INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Licenciatura em Matemática Aplicada à Tecnologia e à Empresa

Análise da Velocidade de Vento em Turbinas Eólicas através de uma Mistura de Distribuições Weibull

Xu Bichen

Trabalho Final de Licenciatura

Orientadores:

Alda Carvalho

Daniel Vaz

Julho de 2021

INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Licenciatura em Matemática Aplicada à Tecnologia e à Empresa

Análise da Velocidade de Vento em Turbinas Eólicas através de uma Mistura de Distribuições Weibull

Xu Bichen

Trabalho final de licenciatura

Orientadores:

Alda Carvalho

Daniel Vaz

Julho de 2021

Análise da Velocidade de Vento em Turbinas Eólicas através de uma Mistura de Distribuições Weibull

Análise da Velocidade de Vento em Turbinas Eólicas através de uma Mistura de Distribuições Weibull

Trabalho Final de Licenciatura

Xu Bichen

INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Licenciatura em Matemática Aplicada à Tecnologia e à Empresa

Índice

I. Agradecimentos x
II. Resumo xi
III. Abstract xii
IV. Introdução xiii
Capítulo 1 15
1.Fundamentação Teórica 15
1.1 Energia Eólica 15
1.2 Distribuição de Weibull 17
1.2.1 Distribuição Weibull Biparamétrica 17
1.2.2 Distribuição Weibull Triparamétrica 18
1.2.3 Distribuição Weibull para a Velocidade do Vento 18
1.3 DFVV Bimodais 20
Capítulo 2 21
2. Metodologia 21
2.1 Modelo de Mistura de Distribuições 21
2.1.1 Definição de Modelo de Mistura de Distribuições 21
2.1.2 Mistura de Distribuições Weibull 22
2.2 Método dos Momentos 22
2.2.1 Definição de Método dos Momentos 22
2.2.2 Estimação dos Parâmetros da Weibull Biparamétrica 23
2.2.3 Estimação dos Parâmetros da Mistura de Duas Weibulls Triparamétricas 25
2.3 Algoritmo EM 28
2.3.1 Estimador de Máxima Verosimilhança 28
2.3.2 Especificação do Algoritmo EM 31
2.3.3 Função fitmixture do R 32
2.4 Análise Estatística 34
2.4.1 Erro Absoluto 34
2.4.2 Análise do Melhor Ajustamento 34
Capítulo 3 36
3. Estudo de Simulação 36
3.1 Simulação da Mistura de Distribuições Weibull 36
3.2 Aplicando o MM nos 4 casos 40

3.2.1 Códigos das Funções que usar no MM 40
3.2.2 Resultados dos parâmetros pelo MM 42
3.3 Aplicando o Algoritmo EM nos 4 Casos 49
3.4 Comparação dos Métodos de Estimação 54
Capítulo 4 55
4. Caso de Estudo com Dados Reais 55
4.1 Velocidade de Vento 55
4.2 Aplicando o MM nos dados 56
4.2 Aplicando o EM nos dados 57
Capítulo 5 59
5. Conclusão 59
Bibliografia 60

I. Agradecimentos

Aos meus orientadores, a professora Alda Carvalho e o professor Daniel Vaz por terem tornado possível a realização deste estudo e pelo apoio.

Aos meus pais que me proporcionaram educação e ensinamentos que irão ficar na minha memória, apoiando-me sempre na concretização dos meus objetivos.

Aos meus irmãos que me ajudaram muito com o português e também pelo apoio e incentivos constantes.

Agradeço também a todos os professores do curso de Matemática Aplicada à Tecnologia e à Empresa pela educação.

II. Resumo

O estudo de modelos probabilísticos para a distribuição de frequência da velocidade do vento é essencial para a análise de recursos de energia eólica. A distribuição de probabilidade da velocidade do vento é geralmente expressa pela função de distribuição Weibull de 2 parâmetros. No entanto, uma limitação importante da função de distribuição de Weibull é que não cobre todos os casos observados. Por vezes, observa-se uma distribuição da velocidade de vento com distribuições bimodais que resultam, em muitos casos, de condições climáticas especiais. Uma das conclusões do projeto MoDMaPE (IDICA-IPL, 2018) indica que, em terreno complexo, a distribuição de frequência de velocidade do vento nem sempre segue uma distribuição unimodal.

