0,66 ; 0,74)	(0,66 ; 0,74)
	5	0,50	0,50	(0,43 ; 0,57)	(0,43 ; 0,57)		0,70	0,70	(0,66 ; 0,74)	(0,66 ; 0,74)
		2,00	2,01	(1,86 ; 2,15)	(1,86 ; 2,15)		0,70	0,70	(0,66 ; 0,74)	(0,66 ; 0,74)
		5,00	5,00	(4,84 ; 5,17)	(4,84 ; 5,17)		0,70	0,70	(0,64 ; 0,76)	(0,64 ; 0,76)
	0	0,50	0,50	(0,38 ; 0,62)	(0,38 ; 0,62)		1,20	1,20	(1,10 ; 1,31)	(0,99 ; 1,52)
		1,00	1,00	(0,88 ; 1,12)	(0,88 ; 1,12)			1,20	(1,14 ; 1,26)	(1,11 ; 1,31)
		1,50	1,50	(1,38 ; 1,62)	(1,38 ; 1,62)			1,20	(1,18 ; 1,22)	(1,16 ; 1,25)
0,50	0,50	(0,44 ; 0,56)	(0,43 ; 0,56)	1,20	(1,15 ; 1,25)	(1,13 ; 1,26)				
1,00	1,00	(0,92 ; 1,07)	(0,92 ; 1,07)	1,20	(1,14 ; 1,26)	(1,13 ; 1,26)				
1,50	1,50	(1,40 ; 1,59)	(1,40 ; 1,59)	1,20	(1,15 ; 1,26)	(1,14 ; 1,26)				

Na Tabela 12 são apresentados os resultados das medidas de eficiência calculadas para os estimadores dos parâmetros μ e p . Observando os resultados da razão $\frac{EOM}{VAR}$ e do vício (\mathcal{B}), notamos respectivamente, uma coluna com 1's e outra coluna com valores próximos de zero, indicando que os estimadores $\hat{\mu}$ e \hat{p} são assintoticamente não viciados.

Tabela 12 – Resultados do desempenho dos estimadores $\hat{\mu}$ e \hat{p} para distribuição k -MB.

n	k	μ	$\hat{\mu}$		p	\hat{p}		
			$\frac{EQM(\hat{\mu})}{Var(\hat{\mu})}$	$\mathcal{B}(\hat{\mu})$		$\frac{EQM(\hat{p})}{Var(\hat{p})}$	$\mathcal{B}(\hat{p})$	
200	0	0,50	1,00	$1,0 \times 10^{-1}$	0,70	1,05	$1,5 \times 10^{-1}$	
		2,00	1,00	$1,0 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$	
		5,00	1,00	$1,1 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$	
	2	0,50	1,00	$6,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$	
		2,00	1,00	$9,0 \times 10^{-2}$		1,00	$4,0 \times 10^{-2}$	
		5,00	1,00	$1,1 \times 10^{-1}$		1,00	$3,0 \times 10^{-2}$	
	5	0,50	1,00	$4,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$	
		2,00	1,00	$9,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$	
		5,00	1,00	$1,1 \times 10^{-1}$		1,01	$4,0 \times 10^{-2}$	
	0	0,50	1,02	$5,0 \times 10^{-2}$		1,20	1,00	$9,0 \times 10^{-2}$
			1,00	$6,0 \times 10^{-2}$			1,00	$5,0 \times 10^{-2}$
			1,50	$5,0 \times 10^{-2}$			1,01	$2,0 \times 10^{-2}$
1		0,50	1,00	$4,0 \times 10^{-2}$	1,00		$3,0 \times 10^{-2}$	
		1,00	1,00	$5,0 \times 10^{-2}$	1,00		$4,0 \times 10^{-2}$	
		1,50	1,00	$6,0 \times 10^{-2}$	1,02		$3,0 \times 10^{-2}$	
500	0	0,50	1,05	$6,0 \times 10^{-2}$	0,70	1,03	$7,0 \times 10^{-2}$	
		2,00	1,00	$6,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$	
		5,00	1,00	$2,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$	
	2	0,50	1,00	$4,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$	
		2,00	1,00	$5,0 \times 10^{-2}$		1,00	$3,0 \times 10^{-2}$	
		5,00	1,00	$7,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$	
	5	0,50	1,00	$3,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$	
		2,00	1,01	$6,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$	
		5,00	1,00	$7,0 \times 10^{-2}$		1,00	$2,0 \times 10^{-2}$	
	0	0,50	1,00	$3,0 \times 10^{-2}$		1,20	1,00	$3,0 \times 10^{-2}$
			1,00	$4,0 \times 10^{-2}$			1,00	$3,0 \times 10^{-2}$
			1,50	$3,0 \times 10^{-2}$			1,00	$1,0 \times 10^{-2}$
1		0,50	1,00	$2,0 \times 10^{-2}$	1,00		$2,0 \times 10^{-2}$	
		1,00	1,01	$3,0 \times 10^{-2}$	1,00		$3,0 \times 10^{-2}$	
		1,50	1,00	$4,0 \times 10^{-2}$	1,00		$2,0 \times 10^{-2}$	

5.4 Trabalhando com Dados Artificiais

Nesta seção temos o interesse em verificar se os critérios de seleção de modelos AIC e BIC são capazes de selecionar adequadamente modelos cujos dados tenham distribuição k -MPS.

Inicialmente foram gerados amostras de dados artificiais das distribuições k -MP, k -MG e k -MB todas de tamanho $n = 200$. Consideramos ainda, diferentes pontos de modificação $k = 0, 1$ e 2 além de diferentes valores para o parâmetro $\mu = 0,5$ e 2 . No caso do parâmetro de modificação p , este foi fixado em $0,7$.

O ajuste da distribuição aos dados foi feito da seguinte forma: para cada conjunto de

dados artificiais com modificação em um ponto k realizamos o ajuste considerando o próprio ponto de modificação k além de outros dois pontos diferentes. Os resultados destes ajustes para essas distribuições são apresentados nas Tabelas 13 - 15.

Tabela 13 – Teste sobre a seleção de modelos para a distribuição k -MP.

k -MPS	k	μ	k.test	$\hat{\mu}$	$IC_{\mu}(95\%)$	\hat{p}	$IC_p(95\%)$	AIC	BIC
k -MP	0	0,5	0	0,53	(0,30 , 0,77)	0,75	(0,59 , 0,90)	336,19	342,79
			1	0,44	(0,33 , 0,54)	1,07	(0,98 , 1,15)	336,35	342,94
			2	0,38	(0,28 , 0,49)	0,99	(0,95 , 1,02)	337,75	344,34
		2	0	1,90	(1,62 , 2,19)	0,69	(0,61 , 0,77)	619,24	625,84
			1	1,25	(1,11 , 1,39)	1,26	(1,18 , 1,35)	632,21	638,80
			2	1,36	(1,20 , 1,53)	1,07	(1,00 , 1,14)	655,67	662,27
	1	0,5	0	0,13	(0,04 , 0,21)	5,30	(4,73 , 5,87)	329,59	336,18
			1	0,40	(0,26 , 0,55)	0,56	(0,47 , 0,65)	328,54	335,14
			2	0,83	(0,69 , 0,97)	1,14	(1,11 , 1,17)	365,46	372,06
		2	0	1,32	(1,12 , 1,52)	1,24	(1,19 , 1,30)	586,10	592,70
			1	1,97	(1,70 , 2,23)	0,66	(0,57 , 0,76)	569,61	576,20
			2	1,68	(1,51 , 1,85)	1,12	(1,05 , 1,19)	602,82	609,42
	2	0,5	0	1,19	(0,96 , 1,42)	0,87	(0,77 , 0,97)	522,21	528,81
			1	1,03	(0,90 , 1,15)	1,28	(1,20 , 1,37)	495,34	501,94
			2	0,52	(0,37 , 0,68)	0,65	(0,58 , 0,73)	450,11	456,71
		2	0	1,94	(1,71 , 2,17)	1,02	(0,96 , 1,07)	650,56	657,16
			1	1,80	(1,62 , 1,99)	1,21	(1,14 , 1,28)	630,17	636,77
			2	1,96	(1,72 , 2,19)	0,67	(0,58 , 0,77)	599,46	606,05

Tabela 14 – Teste Sobre a Seleção de Modelos para a distribuição k -MG

k -MPS	k	μ	k.test	$\hat{\mu}$	$IC_{\mu}(95\%)$	\hat{p}	$IC_p(95\%)$	AIC	BIC
k -MG	0	0,5	0	0,58	(0,31 , 0,86)	0,65	(0,49 , 0,81)	324,46	331,06
			1	0,44	(0,33 , 0,55)	1,10	(1,04 , 1,16)	321,01	327,61
			2	0,30	(0,20 , 0,41)	0,95	(0,91 , 0,99)	324,21	330,81
		2	0	2,02	(1,52 , 2,52)	0,71	(0,61 , 0,81)	645,19	651,78
			1	1,38	(1,14 , 1,62)	1,14	(1,08 , 1,21)	649,13	655,72
			2	1,45	(1,19 , 1,71)	1,03	(0,98 , 1,08)	662,16	668,75
	1	0,5	0	0,24	(0,14 , 0,34)	2,93	(2,57 , 3,28)	415,16	421,76
			1	0,55	(0,39 , 0,71)	0,67	(0,58 , 0,76)	406,91	413,51
			2	0,79	(0,62 , 0,96)	1,08	(1,05 , 1,11)	453,36	459,95
		2	0	1,56	(1,25 , 1,88)	1,26	(1,17 , 1,36)	747,98	754,57
			1	2,26	(1,82 , 2,70)	0,77	(0,69 , 0,86)	733,78	740,37
			2	1,97	(1,64 , 2,31)	1,00	(0,95 , 1,06)	763,73	770,32
	2	0,5	0	0,73	(0,52 , 0,94)	1,32	(1,15 , 1,48)	540,30	546,89
			1	0,96	(0,78 , 1,15)	1,09	(1,01 , 1,16)	542,46	549,06
			2	0,49	(0,33 , 0,66)	0,69	(0,62 , 0,76)	454,93	461,52
		2	0	1,59	(1,27 , 1,91)	1,23	(1,13 , 1,33)	748,31	754,91
			1	1,90	(1,58 , 2,22)	1,06	(0,99 , 1,13)	758,02	764,61
			2	1,94	(1,55 , 2,33)	0,72	(0,64 , 0,80)	693,49	700,09

