



**UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”**

MARCO ANTÔNIO MARCHETTI

**DISTRIBUIÇÃO RAYLEIGH DE MARSHALL OLKIN: PROPRIEDADES E
INFERÊNCIA.**

Presidente Prudente

2024

MARCO ANTÔNIO MARCHETTI

DISTRIBUIÇÃO RAYLEIGH DE MARSHALL OLKIN: PROPRIEDADES E
INFERÊNCIA.

Trabalho de Conclusão de Curso de Graduação em Estatística do Departamento de Estatística da Universidade Estadual Paulista "Júlio de Mesquita Filho" como requisito para obtenção do título de Estatístico. Orientador: Prof. Guilherme Aparecido Santos Aguiar.

Presidente Prudente

2024

M317d

Marchetti, Marco Antônio

Distribuição Rayleigh de Marshall-Olkin : propriedades e inferência
/ Marco Antônio Marchetti. -- Presidente Prudente, 2024
33 p. : il., tabs.

Trabalho de conclusão de curso (Bacharelado - Estatística) -
Universidade Estadual Paulista (UNESP), Faculdade de Ciências e
Tecnologia, Presidente Prudente

Orientador: Guilherme Aparecido Santos Aguilar

1. Distribuições de Probabilidade. 2. Análise de Sobrevivência. 3.
Inferência Bayesiana. 4. Estimação de Máxima Verossimilhança. I.
Título.

MARCO ANTÔNIO MARCHETTI

**DISTRIBUIÇÃO RAYLEIGH DE MARSHALL OLKIN: PROPRIEDADES E
INFERÊNCIA.**

Relatório de Final de Trabalho de Conclusão de Curso aprovado como requisito para
obtenção de créditos na disciplina Trabalho de Conclusão do curso de graduação em
Estatística da Faculdade de Ciências e Tecnologia da Unesp, pela seguinte banca
examinadora:

Presidente Prudente, 5 de Dezembro de 2024.

Banca Examinadora:

Prof. Guilherme Aparecido Santos Aguilar
Faculdade de Ciências e Tecnologia - Câmpus de Presidente Prudente - Unesp

Prof. Dr. Fernando Antônio Moala
Faculdade de Ciências e Tecnologia - Câmpus de Presidente Prudente - Unesp

Para José, Mara e Carol.

AGRADECIMENTOS

Agradeço ao Professor Guilherme pela paciência e pelos valiosos ensinamentos, aos amigos que sempre estiveram ao meu lado em todas as situações, e à minha família, por todo o carinho e amor.

RESUMO

A análise de sobrevivência, um dos campos mais importantes e estudados da Estatística, possui uma gama de aplicações e um suficiente número de pesquisas nesse domínio. Partindo dessa perspectiva, as distribuições têm um valor significativo nesta área e a criação de novas disposições é uma aliada forte, uma vez que torna a tarefa de ajustar os dados complexos, deixando-os mais fáceis e, em segundo lugar, fornece as funções de risco nunca abordadas. Neste trabalho, é apresentada a Distribuição Rayleigh de Marshall-Olkin, desenvolvida a partir do método Marshall-Olkin com a adição de um parâmetro à distribuição base. Serão introduzidas a função densidade, função de risco, função acumulada, função sobrevivência e momentos da nova distribuição. Também foram realizadas simulações para diferentes valores de parâmetros para verificação do melhor método de estimação. Neste caso, os métodos utilizados foram Método da Máxima Verossimilhança e Método Bayesiano. Por fim, foi realizado uma aplicação em dois bancos de dados, para efeito de comparação da nova distribuição com outras já conhecidas.

Palavras-chave: Análise de Sobrevivência; Distribuição Rayleigh de Marshall-Olkin; Métodos de estimação.

ABSTRACT

Survival analysis, one of the most important and studied fields in statistics, has a range of applications and a sufficient amount of research in this domain. From this perspective, distributions have a significant value in this area and the creation of new ones is a strong ally, since it makes the task of adjusting complex data easier and, secondly, it provides risk functions that have never been approached before. In this paper, the Marshall-Olkin Rayleigh Distribution is presented, developed from the Marshall-Olkin method with the addition of a parameter to the base distribution. The density function, risk function, cumulative function, survival function and moments of the new distribution will be introduced. Simulations were also carried out for different parameter values to check the best estimation method. In this case, the methods used were the Maximum Likelihood Method and the Bayesian Method. Finally, an application was carried out on two databases in order to compare the new distribution with other known distributions.

Keywords: Survival Analysis; Marshall-Olkin Rayleigh Distribution; Methods of Estimation.

LISTA DE FIGURAS

Figura 1 – Gráficos da função densidade e risco da Distribuição Rayleigh	17
Figura 2 – Gráficos da função acumulada e sobrevivência da Distribuição Rayleigh	18
Figura 3 – Gráfico da função densidade e risco da DRMO	19
Figura 4 – Gráficos da função risco para diferentes valores de α e θ	19
Figura 5 – Gráficos da função acumulada e sobrevivência da DRMO	19
Figura 6 – Histograma dos dados e curva das distribuições ajustadas.	29
Figura 7 – Histograma dos dados e curva das distribuições ajustadas.	30

LISTA DE TABELAS

Tabela 1	–	Resultado da simulação para os parâmetros $\theta = 1$ e $\alpha = 1$		27
Tabela 2	–	Resultado da simulação para os parâmetros $\theta = 2$ e $\alpha = 0.5$		27
Tabela 3	–	Resultado da simulação para os parâmetros $\theta = 0.8$ e $\alpha = 0.02$		27
Tabela 4	–	Resultado da simulação para os parâmetros $\theta = 0.5$ e $\alpha = 1.5$		28
Tabela 5	–	Critérios de seleção do modelo		30
Tabela 6	–	Critérios de seleção do modelo		31

LISTA DE ABREVIATURAS E SIGLAS

DRMO	Distribuição Rayleigh Estendida de Marshall Olkin
------	---

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVO GERAL	13
1.2	OBJETIVOS ESPECÍFICOS	13
1.3	JUSTIFICATIVA	13
2	REFERENCIAL TEÓRICO	15
3	METODOLOGIA	16
4	DISTRIBUIÇÃO RAYLEIGH DE MARSHALL-OLKIN . . .	17
4.1	DISTRIBUIÇÃO RAYLEIGH	17
4.2	DISTRIBUIÇÃO RAYLEIGH ESTENDIDA DE MARSHALL-OLKIN	18
4.2.1	MOMENTOS DA DISTRIBUIÇÃO DRMO	20
4.2.2	FUNÇÃO GERADORA DE MOMENTOS	21
5	ESTIMADORES	22
5.1	MÉTODO DA MÁXIMA VEROSSIMILHANÇA	22
5.2	MÉTODO BAYESIANO	23
6	SIMULAÇÃO	26
6.1	COMPARADOR DE ESTIMADORES	26
6.2	RESULTADOS DAS SIMULAÇÕES	26
7	APLICAÇÃO	29
7.1	TIRO ESPORTIVO	29
7.2	VELOCIDADE DO VENTO	30
8	CONCLUSÃO	32
8.1	CONSIDERAÇÕES FINAIS	32
9	REFERÊNCIAS	33

1 INTRODUÇÃO

Um dos conceitos fundamentais da Estatística é o estudo das distribuições de probabilidade, este é um modelo que descreve a probabilidade de ocorrência de diferentes resultados em um experimento, sendo utilizadas para analisar, prever e consequentemente tomar decisões. Com uma ampla diversidade de aplicações, essa área atrai uma quantidade significativa de pesquisas e interesse, estando entre as mais desenvolvidas e aplicadas em várias outras disciplinas.