Deste modo, para estes casos, a função de mistura de distribuições Weibulls unimodais pode ser considerada.

Neste trabalho foram simulados dados através de uma mistura de duas componentes com distribuição Weibull para aplicar dois métodos distintos para estimar os parâmetros da distribuição de Weibull mistura: método dos momentos e algoritmo EM. Este estudo de simulação permitiu tipificar diferentes possibilidades de ocorrência de distribuições bimodais. Por fim, na busca de resultados com o melhor ajustamento, a metodologia foi aplicada a dados reais da velocidade de vento registada em turbinas eólicas. Para a avaliação da qualidade do ajustamento, foram utilizados a raiz do erro quadrático médio (RMSE) e o coeficiente de determinação (R^2). Comparando os resultados, verificou-se maior precisão no algoritmo EM do que método dos momentos.

Palavras-chave: Energia Eólica, Distribuição de Weibull Mistura, Determinação dos Parâmetros da Distribuição de Weibull Mistura, Método dos Momentos, Algoritmo EM.

III. Abstract

The study of probabilistic models for the frequency distribution of wind speed is essential for the analysis of wind energy resources. The wind speed probability distribution is usually expressed by the 2-parameter Weibull distribution function. However, an important limitation of the Weibull distribution function is that it does not cover all observed cases. Sometimes, there is a distribution of wind speed with bimodal distributions that result, in many cases, from special climatic conditions. One of the conclusions of the MoDMaPE project (IDICA-IPL, 2018) indicates that, in complex terrain, the wind speed frequency distribution does not always follow a unimodal distribution.

In this context, for these cases, the mixture function of Weibulls unimodals distributions can be considered.

In this work, data were simulated using a two-component mixture with a Weibull distribution to apply two different methods to estimate the parameters of the mixed Weibull distribution: method of moments and EM algorithm. This simulation study allowed us to typify different possibilities of occurrence of bimodal distributions. Finally, in search of the best-fit results, the methodology was applied to real data of wind speed recorded in wind turbines. For a quality of fit evaluation, the root-mean-square error (RMSE) and the coefficient of determination (R^2) were applied, this is, the parameters of the estimate were substituted in the probability density function, then the RMSE and R^2 were calculated. After comparing the results, it was found greater precision in the EM algorithm than the method of moments.

Keywords: Wind Energy, Mixed Weibull Distribution, Determination of the Mixed Weibull Distribution Parameters, Method of Moments, EM Algorithm.

IV. Introdução

O investimento em fontes de energia renováveis, incluindo turbinas eólicas, é de particular importância devido ao aumento da eficiência da energia limpa e à necessidade de reduzir a poluição e o consumo de combustíveis fósseis [1]. À medida que as tecnologias de geração eólica melhoram, esta forma de produção de energia torna-se uma alternativa valiosa relativamente às fontes convencionais de energia [2].

Um problema importante no uso da energia eólica é a sua natureza incerta e característica de imprevisibilidade [3]. Para desenvolver e utilizar recursos de energia eólica de forma eficiente, as características dos recursos de energia eólica precisam primeiro ser analisadas e estudadas [4]. A avaliação dos recursos energéticos em parques eólicos é a base para o desenvolvimento. Compreender as características da distribuição da frequência da velocidade do vento em parques eólicos é a chave para a análise de recursos de energia eólica. A distribuição da frequência da velocidade do vento (DFVV) pode ser expressa através da função de densidade de probabilidade da velocidade do vento. Esta função descreve as principais características da velocidade do vento [5].