Tabela 15 – Teste Sobre a Seleção de Modelos para a distribuição k -MB

k -MPS	k	μ	k.test	$\hat{\mu}$	$IC_{\mu}(95\%)$	\hat{p}	$IC_p(95\%)$	AIC	BIC
k -MB	0	0,5	0	0,45	(0,22 , 0,68)	0,80	(0,63 , 0,97)	313,48	320,07
			1	0,39	(0,29 , 0,50)	1,05	(0,97 , 1,13)	313,32	319,92
			2	0,33	(0,23 , 0,43)	0,98	(0,95 , 1,01)	313,35	319,95
		2	0	1,95	(1,70 , 2,21)	0,69	(0,61 , 0,76)	602,88	609,48
			1	1,26	(1,13 , 1,38)	1,34	(1,26 , 1,43)	608,81	615,41
			2	1,37	(1,21 , 1,52)	1,03	(0,95 , 1,11)	651,78	658,38
	1	0,5	0	0,24	(0,11 , 0,36)	2,70	(2,38 , 3,02)	362,42	369,01
			1	0,48	(0,34 , 0,61)	0,68	(0,58 , 0,78)	359,12	365,71
			2	0,80	(0,67 , 0,94)	1,13	(1,10 , 1,16)	369,66	376,26
		2	0	1,48	(1,29 , 1,68)	1,13	(1,07 , 1,18)	599,82	606,42
			1	1,88	(1,66 , 2,10)	0,76	(0,66 , 0,86)	589,04	595,64
			2	1,69	(1,53 , 1,86)	1,08	(1,00 , 1,16)	606,09	612,69
	2	0,5	0	1,18	(0,96 , 1,41)	0,82	(0,73 , 0,92)	505,37	511,96
			1	0,98	(0,86 , 1,10)	1,27	(1,18 , 1,37)	485,61	492,20
			2	0,53	(0,38 , 0,68)	0,70	(0,62 , 0,77)	454,11	460,70
		2	0	1,91	(1,70 , 2,11)	1,02	(0,97 , 1,07)	613,13	619,72
			1	1,82	(1,64 , 1,99)	1,15	(1,07 , 1,23)	604,27	610,87
			2	1,92	(1,72 , 2,12)	0,76	(0,66 , 0,86)	588,53	595,12

Nas Tabelas 13, 14 e 15 verificamos que os menores valores obtidos pelos critérios AIC e BIC são aqueles nos quais os pontos de modificação considerados no ajuste (4° coluna) coincidem com os pontos de modificação dos quais os dados artificiais foram gerados (2° coluna), com exceção das distribuições k -MG e k -MB no ponto de modificação $k = 0$ com parâmetro $\mu = 0.5$ e $k.test = 0$ em que os valores de AIC e BIC ficaram levemente maiores que os demais.

Observamos também que para os pontos de modificação considerados no ajuste (4° coluna) que são diferentes dos pontos de modificação dos dados artificiais (2° coluna), em todos os casos o valor estimado para o parâmetro de modificação p por \hat{p} estão diferentes, além de seus respectivos intervalos de confiança não compreendem o verdadeiro valor do parâmetro. Desses fatos, podemos concluir que.

- i - Os critérios de seleção de modelos AIC e BIC são capazes de selecionar adequadamente modelos cujos dados possuem distribuição k -MPS.
- ii - Uma vez que não se tem o conhecimento prévio de qual é o ponto de modificação em uma amostra com distribuição k -MPS, é possível realizar ajustes considerando diferentes pontos de modificação.

APLICAÇÕES COM DADOS REAIS

Neste capítulo realizamos análises de seis conjunto de dados reais, dos quais seus resultados foram apresentados nas Tabelas 16 - 32 com as estimativas de máxima verossimilhança para os parâmetros μ e p bem como os seus respectivos intervalos com 95% de confiança, o número esperado de observações, denotado por $\hat{n}_{(k)}$, identificando as quantidades das observações k estruturais provenientes da distribuição PS associada, junto com a quantidade de observações k modificadas (k -M) e os resultados obtidos pelos critérios de seleção de modelos AIC e BIC.

Além disso, serão apresentadas Tabelas referentes ao teste de aderência Kolmogorov-Smirnov, no qual utilizamos com um nível de significância $\alpha = 5\%$ sob as hipóteses:

H_0 : Os dados são provenientes da distribuição k-MPS

H_1 : Os dados não são provenientes da distribuição k-MPS

sendo que, se a estatística do teste de Kolmogorov-Smirnov D_n for maior do que o valor crítico tabelado para uma amostra de tamanho n e o nível de significância pré-estabelecido, rejeitamos a hipótese H_0 .

6.1 Notificações de Leptospirose no Estado da Bahia

A contaminação humana por leptospirose ocorre através do contato com água contaminada com urina, principalmente, de ratos. Regiões que sofrem com a falta de saneamento básico e com restrições de acesso à serviços como saúde e educação, são regiões em que a população possui maior risco de contaminação.

Neste sentido, consideramos um conjunto de dados de notificações de leptospirose das

idades do Estado da Bahia (Brasil), em 2004 (SEI,2004). A Tabela 16¹ apresenta a distribuição de frequência das notificações de leptospirose das 415 cidades do Estado da Bahia consideradas no estudo.

Tabela 16 – Distribuição de frequência de notificações de leptospirose das cidades do Estado da Bahia em 2004.

y_i	0	1	2	3	4	5	6	11	12	14	16	111
f_i	365	29	7	3	2	2	1	2	1	1	1	1

Na Figura 9² (A) vemos o mapa do Estado da Bahia, no qual são identificadas 50 cidades que apresentaram notificações de leptospirose em 2004. Observamos a grande frequência de zeros no conjunto de dados corresponde à ausência de notificação da doença em questão, o que não implica na ausência da doença. Como pretendemos mensurar a probabilidade de não-notificação da doença, definimos um grupo de cidades que estão em risco de notificação da doença dentre todas as cidades consideradas no estudo, de modo que, esse grupo é composto por cidades que registraram notificações da doença em 2004 e/ou no ano de 2003 e, que possuam índice de desenvolvimento humano (IDH) inferior a 0,66.

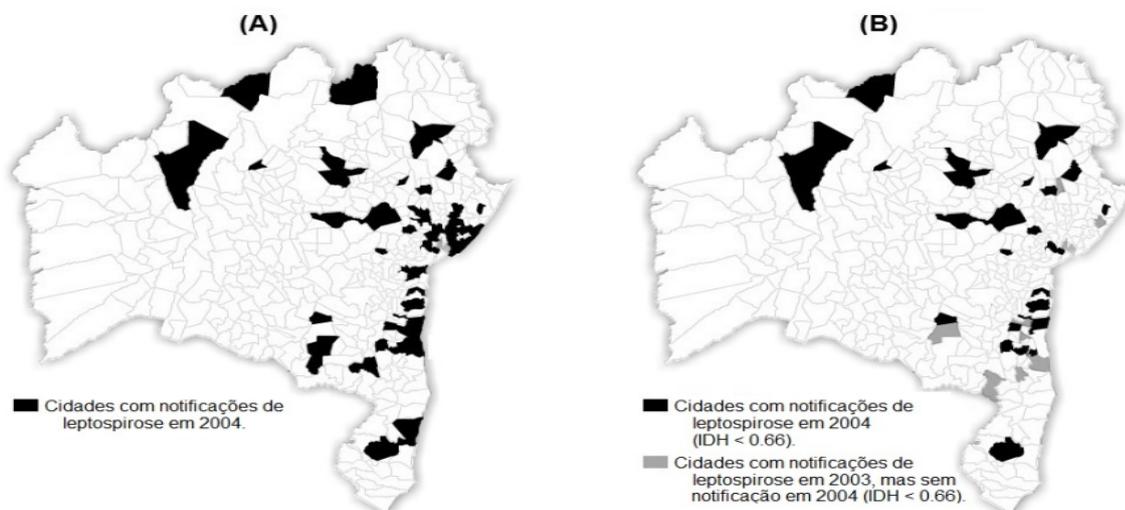


Figura 9 – Mapa do Estado da Bahia. (A) Cidades que apresentaram notificações em 2004. (B) Cidades do grupo de risco em 2004.

Como consequência, é possível mensurar a probabilidade de uma cidade em risco de notificar ou não a doença. Desta forma, obtemos uma sub-amostra de notificações de leptospirose ocorridas em cidades de risco com baixo IDH em 2004, no qual os zeros estão relacionadas apenas às cidades que tiveram casos no ano de 2003, mas não apresentaram casos no ano seguinte. Na Figura 10³, é apresentado o diagrama ilustrando a obtenção da sub-amostra.