A distribuição Rayleigh - proposta pelo engenheiro e físico britânico John William Strutt(1842-1919) – é um caso especial da Distribuição Weibull com dois parâmetros de escala, essa é frequentemente usada nas ciências físicas, na engenharia e na teoria das comunicações.

Beckmann (1964) aborda mais profundamente as propriedades da distribuição e suas generalizações, o que contribui significativamente para o estudo dessa distribuição probabilística. Sendo uma distribuição relacionada com problemas que surgem na propagação de ondas de rádio, suas aplicações abrangem uma variedade de campos, que incluem Imagem por Ressonância Magnética (MRI), precisão de armas, imagens SAR e Oceanografia Física. Ghani (2014) aborda um problema comum em imagens subaquáticas em que a qualidade da imagem é prejudicada devido às propriedades da água, dessa forma ele apresenta métodos relacionados a Distribuição Rayleigh – comumente usada para modelar a intensidade da luz e do ruído em ambientes subaquáticos – como parte do processo de melhora do contraste e qualidade da imagem.

Os extensos estudos no campo das distribuições de probabilidade tornaram a criação de novas distribuições uma aliada importante na pesquisa e desenvolvimento dessa área. As novas distribuições aprimoram os modelos estatísticos, tornando-os mais flexíveis.

Com a complexidade das diversas naturezas de dados atualmente, nem sempre eles vão se encaixar nas distribuições existentes. Dessa forma, as novas distribuições possibilitam uma adaptação mais precisa à natureza específica desses dados, o que as tornam fundamentais em suas aplicações. Junto disso, a inovação na criação das distribuições permite que estatísticos e pesquisadores desenvolvam modelos nunca antes estudados.

Marshall e Olkin (1997) apresentam uma nova maneira de extensão para famílias de distribuições exponenciais e da Weibull. Com diversos estudos e aplicações, é evidente as contribuições do método para pesquisas do campo de Estatística. Meintanis (2007) estuda a adequação de dados a essas distribuições, explorando sua utilidade e avaliando o quão bem elas se ajustam a conjuntos de dados reais.

Outro elemento de extrema importância na Estatística é a utilização dos estimadores, que possuem um papel crucial na obtenção de conclusões informativas e confiáveis a partir de uma amostra de dados. Tais estimadores são escolhidos com base em critérios como eficiência, consistência e viés.

Existem vários métodos de estimação, alguns dos mais conhecidos incluem a Estimação de Máxima Verossimilhança; o Métodos dos Momentos; Estimadores de Mínimos Quadrados; e o Método Bayesiano. Kundu e Raqab (2005) exploram diferentes estimadores e tamanhos de amostras para diferentes valores de parâmetros na Distribuição Rayleigh Generalizada.

Este estudo possibilita identificar o estimador mais adequado para variadas situações, oferecendo uma compreensão valiosa sobre como essas técnicas de estimação se comportam sob diferentes circunstâncias.

1.1 OBJETIVO GERAL

O objetivo geral do trabalho consiste na construção de uma nova distribuição estatística, com o intuito de aprimorar o modelo já estabelecido da Distribuição Rayleigh. Nosso estudo visa fornecer maior flexibilidade e capacidade de ajuste aos dados e, consequentemente, capacitando-a a representar uma variedade mais abrangente de situações.

Para atingirmos esse objetivo, investigaremos o comportamento da nova distribuição aplicaremos métodos de estimação em seus parâmetros, com o intuito de identificar os cenários nos quais ela demonstra o melhor ajuste e a maior eficácia.

1.2 OBJETIVOS ESPECÍFICOS

Para atingir o nosso objetivo, estabelecemos os seguintes objetivos específicos a serem alcançados:

1. Utilização do Método Marshall-Olkin para adicionar um novo parâmetro à Distribuição Rayleigh.
2. Estudo de comportamento da nova distribuição: serão utilizadas as funções de densidade e de risco; faremos uso das visualizações gráficas como curvas de densidade; gráficos de sobrevivência para destacar as características distintivas da nova distribuição.
3. Implementação dos métodos de estimação propostos a fim de determinar os valores dos parâmetros, avaliando seu desempenho em termos de ajuste e eficácia, com o propósito de identificar o método mais eficaz.

1.3 JUSTIFICATIVA

Desenvolver uma nova distribuição estatística amplia as opções disponíveis para modelar uma ampla gama de fenômenos, pois um novo modelo estatístico pode apresentar curvas de probabilidade e comportamentos estatísticos únicos e distintos em comparação com as distribuições já existentes. Dessa forma nosso trabalho se relaciona com a nona e a

décima sétima ODS's da ONU, mais especificamente: fortalecendo a pesquisa científica e a disseminação de tecnologias ambientalmente corretas para os países em desenvolvimento.

Outro ponto relevante é que, mesmo com a introdução de uma nova distribuição, esta proposta mantém a simplicidade, pois requer apenas dois parâmetros para a estimação. Essa simplicidade facilita a estimação dos parâmetros em comparação com modelos mais complexos, tornando-o uma opção atraente em cenários em que a eficiência e a facilidade de uso são fundamentais.

2 REFERENCIAL TEÓRICO

Os materiais estudados para a formulação desse trabalho envolvem principalmente o campo da criação de novas distribuições, bem como os métodos de estimação necessários para sua aplicabilidade prática.

O primeiro artigo analisado foi “Rayleigh distribution and its generalizations” (Beckmann, 1964), sua utilização é essencial pois trata do estudo da Distribuição Rayleigh e suas generalizações. O autor analisa as aplicações da distribuição, o que fornece uma contribuição significativa para o estudo dessa distribuição probabilística. Com o objetivo de adicionar um novo parâmetro a Distribuição de Rayleigh utilizando o Método Marshall-Olkin, uma outra grande referência foi “A new method for adding a parameter to a Family of distributions with application to the exponential and Weibull families” (Marshall e Olkin, 1997) porque os autores apresentam detalhes de uma metodologia, ilustrando também a aplicação as distribuições exponenciais e de Weibull, estudando sua utilidade.

Para maior compreensão dessa família de distribuições foi utilizado o artigo “Test of fit for Marshall-Olkin distributions with applications” (Meintanis, 2007) para estudo da questão de adequação de dados a distribuições Marshall-Olkin, enfatizando sua aplicação prática. É apresentada as aplicações em diversos campos, em áreas como ciências sociais, finanças e engenharia, o que destaca, assim, a relevância dessas distribuições em cenários do mundo real.

No campo dos métodos de estimação o artigo “Zero-truncated poisson exponentiated gamma distribution: Application and estimation methods” (Aguilar, Moala e Cordeiro, 2019) foi utilizado principalmente o estudo de técnicas de estimações aplicadas na nova distribuição proposta. Por último, o artigo “Generalized Rayleigh distribution: different methods of estimations” (Kundu, 2005) foi essencial para o estudo de outra generalização da Distribuição de Rayleigh, tendo como foco principal a discussão de diferentes métodos de estimação para os parâmetros dessa distribuição generalizada.

3 METODOLOGIA

A metodologia se concentrará na aplicação do método Marshall-Olkin para a incorporação de um novo parâmetro na distribuição Rayleigh. Posteriormente, estudaremos o comportamento das funções de densidade e de risco dessa nova distribuição, seguido pela implementação de métodos de estimação e sua aplicação em um conjunto de dados do mundo real.

Como pode servir como concorrente para famílias de distribuições de sobrevivência, o método Marshall-Olkin utiliza a Função de Distribuição Acumulada e sua Função de Sobrevivência para adicionar um novo parâmetro a distribuição, o que contribui na sua flexibilidade e na construção de modelos bivariados. Introduzido por Marshall e Olkin (1997), o método é encontrado em diversas pesquisas no campo da criação de novas distribuições, comumente referidas como Distribuições Estendidas Marshall-Olkin.