As diferentes caraterizações da distribuição da frequência da velocidade do vento para parques eólicos refletem diretamente as diferentes condições dos recursos de energia eólica em um local. A sua racionalidade e a precisão terão uma influência direta nas decisões finais de seleção do modelo de turbinas eólicas, estimativa de geração de energia e avaliação de benefícios económicos de parques eólicos. Observam-se diferenças significativas entre a geração de energia real e projetada em muitos parques eólicos, nomeadamente no que diz respeito à operação prática do parque eólico e sua análise de avaliação após a operação. Uma das principais razões para esta discrepância é existência de um erro significativo na descrição dos recursos de energia eólica para parques eólicos [6]. Portanto, é de grande importância estudar modelos de distribuição de frequência da velocidade do vento para um melhor aproveitamento de parques eólicos.

Os avanços computacionais das últimas décadas aliados à constante melhoria na tecnologia de sensores tornaram a aquisição de dados cada vez mais eficaz. Num parque eólico existem diferentes sensores que medem caraterísticas importantes para a monotorização das turbinas eólicas. Estes sensores fazem parte de um sistema integrado chamado SCADA (*Supervisory Control And Data Acquisition*). SCADA significa supervisão e aquisição de dados, sendo possível controlar remotamente equipamentos em diferentes locais através da recolha e armazenamento dos dados operacionais de interesse.

Uma característica bastante relevante neste estudo é a intensidade do vento. Usando dados SCADA e diferentes técnicas da estatística, é possível estimar parâmetros importantes na caracterização do vento registado em cada turbina.

O principal objetivo deste trabalho foi a estimação dos parâmetros de uma mistura de distribuições Weibull. Inicialmente, usando dados simulados, foi feita uma análise e tipificação de diferentes possibilidades de ocorrência de distribuições bimodais. Para isso utilizou-se uma mistura de Weibull biparamétrica e de Weibull triparamétrica tendo-se optado pela mistura de Weibull biparamétrica quer no estudo prévio com os dados simulados como com os dados reais. No processo de estimação foram comparados dois métodos: método dos momentos e algoritmo EM.

Este relatório está organizado em cinco capítulos.

No capítulo 1 são introduzidos conceitos importantes relativamente à energia eólica e à distribuição Weibull.

No capítulo 2 apresenta-se a metodologia utilizada neste trabalho, incluindo o método dos Momentos e o algoritmo EM.

No capítulo 3 apresenta-se um estudo de simulação de uma mistura da distribuição de Weibull com vista a estudar tipificas as diferentes possibilidades de ocorrência de distribuições bimodais.

No capítulo 4 apresenta-se um caso de estudo com dados reais recolhidos num parque eólico.

O capítulo 5 apresentam-se as principais conclusões do trabalho.

E por fim, apresentam-se as referências bibliográficas utilizadas neste trabalho.

Capítulo 1

1.Fundamentação Teórica

1.1 Energia Eólica

O vento tem a sua origem na associação entre a irradiação solar e a rotação planetária. Todos os planetas envoltos por gases do nosso sistema solar demonstram a existência de distintas formas de circulação atmosférica e apresentam ventos em suas superfícies. Trata-se de um mecanismo solar-planetário permanente; a sua duração é mensurável numa escala de bilhões de anos, por esta razão, o vento é considerado uma fonte renovável de energia [8].

A energia eólica provém da radiação solar, uma vez que os ventos são gerados pelo aquecimento não uniforme da superfície terrestre. Uma estimativa da energia total do vento disponível torno do planeta pode ser feita a partir da hipótese de que, aproximadamente 2% da energia solar absorvida pela Terra é convertida em energia cinética do vento. Esta percentagem, embora pareça pequena, representa cem de vezes a potência anual instalada nas centrais elétricas do mundo [8].

O aproveitamento da energia eólica é a transformação da energia do vento em energia útil, tal como, a utilização de aerogeradores para produzir eletricidade, de moinhos de vento para produzir energia mecânica ou velas para impulsionar os veleiros. A energia eólica, enquanto alternativa aos combustíveis fósseis, é renovável, e na sua utilização não é poluente e não emite gases do efeito estufa.

Enquanto recurso energético, a caraterística mais importante da energia eólica é a sua variabilidade, tanto geograficamente como temporal [9].

Os ventos que sopram em escala global e aqueles que se manifestam em pequena escala são influenciados por diferentes aspetos, em relação à superfície, natureza da superfície (água ou solo). Entre os diferentes aspetos destacam-se a altura, a rugosidade, os obstáculos e o relevo da superfície [8]. Ou seja, os acidentes geográficos podem alterar localmente a direção e a intensidade do vento.