¹ Conjunto de dados da Tabela 16 foi obtido da tese de doutorado de Conceição (2013)

² A Figura 9 foi obtido da tese de doutorado em Conceição (2013).

³ A Figura 10 foi obtida da tese de doutorado em Conceição (2013).

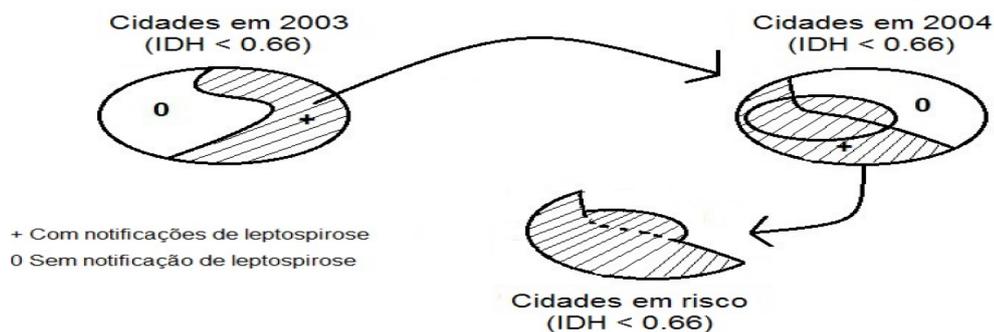


Figura 10 – Diagrama da obtenção da sub-amostra.

Na sub-amostra, temos 13 cidades que apresentaram notificações em 2003, 30 cidades em 2004 e apenas 4 cidades em ambos os anos, resultando em um conjunto de dados com notificações de 39 cidades. Na Tabela 17⁴ temos a distribuição de frequência dos dados da sub-amostra. Na Figura 9 (B), são identificadas as cidades do grupo de risco com IDH<0,66.

Tabela 17 – Distribuição de frequência de notificações de leptospirose das cidades em risco do Estado da Bahia em 2004.

y_i	0	1	2	3
f_i	9	25	4	1

Na análise descritiva dos dados completos detectamos uma possível observação discrepante: a observação 415, correspondente à 111 notificação de leptospirose na cidade de Salvador, capital do Estado da Bahia. Consideramos a distribuição 0-MP para o ajuste aos dados, com e sem a observação 415. Na Tabela 18 são apresentados as estimativas de máxima verossimilhança para os parâmetros μ e p , a frequência esperada das observações resultante do processo de k -modificadas (k -M), a frequência esperada das observações estruturais oriundas da distribuição Poisson tradicional (P), além dos resultados dos critérios de seleção de modelos AIC e BIC .

Tabela 18 – Resultados das estimativas dos parâmetros da distribuição 0-MP.

Dados	$\hat{\mu}$	\hat{p}	$\hat{n}_{(k=0)}$		AIC	BIC
			k -M	P		
Completos	4,98 (4,36 ; 5,61)	0,12 (0,09 ; 0,15)	362	3	1117,98	1126,03
Sem a obs. 415	2,66 (2,16 ; 3,15)	0,13 (0,09 ; 0,16)	336	29	571,92	579,97

Na Tabela 18, notamos que a retirada da observação 415 do conjunto de dados impactou bastante a estimativa do parâmetro μ . No entanto, essa observação retirada pouco afetou na estimativa do parâmetro p , de modo que, em ambos os casos, o conjunto de dados configura-se

⁴ Os dados da Tabela 17 foram obtidos da tese de doutorado em Conceição (2013).

como uma distribuição 0-IP. Embora a retirada da observação 415 não alterou a classificação da distribuição quanto a sua modificação, os resultados dos critérios de seleção de modelos *AIC* e *BIC*, apontam que o ajuste sem a observação 415 é o mais adequado, uma vez que seus valores são menores em relação ao ajuste com os dados completos.

Observamos que a estimativa para o parâmetro p foi $\hat{p} = 0,12$, evidenciando que a amostra é inflacionada de zero. Segundo a distribuição ajustada 0-IP, o valor da frequência estimada de zero é de 365 observações, isso implica que a quantidade de zero estrutural proveniente da distribuição Poisson tradicional com média $\hat{\mu} = 4,98$, é dado por:

$$\hat{n}_{(k=0)}(P) = n \cdot \pi_p(0, \hat{\mu}) = 415 \cdot e^{-4,98} \approx 3.$$

Desta forma, podemos concluir que o processo de zero-modificação (0-M) contribuiu com a adição de 362 observações no processo original de contagem, ocasionando inflação de observações zero no conjunto de dados. Equivalentemente podemos interpretar os resultados para a análise do conjunto de dados sem a observação 415. Na Figura 11 temos a comparação gráfica das frequências observadas e esperadas para as observações da amostra com e sem a observação 415.

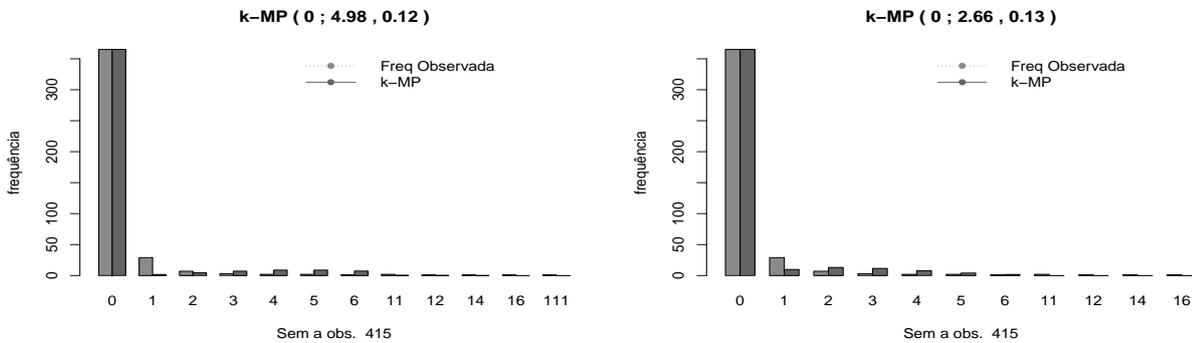


Figura 11 – Comparação entre a frequência observada e a esperada segundo a distribuição 0-IP considerando o conjunto de dados das notificações de leptospirose com e sem a observação 415.

Na Tabela 19 são apresentados os resultados do teste de aderência Kolmogorov-Smirnov considerando o ajuste da distribuição 0-IP ao conjunto de dados de notificações de leptospirose com e sem a observação 415.

Tabela 19 – Teste de aderência Komolgorov Smirnov do ajuste da distribuição 0-IP aos dados de notificações de leptospirose com e sem a observação 415.

Distribuição	Dados	$\hat{\mu}$	\hat{p}	D_n
0-IP	Completos	4,98	0,12	0.06
	Sem a obs. 415	2,66	0,13	0.04

Da Tabela 19 observamos que os valores obtidos pela estatística D_n para a distribuição 0-IP considerando o conjunto de dados de notificações de leptospirose com e sem a observação 415 são menores do que 0,07 que é aproximadamente o valor crítico tabelado para uma amostra de tamanho $n > 50$ e nível de significância $\alpha = 5\%$. Portanto, nos dois casos não rejeitamos a hipótese de que os conjuntos de dados são provenientes das distribuições 0-IP.

Para a sub-amostra de notificações de leptospirose das cidades em risco e com baixo desenvolvimento as distribuições 0-MP e 1-MP foram ajustadas. Na Tabela 20 são apresentadas as estimativas de máxima verossimilhança para os parâmetros μ e p , a frequência esperada das observações resultante do processo de k -modificação (k -M), a frequência esperada das observações estruturais oriundas da distribuição Poisson tradicional, bem como os resultados dos critérios de seleção de modelos AIC e BIC .

Tabela 20 – Resultados do ajuste da distribuições k -MP para a sub-amostra.

k	$\hat{\mu}$	\hat{p}	\hat{n}_k		AIC	BIC
			k-M	P		
0	0,38 (0,08 ; 0,67)	2,45 (2,03 ; 2,87)	-18	27	78,64	81,96
1	0,86 (0,47 ; 1,25)	0,56 (0,33 ; 0,80)	11	14	78,78	82,11

Para o ponto de modificação $k = 0$, a estimativa para o parâmetro p foi $\hat{p} = 2,45$, evidenciando que a sub-amostra é deflacionada de zero. Segundo a distribuição ajustada 0-DP, o valor da frequência estimada de zero é de 9 observações, isso implica que a quantidade de zero estrutural proveniente da distribuição Poisson tradicional com média $\hat{\mu} = 0,86$, é dado por:

$$\hat{n}_{(k=0)}(P) = n \cdot \pi_p(0, \hat{\mu}) = 39 \cdot e^{-0,38} \approx 27.$$

Desta forma, podemos concluir que o processo de zero-modificação (0-M) contribuiu com a retirada de 18 observações no processo de contagem original, ocasionando deflação de observações zero no conjunto de dados.