No estudo de comportamento da nova distribuição, serão utilizadas as funções de densidade e de risco. Para aprimorar nossa compreensão e distinguir a nova distribuição de distribuições preexistentes, recorreremos a ferramentas visuais, como curvas de densidade e gráficos de sobrevivência, possibilitando, assim, uma verificação da capacidade de modelar os dados em comparação com distribuições já estabelecidas.

Os métodos de estimação desempenham um papel essencial na análise e aplicação da nova distribuição, esses serão cruciais para determinar os valores dos parâmetros, o que possibilita a adaptação da distribuição proposta aos dados em estudo. Serão empregados os seguintes métodos de estimação:

- Método de Máxima Verossimilhança: obtém a estimativa mais verossímil dentro de uma amostra para o parâmetro populacional desconhecido. Os parâmetros são estimados a partir da maximização da função de verossimilhança.
- Método Bayesiano: Abordagem que se baseia no teorema de Bayes para estimação dos parâmetros do modelo, ela se difere da Inferência Clássica pois utiliza conhecimentos prévios como informação para estimar. O método considera o parâmetro como variável aleatória e irá possuir uma distribuição.

Após a realização dos métodos de estimação, utilizando o software estatístico R, teremos a capacidade de avaliar qual deles demonstrou um desempenho superior, levando em consideração tanto a eficácia como as aplicações práticas da distribuição.

4 DISTRIBUIÇÃO RAYLEIGH DE MARSHALL-OLKIN

4.1 DISTRIBUIÇÃO RAYLEIGH

Introduzida por John William Strutt, a Distribuição Rayleigh é um caso especial da Weibull, ela representa a soma de um grande número de vetores coplanares com amplitudes aleatórias e fases uniformemente distribuídas. Beckmann (1964) aborda essa condição com problemas práticos, oferecendo análises que ajudam na compreensão do comportamento dessa distribuição.

Sendo amplamente utilizada em problemas de propagação de ondas de rádio, sua importância permeia nas áreas de engenharia, teoria das comunicações e oceanografia física. Existem diversas extensões da distribuição feitas por pesquisadores, Kuruoglu (2004) estuda a modelagem de imagens de radar de abertura sintética (SAR) usando uma generalização da distribuição de Rayleigh, permitindo um ajuste mais flexível das imagens SAR e demonstrando melhor desempenho que outros modelos existentes. Sarti (2005), propõe um modelo e algoritmo para a segmentação de imagens de ultrassom, a distribuição Rayleigh é utilizada na modelagem no comportamento dos níveis de cinza das imagens.

As funções de densidade, acumulada e sobrevivência da Rayleigh possuem as formas, respectivamente

$$f(x|\theta) = \frac{x}{\theta^2} e^{-x^2/(2\theta^2)}, \quad x > 0, \quad \theta > 0 \quad 4.1$$

$$F(x|\theta) = 1 - e^{-x^2/(2\theta^2)} \quad 4.2$$

$$S(x|\theta) = e^{-x^2/(2\theta^2)} \quad 4.3$$

sendo θ o parâmetro de escala da distribuição. Os gráficos das funções de densidade e de risco da Distribuição Rayleigh são dadas pela figura 1 e as funções acumulada e sobrevivência são dadas pela figura 2.

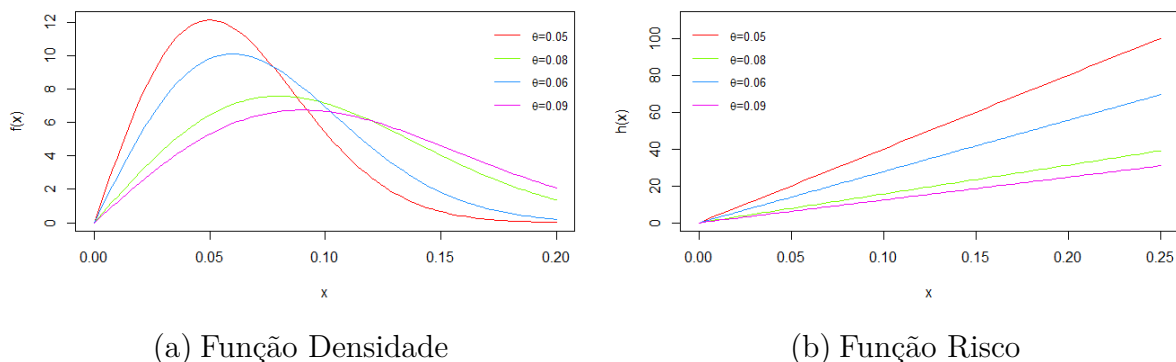
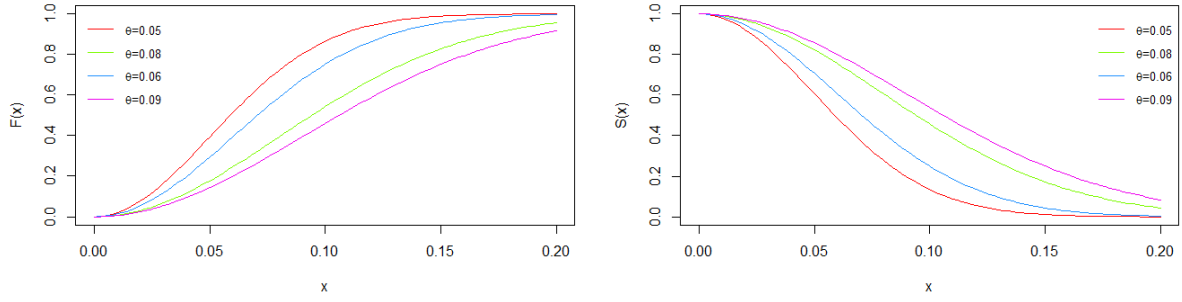


Figura 1 – Gráficos da função densidade e risco da Distribuição Rayleigh



(a) Função Acumulada

(b) Função Sobrevivência

Figura 2 – Gráficos da função acumulada e sobrevivência da Distribuição Rayleigh

4.2 DISTRIBUIÇÃO RAYLEIGH ESTENDIDA DE MARSHALL-OLKIN

Marshall e Olkin (1997) propuseram um método para obter extensões de distribuições de probabilidade, introduzindo um novo parâmetro $\alpha > 0$, utilizando as funções de distribuição acumulada e função de sobrevivência da distribuição base.

Com a adição de um novo parâmetro, o método expande a flexibilidade da distribuição, contribuindo com a extensão de distribuições existentes que competirão com famílias de distribuições amplamente utilizadas. Meintanis(2007) testa a adequação de distribuições Marshall-Olkin com aplicações do método em dados de confiabilidade. O trabalho também apresenta o uso dessas distribuições em dados esportivos, mostrando assim sua importância e versatilidade em uma variedade de contextos.