Numa dada localização, a variação temporal de larga escala significa que a disponibilidade de vento pode variar de um ano para outro, com variações de ainda mais larga escala ocorrendo ao longo de períodos de décadas ou mais. Estas variações de longo termo não são bem conhecidas, o que pode dificultar a previsão da viabilidade económica de um parque eólico [9].

Em escalas de tempo inferiores a um ano, as variações sazonais são mais previsíveis. Há ainda variações importantes em escalas de tempo mais curtas, associadas à passagem de sistemas meteorológicos. Embora as causas destas variações sinópticas [1] sejam relativamente bem compreendidas, elas não se conseguem prever com mais do que uns dias, ou mesmo horas de avanço. Dependendo da localização, as variações diurnas também podem ser importantes, e são razoavelmente previsíveis (àquela escala de tempo) [9].

A escalas de tempo ainda mais reduzidas, da ordem de minutos, segundos, ou ainda menos, as variações da velocidade do vento estão associadas à turbulência. Estas variações têm efeito no projeto e desempenho das turbinas eólicas, bem como na qualidade da energia elétrica injetada na rede [9].

Em 1957, van der Hoven elaborou um espetro da velocidade do vento desde a escala de uma semana até à escala de um segundo (Figura 1.1). O gráfico mostra três picos correspondendo aos efeitos atrás referidos: sinótico, diurno e turbulento. Existirá ainda um quarto pico (fora do gráfico) correspondente ao ano. Note-se a energia muito reduzida no intervalo do espetro entre as 2 horas e os 10 minutos, que se pode denominar de vazio espetral. Este resultado significa que as variações sinóticas e diurnas podem ser tratadas de um modo independente das variações de mais alta-frequência da turbulência. Um período de tempo daquele intervalo, onde as oscilações são menores, será conveniente para definir valores médios da velocidade do vento. Por outras palavras, o valor médio da velocidade do vento depende pouco do período adotado se este estiver compreendido entre 10 minutos e 2 horas [9].

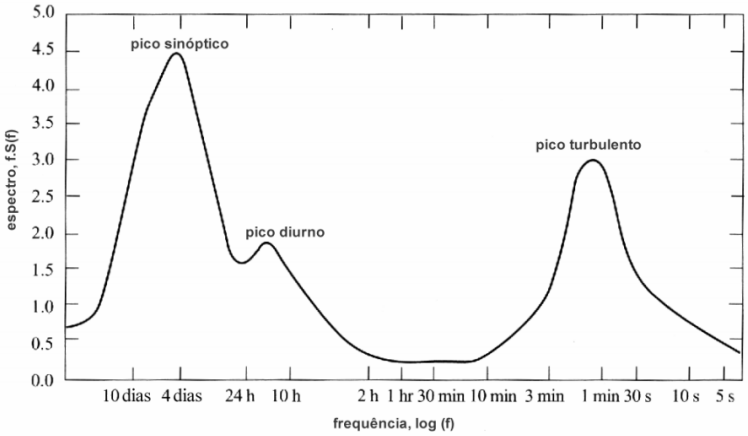

Figura 1. 1 - Espeto de vento para o parque eólico de Brookhaven, Nova Iorque (citado em [9])

[1] Em meteorologia, este termo está conotado com simultaneidade. Refere-se, pois a dados obtidos simultaneamente sobre uma área extensa com o propósito de obter uma imagem (completa) instantânea do estado da atmosfera (citado em [9]).

1.2 Distribuição de Weibull

A distribuição de probabilidade Weibull foi proposta originalmente por Fisher e Tippett em 1928, no estudo de valores extremos. Posteriormente, foi também desenvolvida de modo independente pelo físico sueco Waloddi Weibull, nos seus estudos sobre a falha de materiais [10]. As pesquisas no período pós-segunda guerra deram ênfase às análises de resistência dos materiais, o que resultou na associação do nome de W. Weibull a esta distribuição [11].