Para o ponto de modificação $k = 1$ a estimativa para o parâmetro p foi $\hat{p} = 0,56$, evidenciando que a sub-amostra é inflacionada de 1 (um). Segundo a distribuição ajustada 1-IP, o valor da frequência estimada de 1 (um) é de 25 observações, isso implica que a quantidade de 1 (um) estrutural proveniente da distribuição Poisson tradicional com média $\hat{\mu} = 0,86$, é dado por:

$$\hat{n}_{(k=1)}(P) = n \cdot \pi_p(1, \hat{\mu}) = 39 \cdot e^{-0,86} \approx 14.$$

Desta forma, podemos concluir que o processo de 1-modificação (1-M) contribuiu com a adição de 11 observações no processo de contagem original, ocasionando inflação de observações 1 (um) no conjunto de dados.

Ainda na Tabela 20 podemos observar que os resultados dos critérios de seleção de modelos *AIC* e *BIC*, indicam que não há diferenças significativas entre os distribuições 0-DP e 1-IP no ajuste ao conjunto de dados da sub-amostra. Na Figura 12 é apresentada a comparação gráfica entre a frequência esperada segundo a distribuição de Poisson, a frequência observada dos dados e a frequência esperada segundo a distribuição *k*-MP nos pontos de modificação $k = 0$ e $k = 1$.

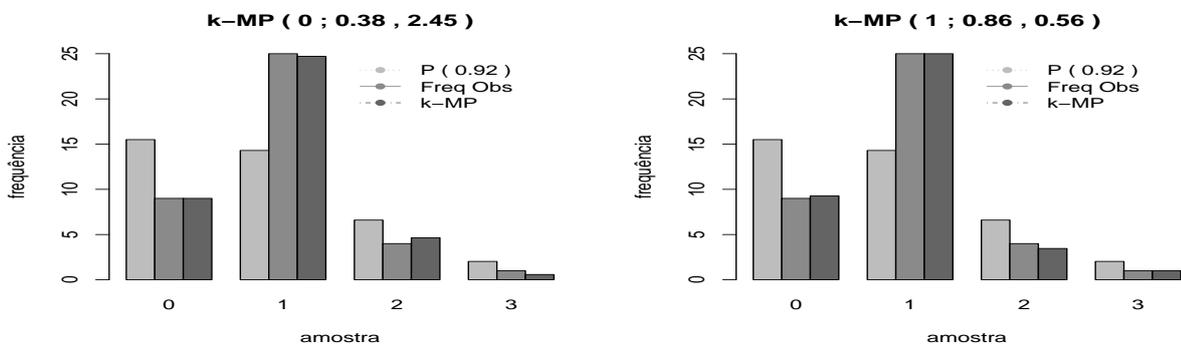


Figura 12 – Comparação gráfica entre a frequência esperada segundo a distribuição de Poisson, frequência das observações e a frequência esperada segundo a distribuição *k*-MP ajustada aos dados da sub-amostra considerando os pontos de modificação $k = 0$ e $k = 1$.

Na Tabela 21 são apresentados os resultados do teste de aderência Kolmogorov-Smirnov considerando o ajuste da distribuição *k*-MP ao conjunto de dados da sub-amostra.

Tabela 21 – Teste de aderência Komolgorov Smirnov para a distribuição *k*-MP, considerando pontos de modificação $k = 0$ e $k = 1$.

Distribuição	k	$\hat{\mu}$	$\hat{\sigma}$	D_n
<i>k</i> -MP	0	0,38	2,45	0,026
	1	0,86	0,56	0,023

Da Tabela 21 observamos que os valores obtidos pela estatística D_n para a distribuição *k*-MP considerando os pontos de modificação $k = 0$ e $k = 1$, são menores do que 0,22 que é aproximadamente o valor crítico tabelado para uma amostra de tamanho $n = 39$ e nível de significância $\alpha = 5\%$. Portanto, não rejeitamos a hipótese de que o conjunto de dados seja proveniente da distribuição 0-DP ou da distribuição 1-IP.

6.2 Número de Gols em Partidas de Futebol

Esta aplicação consiste em analisar o experimento que reuniu o número de gols marcados pelo Barcelona em todos os confrontos com o Real Madrid desde o ano de 1955 até 2015. Neste período, foram realizadas 131 partidas, totalizando 68 vitórias para Barcelona, 62 vitórias do Real Madrid e apenas 1 empate. A Tabela 22 apresenta a distribuição de frequência do número de gols marcados pelo Barcelona em todos os confrontos.

Tabela 22 – Número de gols marcados pelo Barcelona em todos os confrontos com o Real Madrid no período de 1955 a 2015.

y_i	0	1	2	3	4	5	6	Total
f_i	27	39	34	21	5	3	2	131

A distribuição k -MP com modificações nas observações $k = 0$, $k = 1$ e $k = 2$ foi considerada para analisar este conjunto de dados. Na Tabela 23 são apresentados os resultados destas análises, como as estimativas de máxima verossimilhança para os parâmetros μ e p , a frequência esperada das observações resultante do processo de k -modificadas (k -M), a frequência esperada das observações estruturais oriundas da distribuição Poisson tradicional, bem como os resultados dos critérios de seleção de modelos AIC e BIC .

Tabela 23 – Resultados das análises para os dados do Número de Gols de Partidas de Futebol considerando a distribuição k -MP nos pontos de modificação $k = 0$, $k = 1$, $k = 2$.

k	$\hat{\mu}$	\hat{p}	\hat{n}_k		AIC	BIC
			k -M	P		
0	1,70 (1,41 ; 1,99)	0,97 (0,89 ; 1,06)	3	24	427,85	433,60
1	1,63 (1,40 ; 1,86)	1,03 (0,92 ; 1,15)	-3	42	427,94	433,69
2	1,65 (1,43 ; 1,87)	1,00 (0,90 ; 1,10)	0	34	428,17	433,92

Para o ponto de modificação $k = 0$ a estimativa para o parâmetro p foi $\hat{p} = 0,97$, evidenciando não há modificação das observações zero na amostra. Segundo a distribuição ajustada 0-MP, o valor da frequência estimada de zero é de 27 observações, isso implica que a quantidade de zero estrutural proveniente da distribuição Poisson tradicional com média $\hat{\mu} = 1,70$, é dado por:

$$\hat{n}_{(k=0)}(P) = n \cdot \pi_p(0, \hat{\mu}) = 131 \cdot e^{-1,70} \approx 24.$$

Desta forma, concluímos que não há evidência do processo de zero-modificação (0-M) no processo original de contagem para neste conjunto de dados. Para o ponto de modificação $k = 1$ a estimativa para o parâmetro p foi $\hat{p} = 1,03$, evidenciando não há modificação das observações

1 (um) na amostra. Segundo a distribuição ajustada 1-MP, o valor da frequência estimada de 1 (um) é de 39 observações, isso implica que a quantidade de 1 (um) estrutural proveniente da distribuição Poisson tradicional com média $\hat{\mu} = 1,63$, é dado por:

$$\hat{n}_{(k=1)}(P) = n \cdot \pi_p(1, \hat{\mu}) = 131 \cdot e^{-1,63} \cdot 1,63 \approx 42.$$

Desta forma, concluímos que não há evidência do processo de 1-modificação (1-M) no processo original de contagem para neste conjunto de dados. Uma vez que os resultados das análises para o ponto $k = 2$ ficaram bastante semelhantes aos dos pontos $k = 0$ e $k = 1$, a interpretação dos resultados deve ser a mesma. Ou seja, como a estimativa para o parâmetro de modificação p foi $\hat{p} = 1$, podemos considerar que para as observações 2 (dois), não há evidência de modificação na probabilidade de sua ocorrência.

Ainda na Tabela 23, notamos que todas as distribuições ajustadas evidenciaram que os dados podem ser representados pela distribuição de Poisson tradicional, uma vez que os valores estimados de p ficaram próximos de 1 (um), além de seus respectivos intervalos com 95% de confiança incluírem o valor 1 (um). Os resultados obtidos pelos critérios de seleção de modelos *AIC* e *BIC* não apresentaram diferenças significativas.

Na Figura 13 são apresentados três gráficos do comparativo das frequências esperadas das distribuições ajustadas Poisson tradicional, k -MP e da frequência observada dos dados.

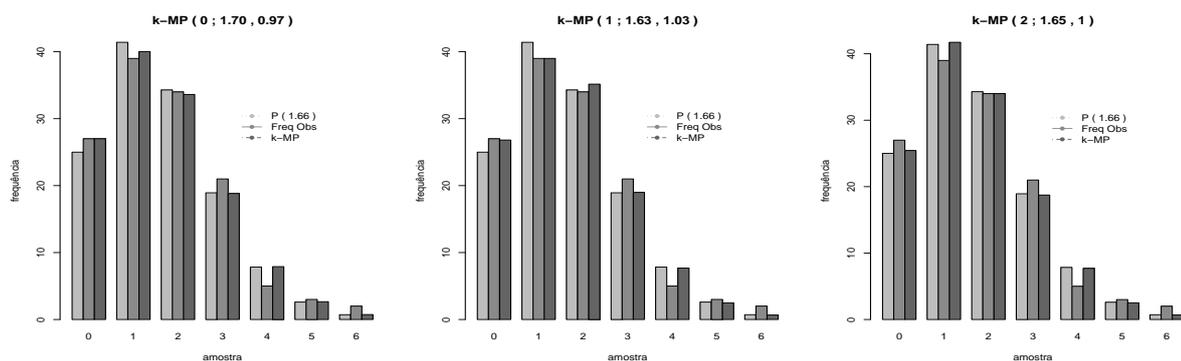


Figura 13 – Gráfico comparativo entre as frequências das distribuições Poisson e k -MP, considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$.