Sendo $f(x)$ e $h(x)$ as funções de densidade e risco da distribuição base respectivamente, a nova função de densidade de probabilidade e de risco, serão dadas respectivamente por:

$$f * (x|\theta, \alpha) = \frac{\alpha f(x|\theta)}{[1 - \bar{\alpha} S(x|\theta)]^2} ; \quad h * (x|\theta, \alpha) = \frac{h(x|\theta)}{1 - \bar{\alpha} S(x|\theta)}. \quad 4.4$$

Então, a nova distribuição será obtida da seguinte maneira:

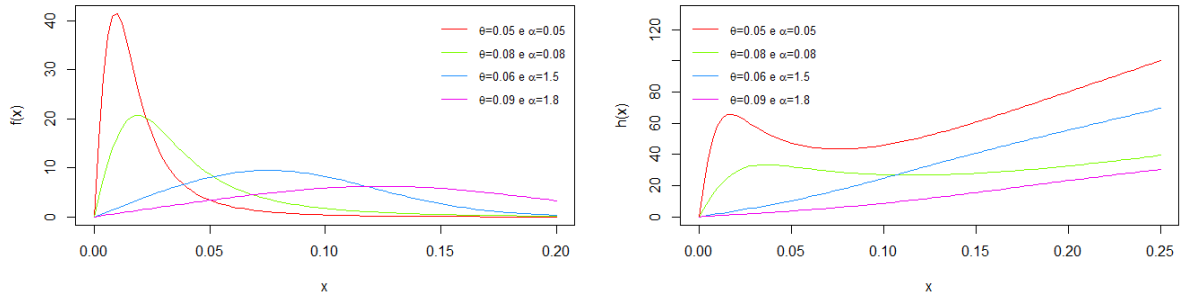
$$f * (x|\theta, \alpha) = \frac{\alpha f(x|\theta)}{[1 - \bar{\alpha} S(x|\theta)]^2} = \frac{\alpha \frac{x}{\theta^2} e^{-x^2/(2\theta^2)}}{[1 - \bar{\alpha} (e^{-x^2/(2\theta^2)})]^2}, \quad x > 0, \quad \theta > 0, \quad \alpha > 0. \quad 4.5$$

Sendo a função de risco da Distribuição Rayleigh dada por :

$$h(x|\theta) = \frac{f(x|\theta)}{S(x|\theta)} = \frac{\frac{x}{\theta^2} e^{-x^2/(2\theta^2)}}{e^{-x^2/(2\theta^2)}} = \frac{x}{\theta^2}. \quad 4.6$$

Logo, a função de risco da Distribuição Rayleigh Estendida de Marshall Olkin será:

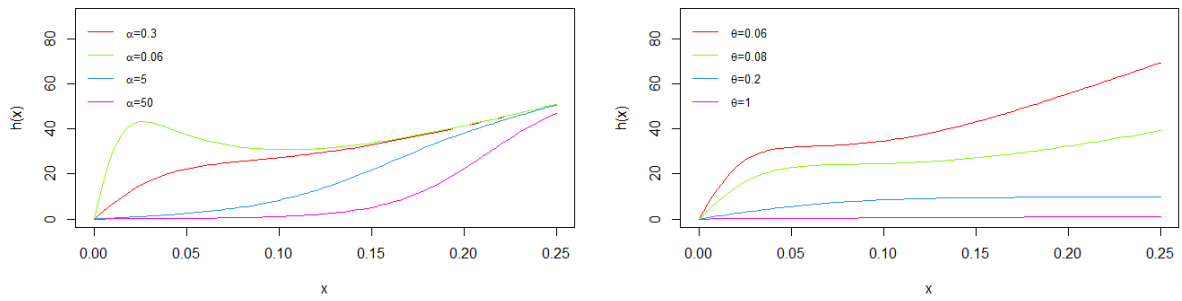
$$h * (x|\theta, \alpha) = \frac{h(x|\theta)}{1 - \bar{\alpha} S(x|\theta)} = \frac{\frac{x}{\theta^2} e^{-x^2/(2\theta^2)}}{(e^{-x^2/(2\theta^2)})(1 - \bar{\alpha} (e^{-x^2/(2\theta^2)}))} = \frac{x}{\theta^2 (1 - \bar{\alpha} (e^{-x^2/(2\theta^2)}))}. \quad 4.7$$



(a) Função Densidade

(b) Função Risco

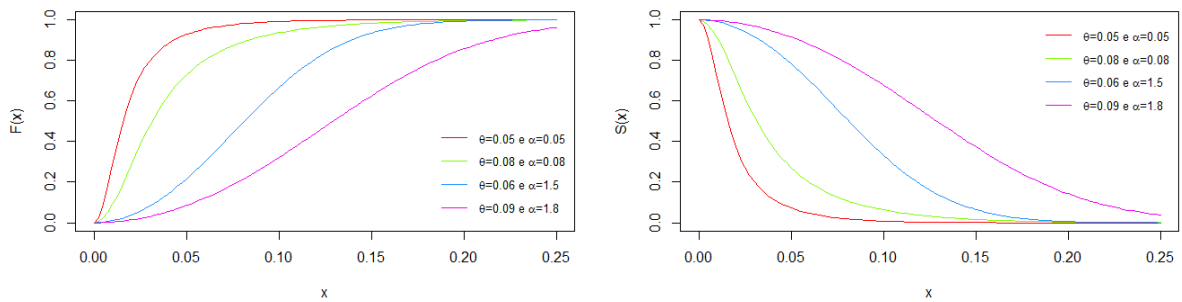
Figura 3 – Gráfico da função densidade e risco da DRMO

(a) $\theta = 0.07$ (b) $\alpha = 0.2$ Figura 4 – Gráficos da função risco para diferentes valores de α e θ .

A função distribuição e a função de sobrevivência são dadas por:

$$F(x|\theta, \alpha) = \frac{F(x|\theta)}{1 - \bar{\alpha}S(x|\theta)} = \frac{1 - e^{-x^2/(2\theta^2)}}{1 - \bar{\alpha}(e^{-x^2/(2\theta^2)})} \quad 4.8$$

$$S(x|\theta, \alpha) = \frac{\alpha S(x|\theta)}{1 - \bar{\alpha}S(x|\theta)} = \frac{\alpha(e^{-x^2/(2\theta^2)})}{1 - \bar{\alpha}(e^{-x^2/(2\theta^2)})} \quad 4.9$$



(a) Função Acumulada

(b) Função Sobrevivência

Figura 5 – Gráficos da função acumulada e sobrevivência da DRMO

É possível observar uma melhora na flexibilização da função risco com a adição do parâmetro α . Enquanto a distribuição base possuía um crescimento linear em relação a x , a estendida tem formas mais variadas com crescimento não lineares e comportamentos de pico.

Assim, a DRMO consegue capturar comportamentos de risco que não seriam possíveis com a função de risco linear simples, proporcionando uma boa vantagem na modelagem de riscos e comportamentos de sobrevivência.

4.2.1 MOMENTOS DA DISTRIBUIÇÃO DRMO

Para obtermos o n -ésimo momento ou momento de ordem n de uma variável aleatória X , utilizamos a expressão $E(X^n)$. Dessa forma, podemos obter a média e a variância da DRMO, caracterizando o seu primeiro e segundo momento.

Então, desejamos calcular:

$$E(X^n) = \int_0^\infty x^n f(x) dx \quad 4.10$$

Para este cálculo, utilizaremos o método de expansão de Taylor. Primeiramente definindo:

$$(1 - \bar{\alpha}z)^2 = \sum_{k=0}^{\infty} (k+1)\alpha^k x^k \quad 4.11$$

Sendo $\bar{\alpha} = (1 - \alpha)$ e $z = e^{\frac{-x^2}{2\theta^2}}$. Assim, temos :

$$\begin{aligned} f(x|\theta, \alpha) &= \left(\frac{\alpha}{\theta^2}\right) x \sum_{k=0}^{\infty} (k+1)(1-\alpha)^k e^{\frac{-kx^2}{2\theta^2}} e^{\frac{-x^2}{2\theta^2}} \\ &= \left(\frac{\alpha}{\theta^2}\right) x \sum_{k=0}^{\infty} (k+1)(1-\alpha)^k e^{\frac{-x^2}{2\theta^2}(1+k)} \end{aligned} \quad 4.12$$

Logo, $E(X^n)$ será dado por:

$$E(X^n) = \left(\frac{\alpha}{\theta^2}\right) \sum_{k=0}^{\infty} (k+1)(1-\alpha)^k \int_0^\infty x^{1+n} e^{\frac{-x^2}{2\theta^2}(1+k)} dx \quad 4.13$$