Por exemplo, Justus, Patel e Manwell indicam a distribuição de Weibull como a que melhor se adequa ao comportamento dos ventos, sendo a velocidade caracterizada por duas funções – uma função densidade de probabilidade (f.d.p.) e uma função de distribuição (f.d.) [12-14]. A f.d.p. descreve a verosimilhança de uma variável aleatória tomar um valor dado, e a f.d. descreve completamente a distribuição da probabilidade de uma variável aleatória do valor real.

1.2.1 Distribuição Weibull Biparamétrica

A distribuição de Weibull biparamétrica [15] é descrita pela f.d.p.:

$$f(x) = \frac{k}{c}\left(\frac{x}{c}\right)^{k-1} e^{-\left(\frac{x}{c}\right)^{k}}, x \geq 0 \tag{1.1}$$

E a f.d. é dada por:

$$F(x) = 1 - e^{-\left(\frac{x}{c}\right)^{k}}, x \geq 0 \tag{1.2}$$

sendo x é o valor da variável aleatória, c é o parâmetro de escala ($c > 0$) e k é o parâmetro de forma ($k > 0$).

1.2.2 Distribuição Weibull Triparamétrica

Em comparação com a função Weibull biparamétrica, na função Weibull triparamétrica [15] há mais um parâmetro, designado de parâmetro de localização que descreve o deslocamento da distribuição biparamétrica.

A f.d.p. da distribuição Weibull triparamétrica é definida como

$$f(x) = \frac{k}{c}\left(\frac{x-u}{c}\right)^{k-1} e^{-\left(\frac{x-u}{c}\right)^{k}}, x \geq u \tag{1.3}$$

e a f.d. da distribuição Weibull triparamétrica é descrita como

$$F(x) = 1 - e^{-\left(\frac{x-u}{c}\right)^{k}}, x \geq u \tag{1.4}$$

onde x é o valor da variável aleatória, c é o parâmetro de escala ($c > 0$), u é o parâmetro de localização ($u \geq 0$) e k é o parâmetro de forma ($k > 0$).

1.2.3 Distribuição Weibull para a Velocidade do Vento

1.2.3.1 Distribuição Weibull Biparamétrica

Segundo Justus, a distribuição de Weibull para uma determinada velocidade de vento v, pode ser expressa pela f.d.p. biparamétrica (1.5) [12].

$$f(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} e^{-\left(\frac{v}{c}\right)^{k}} \tag{1.5}$$

E a f.d. é dada por:

$$F(v) = 1 - e^{-\left(\frac{v}{c}\right)^{k}} \tag{1.6}$$

sendo v a variável velocidade do vento em $m/s (v \geq 0)$, c é o parâmetro de escala ($c > 0$ e expresso em m/s) e k é o parâmetro de forma ($k > 1$ e adimensional).

O exemplo de uma função de densidade de probabilidade Weibull com os valores de k diferentes, para um dado valor c ($c = 2$) pode ver-se na Figura 1.2. Maiores valores de k indicam maior constância dos ventos, com menor ocorrência de valores extremos [16].