Na Tabela 24 são apresentados os resultados do teste de aderência Kolmogorov-Smirnov, considerando o ajuste da distribuições k -MP ao conjunto de dados do números de gols.

Tabela 24 – Teste de aderência Kolmogorov-Smirnov para a distribuição k -MP, considerando pontos de modificação $k = 0$, $k = 1$ e $k = 2$.

Distribuição	k	$\hat{\mu}$	\hat{p}	D_n
Poisson	-	1,66	-	0,024
k -MP	0	1,70	0,97	0,024
	1	1,63	1,03	0,024
	2	1,65	1,00	0,024

Da Tabela 24 observamos que os valores obtidos pela estatística D_n para a distribuição de Poisson e k -MP considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$, são menores do que 0,119 que é aproximadamente o valor crítico tabelado para uma amostra de tamanho $n > 50$ e nível de significância $\alpha = 5\%$. Portanto, não rejeitamos a hipótese H_0 .

Observação: Ressaltamos através desta análise a grande flexibilidade da família de distribuição k -MP, pois como vimos, a distribuição k -MP ajustada com diferentes pontos de modificação ($k = 0, 1, 2$), apontaram para a distribuição mais simples, a distribuição Poisson tradicional, que é um caso particular da distribuição k -MP ($\hat{p} \approx 1$ para todos os casos).

6.3 Variação Global da Temperatura

A Figura 14 refere-se ao estudo realizado em 1978 pelos pesquisadores (ANGELL; KORSHOVER, 1978) no qual publicaram um trabalho sobre suas análises referente a variação da temperatura (em graus Celsius, C°) média global entre os anos de (1958 a 2008).

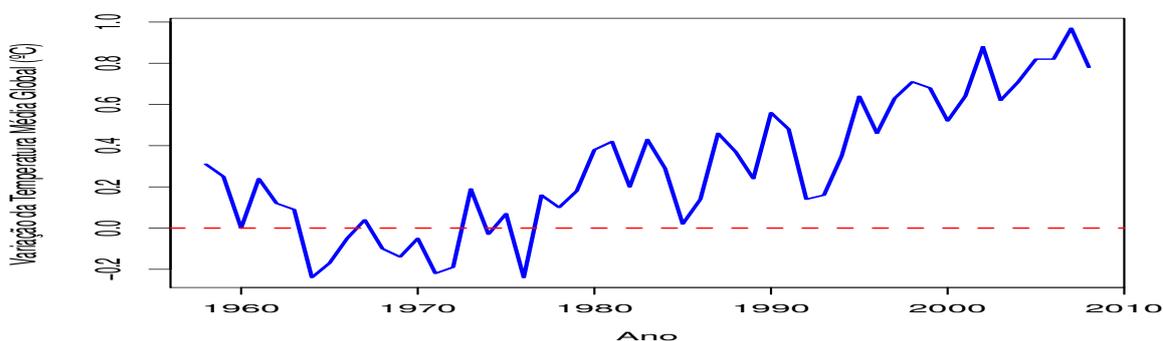


Figura 14 – Gráfico da Variação Global de Temperatura.

Realizamos um experimento com o objetivo de verificar se em determinado ano ocorreu uma variação positiva, a qual denominamos por sucesso, referindo-se ao aumento relativo da temperatura média anual, ou uma variação negativa, a qual denominamos por fracasso referindo-se a diminuição relativa da temperatura média anual. Denotamos por θ a probabilidade de

sucesso no processo de *Bernoulli* ($\theta = \frac{1}{(1-\mu)}$, com $\theta \in (0, 1)$), de modo que para valores de θ próximos de um, maiores serão as chances de ocorrer variações positivas da temperatura em determinados anos, já quando os valores de θ estiverem mais próximos de zero, menores são essas chances.

Do processo de *Bernoulli* definimos a variável aleatória Y que representa o número de anos consecutivos que houve variação negativa de temperatura média anual (fracasso) até a ocorrência de uma variação positiva (sucesso), possibilitando a suposição de um processo geométrico. Na Tabela 25 é apresentado a distribuição de frequência dos dados obtidos deste experimento, na qual observamos indícios de inflação da observação zero.

Tabela 25 – Distribuição de frequência do número de anos consecutivos com variação negativa da temperatura média anual até a ocorrência de uma variação positiva no período entre 1958 e 2008.

y_i	0	1	3	5	Total
f_i	34	3	1	1	39

Ajustamos a distribuição 0-MG a este conjunto de dados. Na Tabela 26 são apresentados as estimativas de máxima verossimilhança para os parâmetros μ e p , a frequência esperada das observações resultante do processo de k -modificadas (k -M), a frequência esperada das observações estruturais oriundas da distribuição Geométrica tradicional, bem como os resultados dos critérios de seleção de modelos *AIC* e *BIC*.

Tabela 26 – Estimativas de máxima verossimilhança dos parâmetros.

k	$\hat{\mu}$	\hat{p}	$\hat{n}_{(k=0)}$		AIC	BIC
			0-M	G		
0	1,20 (0,00 ; 2,62)	0,24 (0,04 ; 0,43)	16	18	49,03	52,36

Temos que a estimativa para o parâmetro p foi $\hat{p} = 0,24$, evidenciando que a amostra é inflacionada de zero. Segundo a distribuição ajustada 0-IG, o valor da frequência estimada de zero é de 34 observações, isso implica que a quantidade de zero estrutural proveniente da distribuição Geométrica tradicional com média $\hat{\mu} = 1,20$, é dado por:

$$\hat{n}_{(k=0)}(G) = n \cdot \pi_G(0, \hat{\mu}) = 39 \cdot \frac{(1,20)^0}{(1 + 1,20)^1} \approx 18.$$

Desta forma, podemos concluir que o processo de zero-modificação (0-M) contribuiu com a adição de 16 observações no processo de original contagem, ocasionando inflação de observações zero no conjunto de dados.

Na Figura 15 é apresentado o comparativo das frequências esperadas das distribuições ajustadas Geométrica tradicional e distribuição 0-IG com a frequência observada dos dados.

Observamos que há maior proximidade entre a frequência observada com frequência esperada da distribuição 0-IG ajustada.

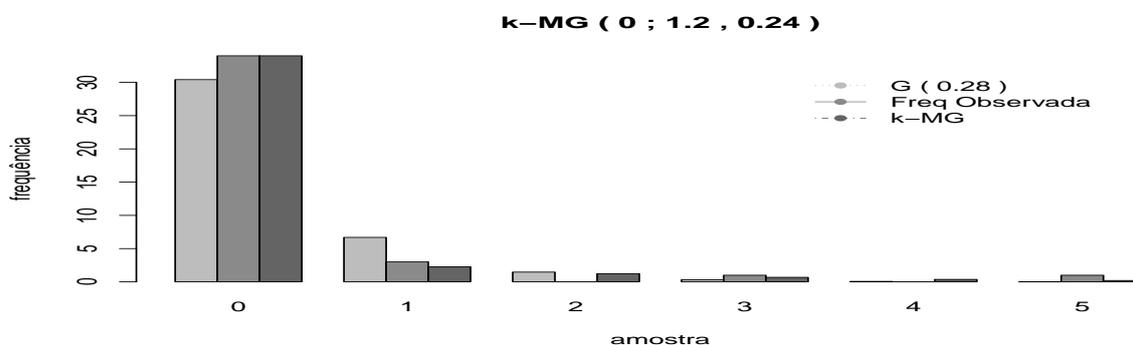


Figura 15 – Gráfico comparativo entre as frequências da distribuição Geométrica, observadas e a distribuição 0-IG.

Na Tabela 27 é apresentado o resultado do teste de aderência Kolmogorov-Smirnov, considerando o ajuste da distribuição k -MG ao conjunto de dados sobre a variação global da temperatura.

Tabela 27 – Teste de aderência Komolgorov Smirnov para a suposição da distribuição 0-IG.

Distribuição	k	$\hat{\mu}$	\hat{p}	D_n
k -MG	0	1,20	0,24	0,051

Da Tabela 27 observamos que o valor obtido pela estatística D_n para a distribuição 0-IG, é menor do que 0,22, que é aproximadamente o valor crítico tabelado para uma amostra de tamanho $n = 39$ e nível de significância $\alpha = 5\%$. Portanto, não rejeitamos a hipótese de que os dados sejam provenientes da distribuição 0-IG.

6.4 Significado de Palavras Técnicas de Estatística

Esta aplicação analisa o um conjunto de dados referente a um experimento no qual foram selecionados 59 alunos de Graduação em Estatística de uma universidade, com pelo menos 1 ano de curso. A fim de verificar seus conhecimentos sobre inglês técnico relacionados à termos estatísticos, foi-lhes então solicitado o significado das seguintes palavras em inglês “Average” e “Standard Deviation”.

Considerando que o número de acerto de cada palavra por um aluno consiste em um ensaio de *Bernoulli* em que a probabilidade de sucesso é dada por θ tal que $\theta \in (0, 1)$, definimos Y como sendo uma variável aleatória de interesse que representa o número total de acertos obtidos por um aluno. A distribuição de frequência do número de acerto pelos 59 alunos é apresentada na Tabela 28.