Utilizando o método de substituição, iremos obter:

$$\left(\frac{2\theta^2}{k+1}\right)^{\frac{1+n}{2}} \frac{\theta^2}{\sqrt{2\theta^2(k+1)}} \int_0^\infty \frac{u^{\frac{1+n}{2}}}{u^{1/2}} e^{-u} du \quad 4.14$$

Aplicando a integral definida da função Gamma, temos por fim:

$$\begin{aligned}
E(X^n) &= \left(\frac{2\theta^2}{k+1} \right)^{\frac{1+n}{2}} \frac{\theta^2}{\sqrt{2\theta^2(k+1)}} \int_0^\infty u^{n/2} e^{-u} du \\
&= \left(\frac{2\theta^2}{k+1} \right)^{\frac{1+n}{2}} \frac{\theta^2}{\sqrt{2\theta^2(k+1)}} \Gamma\left(\frac{n+2}{2}\right)
\end{aligned} \tag{4.15}$$

Para obter a média da DRMO, precisamos do momento de primeira ordem, ou seja, quando $n = 1$. Sendo $k^* = \left(\frac{2\theta^2}{k+1} \right)^{\frac{1+n}{2}} \frac{\theta^2}{\sqrt{2\theta^2(k+1)}}$ e k representando o índice do somatório, temos:

$$E(X) = k^* \Gamma\left(\frac{3}{2}\right) = k^* \frac{\sqrt{\pi}}{2} \tag{4.16}$$

Sendo o momento de segunda ordem, definido por $k^* \Gamma(2)$ e utilizando a definição de variância. Teremos a equação:

$$Var(X) = k^* - \frac{(k^*)^2 \pi}{4} \tag{4.17}$$

Portanto, a partir dos parâmetros θ e α podemos determinar a constante k^* e consequentemente as expressões de média e variância.

4.2.2 FUNÇÃO GERADORA DE MOMENTOS

Para encontrar todos os momentos da variável aleatória X podemos utilizar a expressão definida como:

$$M_X(t) = E(e^{tX}) \tag{4.18}$$

Que por meio da expansão em série, podemos escrever da seguinte maneira:

$$M_X(t) = E(e^{tX}) = E\left(\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}\right) = \sum_{n=0}^{\infty} \frac{t^n E(X^n)}{n!} \tag{4.19}$$

Com este resultado e pela equação (4.15), a fgm será dada por:

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} k^* \Gamma\left(\frac{n+2}{2}\right) \tag{4.20}$$

5 ESTIMADORES

Serão apresentados os dois métodos de estimação para estimar os parâmetros θ e α da distribuição Estendida Marshall-Olkin, sendo eles: o Método de Máxima Verossimilhança e o método de estimação Bayesiana.

5.1 MÉTODO DA MÁXIMA VEROSSIMILHANÇA

O método da máxima verossimilhança é o método mais tradicional de estimação estatística, sendo um processo de obtenção da informação sobre um vetor de parâmetro θ a partir de um ponto y do espaço amostral.

Assume-se que a amostra é representativa da população e escolhe-se o estimador o qual valor do parâmetro maximiza a função $L(\theta|y)$. Os estimadores de máxima verossimilhança apresentam propriedades desejáveis e são usados para obtenção de intervalos de confiança para os parâmetros do modelo.

Seja $Y_1 \cdots Y_n$ uma amostra aleatória com valores observados $y_1 \cdots y_n$ da DRMO $f * (y|\theta, \alpha)$ com vetor de parâmetros $(\theta, \alpha)'$. A função de verossimilhança é dada por:

$$L(\theta, \alpha|y) = \prod_{i=1}^n f(y_i|\theta, \alpha) \quad 5.1$$

substituindo a expressão para $f(y|\theta, \alpha)$:

$$L(\theta, \alpha|y) = \prod_{i=1}^n \frac{\alpha \frac{y_i}{\theta^2} e^{-y_i^2/(2\theta^2)}}{[1 - \bar{\alpha}(e^{-y_i^2/(2\theta^2)})]^2} \quad 5.2$$

Sendo $\bar{\alpha} = (1 - \alpha)$, a função da verossimilhança da distribuição fica:

$$L(\theta, \alpha) = \left(\frac{\alpha}{\theta^2}\right)^n \prod_{i=1}^n \frac{y_i e^{-y_i^2/(2\theta^2)}}{(1 - e^{-y_i^2/(2\theta^2)} + \alpha e^{-y_i^2/(2\theta^2)})^2} \quad 5.3$$

com função log-verossimilhança:

$$\log L(\theta, \alpha) = n(\log \alpha - 2 \log \theta) + \sum_{i=1}^n \left(\log y_i - \frac{-y_i^2}{2\theta^2} - 2 \log(1 - e^{-y_i^2/(2\theta^2)} + \alpha e^{-y_i^2/(2\theta^2)}) \right) \quad 5.4$$

em que $\alpha > 0$ e $-\infty < y < \infty$.

A partir de (5.4), podemos obter as derivadas em relação aos parâmetros θ e α :

$$\frac{\partial \log L(\theta, \alpha)}{\partial \theta} = -\frac{2n}{\theta} + \sum_{i=1}^n \left(\frac{y_i^2}{\theta^3} + 2 \cdot \frac{y_i^2 e^{-y_i^2/(2\theta^2)}}{\theta^3 (1 - e^{-y_i^2/(2\theta^2)} + \alpha e^{-y_i^2/(2\theta^2)})} \right) \quad 5.5$$

$$\frac{\partial \log L(\theta, \alpha)}{\partial \alpha} = \frac{n}{\alpha} - 2 \sum_{i=1}^n \frac{e^{-y_i^2/(2\theta^2)}}{1 - e^{-y_i^2/(2\theta^2)} + \alpha e^{-y_i^2/(2\theta^2)}} \quad 5.6$$

A matriz de informação de Fisher, é dada por $I(\Theta)$.

$$I(\Theta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} \quad 5.7$$

Sendo a sua inversa dada por

$$I^{-1}(\Theta) = \begin{pmatrix} \text{var}(\hat{\alpha}_{EMV}) & \text{cov}(\hat{\alpha}_{EMV}\hat{\theta}_{EMV}) \\ \text{cov}(\hat{\theta}_{EMV}\hat{\alpha}_{EMV}) & \text{var}(\hat{\theta}_{EMV}) \end{pmatrix} \quad 5.8$$

Sendo assim, vamos determinar a derivada de segunda ordem dos parâmetros θ e α .

Primeiro, para a derivada de segunda ordem de θ , definimos :

$$u(\theta) = y_i^2 e^{-y_i^2/(2\theta^2)} \quad ; \quad v(\theta) = \theta^3(1 - e^{-y_i^2/(2\theta^2)} + \alpha e^{-y_i^2/(2\theta^2)})$$

$$u'(\theta) = \frac{-y_i^4}{\theta^3} e^{-y_i^2/(2\theta^2)} \quad ; \quad v'(\theta) = 3\theta^2(1 - e^{-y_i^2/(2\theta^2)} + \alpha e^{-y_i^2/(2\theta^2)}) + \frac{\alpha y_i^4 e^{-y_i^2/(2\theta^2)}}{\theta^6}$$

Temos então:

$$\frac{\partial^2 \log L(\theta, \alpha)}{\partial \theta^2} = \frac{2n}{\theta^2} - \sum_{i=1}^n \left(-\frac{3y_i^2}{\theta^4} + 2 \cdot \frac{u'(\theta)v(\theta) - u(\theta)v'(\theta)}{v(\theta)^2} \right) \quad 5.9$$

$$\frac{\partial^2 \log L(\theta, \alpha)}{\partial \alpha^2} = -\frac{n}{\alpha^2} + 2 \sum_{i=1}^n \frac{e^{-y_i^2/\theta^2}}{(1 - e^{-y_i^2/(2\theta^2)} + \alpha e^{-y_i^2/(2\theta^2)})^2} \quad 5.10$$