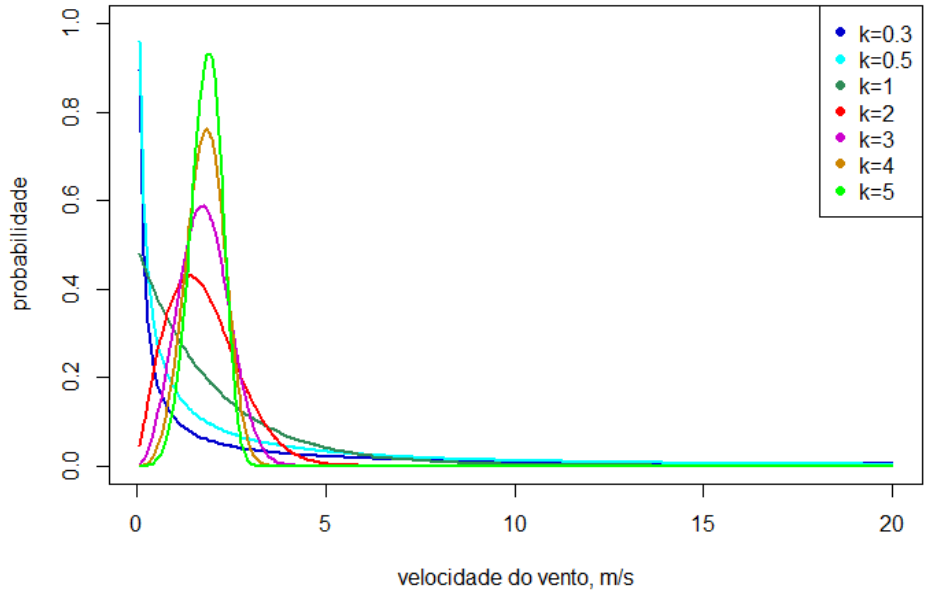

Figura 1. 2 - f.d.p Weibull para diferentes valores de k

Para $0 < k \leq 1$, a f.d.p. tende para ∞ quando v se aproxima de zero à esquerda e é estritamente decrescente. Portanto, $0 < k \leq 1$ não é possível para a velocidade do vento, visto que não é fisicamente realista que o vento nulo tenha a probabilidade máxima, nota-se que o vento nulo é a velocidade do vento igual a zero.

1.2.3.2 Distribuição Weibull Triparamétrica

A f.d.p. da distribuição Weibull triparamétrica para a velocidade do vento é dado por

$$f(v) = \frac{k}{c}\left(\frac{v-u}{c}\right)^{k-1} e^{-\left(\frac{v-u}{c}\right)^{k}} \tag{1.7}$$

e a f.d. da distribuição Weibull triparamétrica para a velocidade do vento é descrita como

$$F(v) = 1 - e^{-\left(\frac{v-u}{c}\right)^k} \tag{1.8}$$

sendo v a variável velocidade do vento em $m/s (v \geq u)$, c é o parâmetro de escala ($c > 0$ e expresso em m/s), u é o parâmetro de localização ($u \geq 0$) e k é o parâmetro de forma ($k > 1$ e adimensional).

1.3 DFVV Bimodais

Em alguns casos a distribuição de frequências da velocidade do vento não segue uma distribuição de Weibull. Por exemplo, existem ventos com distribuições bimodais ou mesmo multimodais que resultam, em muitos casos, de condições climáticas especiais. Em terreno complexo, a distribuição de frequência de velocidade do vento nem sempre segue uma distribuição de unimodal (MoDMaPE (IDICA-IPL, 2018)).

Na Figura 1.2 mostra-se um exemplo de uma DFVV que, não seguindo uma distribuição de Weibull, pode ser decomposta em duas distribuições de Weibull.

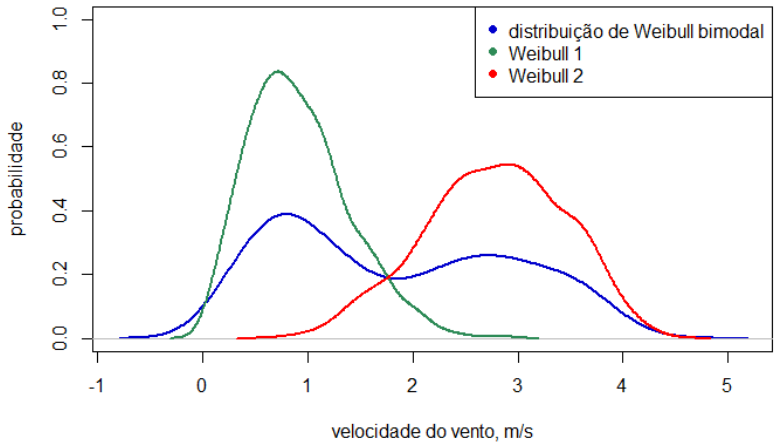

Figura 1. 3 - Exemplo de uma DFVV que não seguindo uma distribuição de Weibull, pode ser decomposta em duas distribuições de Weibull.

Deste modo, uma mistura de distribuições Weibulls unimodais pode ser considerada como uma possibilidade, quando existe uma mistura de duas componentes distintas.