Tabela 28 – Distribuição de frequência do número de acertos por aluno nas traduções das palavras.

y_i	0	1	2	Total
f_i	18	15	26	59

Consideramos o ajuste da distribuição k -MB aos dados nos pontos de modificação $k = 0$, $k = 1$, $k = 2$ e $m = 2$. Na Tabela 29 são apresentados os resultados deste ajuste, como as estimativas de máxima verossimilhança para os parâmetros μ e p , a frequência esperada das observações resultante do processo de k -modificadas (k -M), a frequência esperada das observações estruturais oriundas da distribuição Binomial tradicional, bem como os resultados dos critérios de seleção de modelos AIC e BIC .

Tabela 29 – Resultados da análise para a distribuição k -MB.

k	$\hat{\mu}$	\hat{p}	\hat{n}_k		AIC	BIC
			k-M	B		
0	1,55 (1,33 ; 1,77)	0,73 (0,61 ; 0,86)	15	3	130,43	134,59
1	1,09 (0,94 ; 1,24)	1,48 (1,26 ; 1,70)	-14	29	130,43	134,59
2	0,59 (0,30 ; 0,87)	0,61 (0,47 ; 0,75)	21	5	130,43	134,59

Considerando $k = 0$, temos que a estimativa para o parâmetro p foi $\hat{p} = 0,73$, evidenciando a possibilidade do ajuste pela distribuição 0-IB. Da suposição de que as observações são oriundas da distribuição 0-IB, o valor da frequência estimada de zero é de 18 observações, isso implica que a quantidade de zero estrutural proveniente da distribuição Binomial tradicional com média $\hat{\mu} = 1,55$, é dado por:

$$\hat{n}_{(k=0)}(B) = n \cdot \pi_B(0, \hat{\mu}) = 59 \binom{2}{0} \frac{1,55^0}{2^2} (2 - 1,55)^{2-0} \approx 3.$$

Desta forma, podemos concluir que o processo de 0-modificação (0-M) contribuiu com a adição de 15 observações no processo de contagem original, ocasionando inflação de observações zero no conjunto de dados.

Ainda na Tabela 29, observamos que a estimativa para o parâmetro p considerando o ponto de modificação $k = 1$ foi $\hat{p} = 1,48$, indicando que os dados podem ser representados pela distribuição 1-DB. Admitindo que as observações são oriundas da distribuição 1-DB, o valor da frequência estimada de 1 (um) é de 15 observações, isso implica que a quantidade de 1 (um) estrutural oriundos da distribuição Binomial tradicional com $\hat{\mu} = 1,09$, é dada por:

$$\hat{n}_{(k=1)}(B) = n \cdot \pi_B(1, \hat{\mu}) = 59 \cdot \binom{2}{1} \frac{1,09^1}{2^2} (2 - 1,09)^{2-1} \approx 29.$$

Desta forma, podemos concluir que o processo de 1-modificação (1-M) contribuiu com a retirada de 14 observações no processo de contagem original, ocasionando deflação de observações 1 (um) no conjunto de dados.

Considerando ainda que a distribuição 2-MB foi ajustada aos dados, observamos que a estimativa para o parâmetro de modificação p foi $\hat{p} = 0,61$, evidenciando que os dados podem ser representados pela distribuição 2-IB. Sob a suposição de que os dados são provenientes da distribuição 2-IB, o valor da frequência estimada de 2 (dois) é de 26 observações, isso implica que a frequência esperada de dois estruturais oriundos da distribuição Binomial tradicional com $\hat{\mu} = 0,59$, é dada por:

$$\hat{n}_{(k=2)}(B) = n \cdot \pi_B(2, \hat{\mu}) = 59 \cdot \binom{2}{2} \frac{0,59^2}{2^2} (2 - 0,59)^{2-2} \approx 5.$$

Desta forma, podemos concluir que o processo de 2-modificação (2-M) contribuiu com a adição de 21 observações no processo de contagem original, ocasionando inflação de observações 2 (dois) no conjunto de dados.

Os resultados obtidos pelos critérios de seleção de modelos *AIC* e *BIC* indicam que não há diferenças significativas entre os modelos ajustados. Dessa forma, podemos considerar que os três modelos ajustados representam adequadamente a distribuição deste conjunto de dados.

Na Figura 16 são apresentadas três gráficos do comparativo das frequências esperadas das distribuições Binomial tradicional e k -MB considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$ com a frequência das observações.

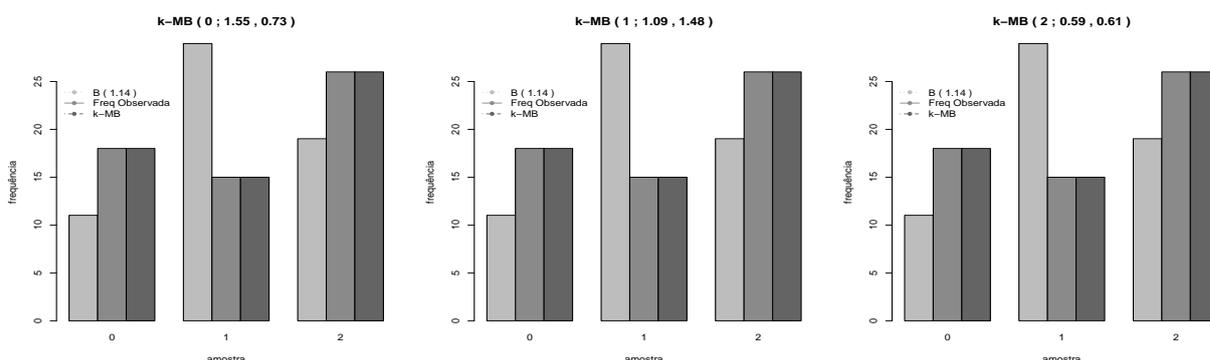


Figura 16 – Gráfico comparativo das frequências esperadas das distribuições Binomial tradicional e k -MB com a frequência das observações.

Na Tabela 30 são apresentados os resultados do teste de aderência Kolmogorov-Smirnov, considerando o ajuste da distribuição k -MB para os pontos de modificação $k = 0$, $k = 1$ e $k = 2$ ao conjunto de dados sobre o significado de palavras técnicas de estatística.

Tabela 30 – Teste Komolgorov-Smirnov de aderência para a distribuição k -MB para os pontos de modificação $k = 0$, $k = 1$ e $k = 2$.

Distribuição	k	$\hat{\mu}$	\hat{p}	D_n
k-MB	0	1,55	0,73	0,00
	1	1,09	1,48	
	2	0,59	0,61	

Da Tabela 30 observamos que o valor obtido pela estatística D_n para a distribuição k -MB considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$, são menores do que 0,18 que é aproximadamente o valor crítico tabelado para uma amostra de tamanho $n > 50$ e nível de significância $\alpha = 5\%$. Portanto, não rejeitamos a hipótese H_0 de que os dados sejam provenientes da distribuição k -MB.

6.5 Estudo que Caracteriza o Estilo de um Autor.

Em estudos para caracterizar o estilo da escrita de autores, são amostradas n palavras e então conta-se o número de vezes em que determinadas palavras aparecem. Os dados deste exemplo referem-se ao número de ocorrências combinadas dos artigos "the", "a" e "an" em amostras de 5 palavras de Macaulay no "Ensaio sobre Milton", retirada da edição Oxford de Macaulay ensaios literários de Bailey. Das palavras, foram retiradas amostras sem reposição em duas linhas escolhidas aleatoriamente de um texto impresso com 50 páginas. Neste contexto, definimos Y como sendo a variável aleatória que representa o número de ocorrência combinada dos artigos em inglês "the", "a" e "an" a cada 5 palavras. Na Tabela 31 é apresentada a distribuição de frequência dos dados deste experimento.

Tabela 31 – Distribuição de frequência da ocorrência combinadas dos artigos "the", "a" e "an".

y_i	0	1	2
f_i	45	49	6

A distribuição k -MB foi ajustada a este conjunto de dados considerando os pontos de modificação $k = 0$, $k = 1$, $k = 2$ e o parâmetro m referente ao números de ensaios de *Bernoulli* igual a 5.

Na Tabela 32, são apresentados os resultados das estimativas de máxima verossimilhança dos parâmetros μ e p da distribuição k -MB, a frequência estimada das observações referente ao processo de modificação (k -M), a frequência estimada dos k -estruturais considerando a distribuição Binomial tradicional, além dos critérios de seleção de modelos *AIC* e *BIC*.

Tabela 32 – Estimativas para o conjunto de dados que Caracteriza o Estilo de um Autor.

k	$\hat{\mu}$	\hat{p}	\hat{n}_k		AIC	BIC
			k-M	B		
0	0,26 (0,06 ; 0,46)	2,36 (1,94 ; 2,77)	-32	77	180,21	185,42
1	0,48 (0,30 ; 0,66)	0,75 (0,61 ; 0,90)	17	32	180,83	186,04
2	0,71 (0,52 ; 0,90)	1,08 (1,02 ; 1,13)	-7	13	186,30	191,51

Observamos que para o ajuste da distribuição 0-MB, a estimativa para o parâmetro de modificação p foi $\hat{p} = 2,36$, evidenciando que os dados podem ser representados pela distribuição 0-DB. Considerando a suposição de que os dados são provenientes da distribuição 0-DB, o valor da frequência estimada de zero é de 45 observações, isso implica que a quantidade de zeros estruturais oriundos da distribuição Binomial tradicional com $\hat{\mu} = 0,26$, é dada por:

$$\hat{n}_{(k=0)}(B) = n \cdot \pi_B(0, \hat{\mu}) = 100 \cdot \binom{5}{0} \frac{0,26^0}{5^5} (5 - 0,26)^{5-0} \approx 77.$$

Desta forma, podemos concluir que o processo de zero modificação (0-M) contribuiu com a retirada de 32 observações do processo original de contagem, ocasionando deflação de observações zero no conjunto de dados.