Para a derivada mista, derivaremos em relação a θ a equação (5.10). Sendo assim, definimos:

$$a(\theta) = e^{-y_i^2/(2\theta^2)} \quad ; \quad b(\theta) = 1 - e^{-y_i^2/(2\theta^2)} + \alpha e^{-y_i^2/(2\theta^2)}$$

$$a'(\theta) = \frac{y_i^2}{\theta^3} e^{-y_i^2/(2\theta^2)} \quad ; \quad b'(\theta) = \frac{y_i^2}{\theta^3} e^{-y_i^2/(2\theta^2)}(1 + \alpha)$$

Logo,

$$\frac{\partial^2 \log L(\theta, \alpha)}{\partial \theta \partial \alpha} = -2 \sum_{i=1}^n \frac{a'(\theta)b(\theta) - a(\theta)b'(\theta)}{b(\theta)^2} \quad 5.11$$

5.2 MÉTODO BAYESIANO

Na estimação Bayesiana, a estimativa dos parâmetros não depende apenas dos dados, levando em consideração a opinião do especialista que nos fornece uma informação a priori sobre o assunto estudado. Do ponto de vista Bayesiano, θ é incerto e essa incerteza é representada através de modelos probabilísticos para θ .

Os parâmetros são tratados como variáveis aleatórias e assumem uma distribuição à priori, eles serão atualizados pelo Teorema de Bayes utilizando a função de Verossimilhança proveniente dos dados. Utilizamos a seguinte equação :

$$p(\theta, y) = \frac{p(\theta)p(x|\theta)}{p(x)} = \frac{p(\theta)p(x|\theta)}{\int p(\theta)p(x|\theta)d\theta} \quad 5.12$$

Dessa forma, teremos a distribuição a posteriori dada por :

$$p(\theta|x) \propto \pi(\theta)L(\theta|x), \quad 5.13$$

sendo $\pi(\theta)$ a distribuição à priori e $L(\theta|x)$ a função de verossimilhança dos dados.

Para os parâmetros desconhecidos da DRMO $\sim (\theta, \alpha)$ será assumido que θ e α tenham como distribuições a priori independentes a Gama:

$$\begin{aligned} g(\theta) &\propto \theta^{a-1}e^{-\theta b}, \quad \theta > 0 \\ g(\alpha) &\propto \alpha^{c-1}e^{-\alpha d}, \quad \alpha > 0 \end{aligned}$$

Sendo a, b, c e d hiperparâmetros conhecidos e não negativos.

Utilizamos o algoritmo Metropolis-Hastings para obtenção dos estimadores. Este é um método alternativo para casos onde é difícil a obtenção de forma analítica ou numérica de $f(x)$.

Sendo generalizado por Hastings (1970) e originalmente proposto por Nicholas e Rosenbluth (1953), o algoritmo gera uma amostra $X_1, X_2, \dots, X_m \sim f(x)$. Com uma amostra suficientemente grande, podemos calcular a média, variância ou qualquer outra característica da $f(x)$.

Ao inicializar o MCMC, utilizamos uma distribuição de probabilidade auxiliar $q(\theta, \cdot)$ para gerar possíveis valores da distribuição a posteriori. O valor gerado será aceito, se cumprir a condição abaixo.

$$\alpha(\theta^{(t)}, \theta^{(*)}) = \min \left[1; \frac{p(\theta^{(*)})q(\theta^{(t)}|\theta^{(*)})}{p(\theta^{(t)})q(\theta^{(*)}|\theta^{(t)})} \right] \geq u(0, 1). \quad 5.14$$

Em seguida, serão realizados os seguintes passos :

- i) Introduza a distribuição auxiliar que irá gerar possíveis valores a θ a posteriori;
- ii) Inicie o contador $t = 1$;
- iii) Escolha um valor inicial para $\theta^{(t)}$;
- iv) Gere um valor $\theta^{(t)}$ utilizando $q(\theta, \cdot)$;
- v) Teste a condição $\alpha(\theta^{(t)}, \theta^{(*)}) \geq u(0, 1)$;
- vi) Se a condição for satisfeita: $\theta^{(t+1)} = \theta^{(*)}$, $t = t + 1$ e volte ao passo iv;
- vii) Se a condição não for satisfeita: $\theta^{(t+1)} = \theta^{(t)}$, $t = t + 1$ e volte ao passo iv;

viii) Repita o procedimento até alcançar a convergência.

O algoritmo foi desenvolvido em linguagem de programação R, para cada uma das 1000 amostras foram realizadas 15000 iterações, descartando as 5000 primeiras para o aquecimento inicial (burn-in), e o resultado utilizado foi a média dos valores gerados para θ e α .

6 SIMULAÇÃO

Para obter amostras da DRMO, utilizamos o método da transformação inversa. Esta técnica consiste na geração de um vetor u de tamanho 200000 tal que $u \sim U(0, 1)$, em seguida encontraremos suas raízes para $F(x) = u \rightarrow F^{-1}(u)$, utilizando o método numérico Newton-Raphson.

Este método iterativo é definido pela iteração $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$, sendo o nosso caso $f(x) = F(x) - u$. Assim, as raízes serão armazenadas em uma matriz X de 200 linhas e 1000 colunas (número de amostras).

Foram geradas $B=1000$ amostras da DRMO para diferentes valores dos parâmetros e para quatro diferentes tamanhos amostrais, são eles $n = 20, 50, 100$ e 200 .

6.1 COMPARADOR DE ESTIMADORES

Para avaliar o desempenho dos métodos de estimação mencionados, foi calculado para cada uma das estimativas o vício (6.1), o erro absoluto médio (6.2), o erro quadrático médio (6.3) e a diferença absoluta média entre as funções de distribuições verdadeira e estimada (6.4).

$$BIAS(\hat{\theta}) = \frac{1}{B} \sum_{i=1}^B (\hat{\theta}_i - \theta) ; \quad BIAS(\hat{\alpha}) = \frac{1}{B} \sum_{i=1}^B (\hat{\alpha}_i - \alpha) \quad 6.1$$

$$MAE(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n |\hat{\theta}_i - \theta| ; \quad MAE(\hat{\alpha}) = \frac{1}{n} \sum_{i=1}^n |\hat{\alpha}_i - \alpha| \quad 6.2$$

$$RMSE(\hat{\theta}) = \sqrt{\frac{1}{B} \sum_{i=1}^B (\hat{\theta}_i - \theta)^2}; \quad RMSE(\hat{\alpha}) = \sqrt{\frac{1}{B} \sum_{i=1}^B (\hat{\alpha}_i - \alpha)^2} \quad 6.3$$

$$D_{abs} = \frac{1}{Bn} \sum_{i=1}^B \sum_{j=1}^n \left| F(y_{ij}|\theta, \alpha) - F(y_{ij}|\hat{\theta}, \hat{\alpha}) \right| \quad 6.4$$

6.2 RESULTADOS DAS SIMULAÇÕES

Nas Tabelas de 1 até 4 estão os valores dos parâmetros obtidos pelos estimadores de Máxima Verossimilhança e Bayesiano.

Podemos observar que todas as tabelas apresentaram comportamentos semelhantes, para todos os pares de valores de θ e α em ambos métodos de estimação. Enquanto o tamanho de amostra aumenta, os estimadores produzem estimativas mais precisas e com menos variação. A estimação Bayesiana demonstrou superioridade em estimar α , já o parâmetro θ obteve melhor desempenho quando estimado pelo método de Máxima Verossimilhança.