Considerando o ajuste da distribuição 1-MB aos dados, observamos na Tabela 32 que a estimativa para o parâmetro p foi $\hat{p} = 0,75$, evidenciando que os dados podem ser representados pela distribuição 1-IB. Admitindo que os dados são provenientes da distribuição 1-IB, o valor da frequência estimada de 1 (um) é de 49 observações, isso implica que a frequência esperada de 1 (um) estrutural oriundos da distribuição Binomial tradicional com $\hat{\mu} = 0,48$, é dada por:

$$\hat{n}_{(k=1)}(B) = n \cdot \pi_B(1, \hat{\mu}) = 100 \cdot \binom{5}{1} \frac{0,48^1}{5^5} (5 - 0,48)^{5-1} \approx 32.$$

Desta forma, podemos concluir que o processo de 1-modificação (1-M) contribuiu com a adição de 17 observações do processo original de contagem, ocasionando inflação de observações 1 (um) no conjunto de dados.

Ainda da Tabela 32, considerando o ajuste da distribuição 2-MB aos dados, observamos que a estimativa para o parâmetro p foi $\hat{p} = 1,08$, indicando que os dados podem ser representados pela distribuição 2-DB. Admitindo que os dados são provenientes da distribuição 2-DB, o valor da frequência estimada de 2 (dois) é de 6 observações, isso implica que a frequência esperada de 2 (dois) estrutural oriundos da distribuição Binomial tradicional com $\hat{\mu} = 0,71$, é dada por:

$$\hat{n}_{(k=2)}(B) = n \cdot \pi_B(2, \hat{\mu}) = 100 \cdot \binom{5}{2} \frac{0,71^2}{5^5} (5 - 0,71)^{5-2} \approx 13.$$

Dessa forma, podemos concluir que no processo de modificação das observações dois (2-M) contribuiu com a retirada de 7 observações do processo original de contagem, ocasionando deflação de observações 2 (dois) no conjunto de dados.

Analisando os critérios de seleção de modelos *AIC* e *BIC*, podemos observar que não há diferenças significativas entre os ajustes considerando os pontos de modificação $k = 0$ e $k = 1$, porém os valores de *AIC* e *BIC* considerando o ponto de modificação $k = 2$ ficaram maiores em relação aos obtidos pelos pontos de modificação $k = 0$ e $k = 1$. Desta forma, concluímos que para representar este conjunto de dados o uso da distribuição 0-DB ou da distribuição 1-IB é adequado.

Na Figura 17 são apresentados três gráficos do comparativo das frequências esperadas obtidas pela distribuição Binomial tradicional e distribuição k -MB considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$ com a frequência dos dados observados.

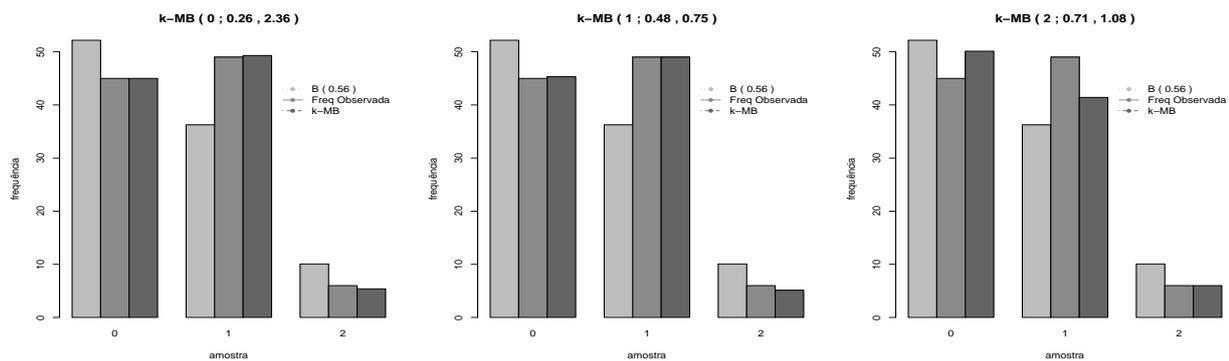


Figura 17 – Gráfico comparativo entre as frequências das distribuições Binomial tradicional, dos dados observados e da distribuição k -MB considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$.

Na Tabela 33 são apresentados os resultados do teste de aderência Kolmogorov-Smirnov, considerando o ajuste da distribuição k -MB considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$ ao conjunto de dados do estudo que caracteriza o estilo de um autores.

Tabela 33 – Teste de aderência Komolgorov Smirnov para a distribuição k -MB.

Distribuição	k	$\hat{\mu}$	\hat{p}	D_n
k -MB	0	0,26	2,36	0,009
	1	0,48	0,75	0,009
	2	0,71	1,08	0,065

Na Tabela 33 observamos que os valores obtidos pela estatística D_n para a distribuição k -MB considerando os pontos de modificação $k = 0$, $k = 1$ e $k = 2$, são menores do que 0,14

que é aproximadamente o valor crítico tabelado para uma amostra de tamanho $n = 100$ e nível de significância $\alpha = 5\%$. Portanto, não rejeitamos a hipótese H_0 .

CONCLUSÃO

Neste trabalho, apresentamos a família de distribuição série de potência k -modificadas cujo principal objetivo foi o de estender a ideia da modificação na probabilidade de observação zero, adotada nas distribuições ZMPS em [Conceição \(2013\)](#), para modificar a probabilidade da observação de qualquer ponto do suporte da família de distribuição PS uniparamétrica cuja função massa de probabilidade possa ser escrita em função de sua média.

A vantagem de utilizar esta família de distribuição para modelar conjunto de dados de contagem, é o fato de sem o conhecimento prévio podermos modelar conjuntos de dados que apresentam qualquer tipo de modificação (inflação ou deflação) na frequência de determinada observação k em relação à família de distribuição PS.

No Capítulo 3 apresentamos as Distribuições Série de Potência k -Modificadas através de sua função massa de probabilidade juntamente com um estudo sobre seus casos particulares, suas caracterizações além das propriedades teóricas relacionadas à teoria das probabilidades.

No Capítulo ?? foram apresentados o Método dos Momentos e o Método de Máxima Verossimilhança para estimação dos parâmetros da distribuição k -MPS($\mu; p$). Quanto ao Método dos Momentos, podemos dizer que: pela sua simplicidade, optamos apenas por apresentar o sistema de equações que torna o processo de estimação possível. Em relação ao método de máxima verossimilhança, além de apresentarmos todo o desenvolvimento teórico, optamos por implementá-lo, por ser o método mais utilizado na literatura, a fim de obter as estimativas de máxima verossimilhança para os parâmetros μ e p , bem como seus respectivos intervalos com 95% de confiança.

Do processo de simulação apresentado no Capítulo 5, consideramos que de forma geral, os resultados obtidos para as estimativas pontuais e intervalares, considerando o procedimento assintótico e de *bootstrap*, mostraram-se satisfatórios. Neste sentido, os resultados obtidos para a probabilidade de cobertura (PC) e medidas de eficiência demonstram respectivamente que o nível de confiança esteve, de forma geral, próximo do nível nominal de 95% confiança e que os

estimadores $\hat{\mu}$ e \hat{p} são consistentes.

Na análise de dados reais, trabalhamos com conjuntos de dados inflacionados, deflacionados e também com conjunto de dados que não apresentaram qualquer modificação na frequência da observação de interesse. Para estas análises, foram plotados gráficos de barras dos ajustes obtidos pelas distribuições PS e k -MPS em comparação com a frequência dos dados observados. De forma geral, notamos que para as observações cuja frequência havia algum tipo de modificação, o ajuste considerando a distribuição k -MPS ficou adequado, ou seja, após o ajuste, a frequência esperada ficou bastante próxima da frequência observada.

Em suma, temos que esta nova família de distribuições k -Modificada é uma alternativa dentre as existentes para modelar conjunto de dados de contagem, com a vantagem de ser capaz de modelar conjuntos de dados em que há modificações nas frequências de suas observações. Neste sentido, podemos dizer que a família de distribuições k -MPS teve seu objetivo alcançado, uma vez que os resultados apresentados pelo desenvolvimento teórico da k -MPS, suas caracterizações, o estudo de simulação e os resultados obtidos pela aplicação em conjunto de dados reais, tiveram um bom desempenho em relação as medidas estatísticas adotadas.

7.1 Trabalhos Futuros

Como propostas de trabalhos futuros elencamos os seguintes tópicos como sugestão de pesquisa.

- 1 - Incluir as distribuições k -Modificadas no contexto dos Modelos de Regressão.
- 2 - Procedimento para identificar o ponto(s) de modificação.
- 3 - Abordagem Bayesiana para o procedimento de estimação dos parâmetros.
- 4 - Incluir as distribuições de Série de Potência Biparamétricas.
- 5 - Incluir procedimento de modificação conjunta (dois ou mais pontos, k_1 e k_2 por exemplo).

APÊNDICE

.1 Derivada de Ordem Primeira da PS

$$\begin{aligned}
 \frac{\partial (\pi_{PS}(k; \mu))}{\partial \mu} &= \frac{\partial}{\partial \mu} \left(\frac{a(k)g(\mu)^k}{f(\mu)} \right) \\
 &= \frac{ka(k)g(\mu)^{k-1}g'(\mu)f(\mu) - a(k)g(\mu)^k f'(\mu)}{f^2(\mu)} \\
 &= \frac{ka(k)g(\mu)^k g'(\mu)f(\mu) - a(k)g(\mu)^k g(\mu)f'(\mu)}{f^2(\mu)} \\
 &= \left(\frac{a(k)g(\mu)^k}{f(\mu)} \right) \left(\frac{kg'(\mu)f(\mu) - g(\mu)f'(\mu)}{g(\mu)f(\mu)} \right) \\
 &= \left(\frac{a(k)g(\mu)^k}{f(\mu)} \right) \left(\frac{kg'(\mu)}{g(\mu)} - \frac{f'(\mu)}{f(\mu)} \right) \\
 &= \frac{\pi_{PS}(k; \mu)g'(\mu)(k - \mu)}{g(\mu)} \\
 &= \frac{\pi_{PS}(k; \mu)(k - \mu)}{\sigma^2}.
 \end{aligned}$$

.2 Derivada de Segunda Ordem da PS

$$\begin{aligned}
\frac{\partial^2 (\pi_{PS}(k; \mu))}{\partial \mu^2} &= \frac{\partial}{\partial \mu} \left(\frac{\pi_{PS}(k; \mu) g'(\mu) (k - \mu)}{g(\mu)} \right) \\
&= \frac{\partial}{\partial \mu} \left[\pi_{PS}(k; \mu) \left(\frac{g'(\mu) (k - \mu)}{g(\mu)} \right) \right] \\
&= \frac{\partial}{\partial \mu} (\pi_{PS}(k; \mu)) \left(\frac{g'(\mu) (k - \mu)}{g(\mu)} \right) + (\pi_{PS}(k; \mu)) \frac{\partial}{\partial \mu} \left(\frac{g'(\mu) (k - \mu)}{g(\mu)} \right) \\
&= \pi_{PS}(k; \mu) \left[\frac{(k - \mu)^2}{(\sigma^2)^2} + \left(\frac{(g'(\mu) (k - \mu))' g(\mu) - g'(\mu) (k - \mu) g'(\mu)}{g(\mu)^2} \right) \right] \\
&= \pi_{PS}(k; \mu) \left[\frac{(k - \mu)^2}{(\sigma^2)^2} + \left(\frac{(g''(\mu) (k - \mu) + g'(\mu) (k - \mu)') g(\mu) - g'(\mu)^2 (k - \mu)}{g(\mu)^2} \right) \right] \\
&= \pi_{PS}(k; \mu) \left[\frac{(k - \mu)^2}{(\sigma^2)^2} + \left(\frac{(g''(\mu) (k - \mu))}{g(\mu)} - \frac{\mu g'(\mu)}{g(\mu)} - \left(\frac{g'(\mu)}{g(\mu)} \right)^2 (k - \mu) \right) \right] \\
&= \pi_{PS}(k; \mu) \left[\left(\frac{g'(\mu)}{g(\mu)} \right)^2 (k - \mu)^2 + \frac{g''(\mu) (k - \mu)}{g(\mu)} - \frac{\mu g'(\mu)}{g(\mu)} - \left(\frac{g'(\mu)}{g(\mu)} \right)^2 (k - \mu) \right].
\end{aligned}$$

.3 Código

Funções de massa das distribuições PS escrita em função da média (mu).

```
PS<-function(y,mu,ps,m){

  erro.ps(mu)

  if(ps=='pois'){ p.s<-dpois(y,mu)}
  if(ps=='geom'){ p.s<-dgeom(y,1/(1+mu))}
  if(ps=='binom'){ p.s<-dbinom(y,m,mu/m)}

  return(p.s)
}
```

Variância da PS escrita em função da média (mu)

```
var.ps<-function(mu,m,ps){

  erro.ps(mu)

  if(ps=='pois'){ sigma.2<-mu }
  if(ps=='geom'){ sigma.2<-mu*(1+mu) }
  if(ps=='binom'){ sigma.2<-mu*(m-mu)/m }

  return(sigma.2)
}
```

Função para calcular a variância da k-MPS

```
var.kmps<-function(mu,m,ps,p,k){

  erro.ps(mu)

  if(ps=='pois'){ sigma.2.kmps<- p*( (1-p)*(k-mu)^2+ mu) }
  if(ps=='geom'){ sigma.2.kmps<- p*( (1-p)*(k-mu)^2+ mu*(1+mu) ) }
  if(ps=='binom'){ sigma.2.kmps<- p*( (1-p)*(k-mu)^2+ mu*(1-mu/m) ) }

  return(sigma.2.kmps)
}
```

```

# Função de massa da k-MPS
dkmps<-function(y,k,mu,p,ps,m){

  d.v<-vector()

  for(i in length(y)) d.v<-c(d.v,(1 - p)*Ik(y,k) + p*PS(y,mu,ps,m))

  return(d.v)
}

# --- Gerador de Número Pseudo-Aleatório da Distribuição k-MPS(mu,p) -----#
# n : Tamanho da Amostra
# k : Ponto de modificação
# p : Parâmetro de modificação
# mu: Parâmetro de Média da PS
# m : Parâmetro referente a distribuição Binomial

rkmps<-function(n,k,p,mu,ps,m){

  f.erro(k,mu,p,ps,m) # verifica as restrições para mu, k , p e m
  s<-s.start(ps)      # inicia s como o menor valor da reta suporte.
  v.kmps<-vector()

  for(i in 1:n){

    y<-s ; py<-dkmps(y,k,mu,p,ps,m) ; FY<-py

    u <- runif(1)

    while(u > FY){

      y = y + 1
      FY = FY + dkmps(y,k,mu,p,ps,m)
    }

    v.kmps<-c(v.kmps,y)
  }
}

```

```
    return(v.kmps)  
}
```


REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **Automatic Control, IEEE Transactions on**, Ieee, v. 19, n. 6, p. 716–723, 1974. Citado na página [56](#).

ANGELL, J.; KORSHOVER, J. Estimate of global temperature variations in the 100-30 mb layer between 1958 and 1977. **Monthly Weather Review**, v. 106, n. 10, p. 1422–1432, 1978. Citado na página [79](#).

BAYARRI, M.; BERGER, J. O.; DATTA, G. S. *et al.* Objective bayes testing of poisson versus inflated poisson models. In: **Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh**. [S.l.]: Institute of Mathematical Statistics, 2008. p. 105–121. Citado na página [17](#).

CONCEIÇÃO, K. S. **Modelos Séries de Potência Zero-Modificados**. Tese (Tese de Doutorado) — DEs – Universidade Federal de São Carlos, São Carlos – SP, 2013. Citado nas páginas [17](#), [18](#), [19](#), [72](#), [73](#) e [89](#).

CONSUL, P. New class of location-parameter discrete probability distributions and their characterizations. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 19, n. 12, p. 4653–4666, 1990. Citado na página [37](#).

CONSUL, P.; SHENTON, L. Some interesting properties of lagrangian distributions. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 2, n. 3, p. 263–272, 1973. Citado na página [37](#).

CONSUL, P. C.; FAMOYE, F. **Lagrangian Probability Distributions**. Boston: Birkhäuser, 2006. Citado na página [37](#).

CORDEIRO, G. M.; ANDRADE, M. G.; CASTRO, M. de. Power series generalized nonlinear models. **Computational Statistics & Data Analysis**, North-Holland, v. 53, n. 4, p. 1155–1166, 2009. Citado na página [17](#).

FROME, E. L.; CHECKOWAY, H. Use of poisson regression models in estimating incidence rates and ratios. **American Journal of Epidemiology**, Oxford Univ Press, v. 121, n. 2, p. 309–323, 1985. Citado na página [17](#).

GUPTA, P. L.; GUPTA, R. C.; TRIPATHI, R. C. Inflated modified power series distributions with applications. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 24, n. 9, p. 2355–2374, 1995. Citado na página [18](#).

GUPTA, R. C. Modified power series distribution and some of its applications. **Sankhyā: The Indian Journal of Statistics, Series B**, JSTOR, p. 288–298, 1974. Citado nas páginas [7](#), [17](#), [23](#) e [37](#).

JANI, P. On modified power series distributions. **Metron**, v. 36, p. 173–186, 1978. Citado na página [17](#).

- KHATRI, C. On certain properties of power-series distributions. **Biometrika**, Oxford University Press, Biometrika Trust, v. 46, n. 3/4, p. 486–490, 1959. Citado na página 37.
- MURAT, M.; SZYNAL, D. Non-zero inflated modified power series distributions. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 27, n. 12, p. 3047–3064, 1998. Citado nas páginas 18 e 47.
- PATIL, G. Certain properties of the generalized power series distribution ii. **Annals of the Institute of Statistical Mathematics**, Springer, v. 14, n. 1, p. 179–182, 1962. Citado na página 17.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>. Citado na página 19.
- RODRIGUES, C. **Distribuições em série de potências modificadas inflacionadas e distribuição Weibull binominal negativa**. Tese (Doutorado) — Universidade de São Paulo, 2011. Citado na página 18.
- SAITO, M. Y.; RODRIGUES, J. Análise bayesiana de dados de contagem com excesso de zeros e uns. **Rev. Mat. Estat.**, v. 23, n. 1, p. 47–57, 2005. Citado na página 18.
- SCHWARZ, G. *et al.* Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. Citado na página 56.