Tabela 1 – Resultado da simulação para os parâmetros $\theta = 1$ e $\alpha = 1$

Método	n	BIAS θ	BIAS α	MAE θ	MAE α	RMSE θ	RMSE α	D_{abs}
MV	20	-0.0273	0.7473	0.1753	1.0512	0.2245	1.9597	0.0589
BAYES		0.1777	0.2046	0.3048	0.7356	1.1748	1.1153	0.0575
MV	50	-0.0092	0.2450	0.1162	0.4811	0.1448	0.7022	0.0364
BAYES		0.0298	0.1113	0.1277	0.4331	0.1651	0.6073	0.0363
MV	100	-0.0074	0.1153	0.0819	0.3037	0.1005	0.4014	0.0254
BAYES		0.0090	0.0595	0.0845	0.2885	0.1058	0.3767	0.0253
MV	200	-0.0065	0.0617	0.0565	0.2029	0.0703	0.2664	0.0177
BAYES		0.0015	0.0352	0.0576	0.1989	0.0722	0.2593	0.0178

Os resultados da simulação evidenciam que tanto o Método da Máxima Verossimilhança (MV) quanto o Método Bayesiano (BAYES) possuem a propriedade de consistência de um estimador, pois ambos convergem em probabilidade para o valor real do parâmetro populacional que está sendo estimado.

Com n grande, as estimativas devem apresentar valores próximos, isso não acontece com o critério *BIAS* por conta das diferenças negativas e positivas que podem ocorrer nos cálculos.

Tabela 2 – Resultado da simulação para os parâmetros $\theta = 2$ e $\alpha = 0.5$

Método	n	BIAS θ	BIAS α	MAE θ	MAE α	RMSE θ	RMSE α	D_{abs}
MV	20	-0.0890	0.4201	0.4553	0.5673	0.5658	0.9975	0.0574
BAYES		0.2823	0.1647	0.6144	0.4284	1.0695	0.6990	0.0569
MV	50	-0.0405	0.1502	0.2934	0.2714	0.3666	0.4020	0.0354
BAYES		0.1392	0.0616	0.3842	0.2472	0.7209	0.3460	0.0361
MV	100	-0.0267	0.0751	0.2186	0.1768	0.2718	0.2422	0.0242
BAYES		0.0347	0.0323	0.2264	0.1617	0.2849	0.2088	0.0251
MV	200	-0.0158	0.0347	0.1517	0.1159	0.1905	0.1499	0.0176
BAYES		0.0069	0.0203	0.1486	0.1102	0.1870	0.1435	0.0175

Tabela 3 – Resultado da simulação para os parâmetros $\theta = 0.8$ e $\alpha = 0.02$

Método	n	BIAS θ	BIAS α	MAE θ	MAE α	RMSE θ	RMSE α	D_{abs}
MV	20	-0.1889	0.1618	0.4713	0.1664	0.5970	0.3449	0.0571
BAYES		0.2258	0.0735	0.6332	0.0834	1.1393	0.2213	0.0572
MV	50	-0.1119	0.0539	0.3671	0.0591	0.4581	0.1059	0.0350
BAYES		0.2230	0.0263	0.5254	0.0372	0.8454	0.0742	0.0354
MV	100	-0.0660	0.0244	0.2948	0.0301	0.3618	0.0470	0.0238
BAYES		0.1319	0.0115	0.3708	0.0212	0.5542	0.0347	0.0246
MV	200	-0.0431	0.0127	0.2225	0.0179	0.2780	0.0272	0.0166
BAYES		0.0707	0.0069	0.2705	0.0150	0.3860	0.0225	0.0170

Tabela 4 – Resultado da simulação para os parâmetros $\theta = 0.5$ e $\alpha = 1.5$

Método	n	BIAS θ	BIAS α	MAE θ	MAE α	RMSE θ	RMSE α	D_{abs}
MV	20	-0.0120	1.0854	0.0778	1.5435	0.0991	2.9813	0.0593
BAYES		0.1284	0.2352	0.1838	1.0196	0.9554	1.5137	0.0583
MV	50	-0.0040	0.3495	0.0511	0.6994	0.0637	1.0278	0.0366
BAYES		0.0118	0.1527	0.0548	0.6178	0.0705	0.8593	0.0363
MV	100	-0.0033	0.1644	0.0358	0.4412	0.0441	0.5857	0.0256
BAYES		0.0040	0.0813	0.0373	0.4223	0.0464	0.5439	0.0255
MV	200	-0.0029	0.0876	0.0248	0.2955	0.0309	0.3884	0.0179
BAYES		0.0005	0.0503	0.0253	0.2912	0.0317	0.3764	0.0180

7 APLICAÇÃO

Para testar o ajustamento da nova distribuição DRMO, foi realizada sua aplicação em dois conjuntos de dados diferentes, onde também houve a comparação da distribuição com outras já existentes.

7.1 TIRO ESPORTIVO

A base de dados utilizada é proveniente de uma iniciação científica de análise estatística de padrões de tiro esportivo, os dados foram coletados em um clube de tiro de Presidente Prudente - SP. A variável que iremos utilizar é a distância dos tiros em relação ao centro do alvo.

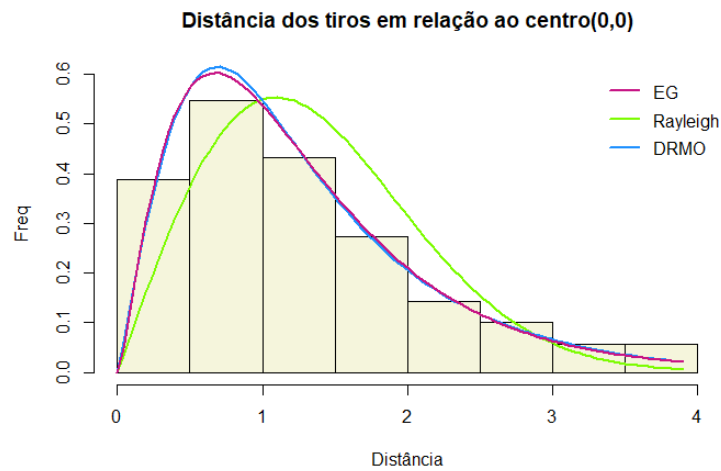


Figura 6 – Histograma dos dados e curva das distribuições ajustadas.

Para ajustar as curvas no histograma, foram estimados os parâmetros de cada distribuição por meio do método de máxima verossimilhança. Os parâmetros foram ajustados utilizando um algoritmo de otimização numérica no software R.

Na seleção dos modelos ajustados, devemos utilizar critérios que levam em consideração a complexidade de cada um deles. Esses critérios, penalizam a função verossimilhança pelo número de parâmetros do modelo e o tamanho da amostra.

O critério de informação de Akaike (AIC) é um dos métodos de seleção de modelos mais utilizados, introduzido por Hirotugu Akaike (1974), a estimativa é baseada na função de log verossimilhança em seu ponto máximo que penaliza conforme o número de parâmetros utilizados. Liang e Zou (2008) definiram uma versão melhorada (CAIC), que adiciona uma ponderação dependente do tamanho amostral.

O critério de informação bayesiano (BIC), é um critério de avaliação de modelos que é definido em termos de probabilidade a posteriori, ele aplica uma penalização mais forte ao número de parâmetros. Utilizando esses critérios, o melhor modelo será aquele que apresentar menor valor para AIC, CAIC E BIC.

$$AIC = -2l(\theta) + 2k$$

$$CAIC = AIC + \frac{2(k+2)(k+3)}{n-k-3}$$

$$BIC = -2l(\theta) + \log(n)$$

As distribuições utilizadas e seus respectivos parâmetros estimados foram a Distribuição Rayleigh com $\theta = 1.096575$, a Distribuição Exponencial-Gama (EG) com $\theta = 2.354054$ e $\alpha = 1.274458$ e a Distribuição Rayleigh Estendida de Marshall Olkin (DRMO) com $\theta = 1.6077657$ e $\alpha = 0.2451225$.

Tabela 5 – Critérios de seleção do modelo

Modelo	AIC	CAIC	BIC
DRMO	355.117	355.416	360.986
EG	355.848	356.147	361.717
Rayleigh	375.611	375.789	378.546

Observando a Tabela 5, notamos que a distribuição DRMO teve um bom ajuste no conjunto de dados estudado. Ela obteve resultado melhores que a distribuição base e ligeiramente melhores que a distribuição EG, que também possui 2 parâmetros.

7.2 VELOCIDADE DO VENTO

A base de dados utilizada é proveniente de um estudo estatístico da velocidade do vento para a geração de energia. Os dados foram coletados na cidade de Presidente Prudente - SP, sendo as observações referentes as médias registradas a cada dez minutos consecutivamente, ao longo dos dias do mês de Agosto no ano de 2002.

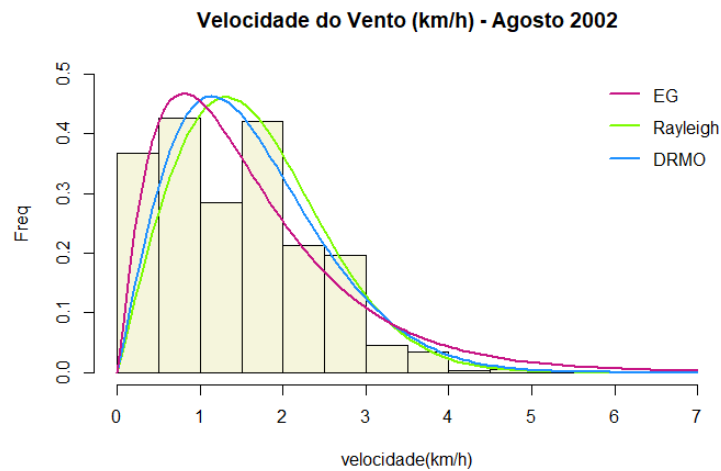


Figura 7 – Histograma dos dados e curva das distribuições ajustadas.

O ajuste das curvas ao histograma vem da estimação dos parâmetros pelo método da máxima verossimilhança, utilizando algoritmos de otimização numérica no software

R. Para seleção dos modelos ajustados, utilizamos os mesmos critérios da aplicação anterior. As distribuições utilizadas e seus respectivos parâmetros estimados foram a Distribuição Rayleigh com $\theta = 1.31547$, a Distribuição Exponencial Gama (EG) com $\theta = 2.1645288$ e $\alpha = 0.9632485$ e a Distribuição Rayleigh Estendida de Marshall Olkin (DRMO) com $\theta = 1.4471689$ e $\alpha = 0.6825817$.

Tabela 6 – Critérios de seleção do modelo

Modelo	AIC	CAIC	BIC
DRMO	21738.65	21738.66	21751.46
EG	21984.87	21984.88	21997.68
Rayleigh	21785.70	21785.71	21792.11

Observando a Tabela 6, notamos que a distribuição DRMO apresenta valores menores do que as distribuições concorrentes, corroborando assim com o ajuste da sua curva na Figura 7.

8 CONCLUSÃO

Com utilização do método de Marshall-Olkin, adicionamos um novo parâmetro α a distribuição Rayleigh, desenvolvendo assim a Distribuição Rayleigh Estendida de Marshall-Olkin (DRMO). Com a introdução de suas funções de densidade e risco, bem como seus respectivos gráficos, foi possível observar uma maior flexibilidade em comparação a distribuição base.

Para estimar os parâmetros da DRMO, foram utilizados os métodos de Máxima Verossimilhança e Bayesiano. O estimador Bayesiano se mostrou mais eficiente conforme o tamanho de amostra aumenta, enquanto o MV apresentou resultados melhores em amostras menores na estimação de θ .

Na aplicação prática da DRMO, foram utilizados duas bases de dados reais. Por meio de critérios de avaliação de modelos, demonstrou-se que a nova distribuição superou tanto a distribuição base quanto uma distribuição exponencial de dois parâmetros.

Portanto, a DRMO se torna uma boa alternativa em estudos de análise de sobrevivência. Com apenas dois parâmetros, juntamente com as características distintas, a nova distribuição torna-se uma opção promissora em comparação com modelos mais complexos.

8.1 CONSIDERAÇÕES FINAIS

O estudo da Distribuição Rayleigh Estendida de Marshall-Olkin mostrou os comportamentos únicos e a capacidade de modelagem da nova distribuição. Durante o desenvolvimento do trabalho, foram adicionadas novas propriedades que não estavam previstas inicialmente, sendo elas o cálculo dos momentos e a matriz de Informação de Fisher.

Além disso, foram realizadas aplicações em dois conjuntos de dados reais, contribuindo na justificativa de sua aplicabilidade e versatilidade. Sendo assim, a criação da Distribuição Rayleigh Estendida de Marshall-Olkin representa uma contribuição significativa para o campo da Estatística.

Para as novas pesquisas da DRMO, recomenda-se que explorem outras aplicações com outros tipos de dados e a comparação com distribuições mais complexas.

9 REFERÊNCIAS

- ABDUL GHANI, Ahmad Shahrizan; MAT ISA, Nor Ashidi. Underwater image quality enhancement through composition of dual-intensity images and Rayleigh stretching. *SpringerPlus*, v. 3, n. 1, p. 1-14, 2014.
- AGUILAR, Guilherme Aparecido Santos. *Família distribuição gama exponenciada*, 2017.
- AGUILAR, Guilherme AS; MOALA, Fernando A.; CORDEIRO, Gauss M. Zerotruncated poisson exponentiated gamma distribution: Application and estimation methods. *Journal of Statistical Theory and Practice*, v. 13, p. 1-20, 2019.
- BECKMANN, Petr. Rayleigh distribution and its generalizations. *Radio Science Journal of Research NBS/USNC-URSI*, v. 68, n. 9, p. 927-932, 1964.
- CRAVEIRO, Vitor Marchetti. *Cálculo de CEP e outras propriedades estatísticas do tiro ao alvo com armas de fogo*, 2024.
- KUNDU, Debasis; RAQAB, Mohammad Z. Generalized Rayleigh distribution: different methods of estimations. *Computational Statistics & Data Analysis*, v. 49, n. 1, p. 187-200, 2005.
- KURUOGLU, Ercan E.; ZERUBIA, Josiane. Modeling SAR images with a generalization of the Rayleigh distribution. *IEEE Transactions on Image Processing*, v. 13, n. 4, p. 527-533, 2004.
- LARSON, Harold J. *Introduction to probability theory and statistical inference*. (No Title), 1969.
- MARIA, Anu. Introduction to modeling and simulation. In: **Proceedings of the 29th conference on Winter simulation**. 1997. p. 7-13.
- MARSHALL, Albert W.; OLKIN, Ingram. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, v. 84, n. 3, p. 641-652, 1997.
- MEINTANIS, Simos G. Test of fit for Marshall–Olkin distributions with applications. *Journal of Statistical Planning and inference*, v. 137, n. 12, p. 3954- 3963, 2007.
- RINALDI, José Gilberto Spasiani. *Caracterização Estatística da Velocidade de Ventos para Geração de Energia*, 2001..
- SARTI, Alessandro et al. Maximum likelihood segmentation of ultrasound images with Rayleigh distribution. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, v. 52, n. 6, p. 947-960, 2005.