Capítulo 2

2. Metodologia

Neste capítulo é apresentada a metodologia utilizada para a realização deste trabalho. Primeiramente, apresenta-se uma definição de um modelo de mistura de distribuições de forma genérica e depois o caso específico de interesse para este trabalho, a função de Weibull de mistura. Seguidamente, apresentam-se os métodos de estimação dos parâmetros: o método dos momentos e algoritmo EM. Por fim, apresenta-se os conceitos básicos do erro absoluto, do RMSE e do coeficiente de determinação. Todos os resultados foram obtidos através da utilização de ferramentas computacionais do RStudio.

2.1 Modelo de Mistura de Distribuições

Os modelos de mistura de distribuições têm, desde há muito, merecido especial atenção dos estatísticos e da comunidade científica em geral, tendo-se assistido na última década a um interesse crescente no seu estudo, quer do ponto de vista teórico quer prático.

A importância destes modelos, deve-se ao facto de os mesmos serem os mais adequados quando a população em estudo é formada por várias subpopulações que estão presentes na população inicial em proporções desconhecidas. Situações destas ocorrem com frequência na prática quando os dados provêm de populações formadas por grupos distintos, cuja existência pode ou não ser conhecida á priori, desconhecendo se quais os dados que pertencem a cada grupo [18].

2.1.1 Definição de Modelo de Mistura de Distribuições

Seja X a variável aleatória com valores num espaço S e cuja f.d.p. é dada por:

$$f_X(x) = \sum_{i=1}^{m} \lambda_i \, f_i(x) \tag{2.1}$$

onde $f_1(x), \ldots, f_m(x)$ são as funções densidades de probabilidade componentes da mistura, λ_i são designadas por proporções ou pesos de mistura, $0 \leq \lambda_i \leq 1, i = 1, \ldots m$ e m é o número de componentes na mistura.

2.1.2 Mistura de Distribuições Weibull

Com base na função Weibull triparamétrica, a f.d.p. de Weibull mistura pode descrita como

$$f_{mix}(v) = \sum_{i=1}^{m} \lambda_i \, f_i(v|k_i, c_i, u_i) \tag{2.2}$$

A f.d. de Weibull mistura pode descrita como

$$F_{mix}(v) = \sum_{i=1}^{m} \lambda_i \, F_i(v|k_i, c_i, u_i) \tag{2.3}$$

onde λ_i são os pesos de mistura, $0 \leq \lambda_i \leq 1$, $\sum_{i=1}^{m} \lambda_i = 1$ e m é o número de componentes na mistura.

O cálculo dos parâmetros da função de distribuição Weibull mista é difícil. A seguir, serão apresentados dois métodos de estimação dos parâmetros.

2.2 Método dos Momentos

O método dos momentos (MM) [20] é provavelmente o mais antigo método de estimação, introduzido por Pearson no final do século XIX. Por ser bastante simples de utilizar, é ainda hoje uma opção a ter em conta, pois permite encontrar estimativas dos parâmetros em situações onde outros métodos de estimação mais complexos não conseguem.

2.2.1 Definição de Método dos Momentos

Seja $X_1, \ldots, X_n$ é uma variável aleatória da função massa de probabilidade (f.m.p.) $p(x)$ ou f.d.p $f(x)$. Para $s = 1,2,3, \ldots$, o s-ésimo momento populacional, ou s-ésimo momento da distribuição, é $E(X^s)$. O s-ésimo momento de amostra é

$$E(X^s) = \sum_i^n \frac{X_i^s}{n}$$

(2.4)

Assim, o primeiro momento populacional é dado pela própria média amostral, isto é,

$$E(X) = \sum_i^n \frac{X_i}{n} = \mu = \bar{x}$$

(2.5)

Se X é uma variável aleatória discreta com f.m.p. $p(x)$, temos que

$$E(X^s) = \sum x^s p(x)$$

(2.6)

e, se X é uma variável aleatória contínua com f.d.p. $f(x)$, temos que

$$E(X^s) = \int_\infty^\infty x^s f(x) dx$$

(2.7)

2.2.2 Estimação dos Parâmetros da Weibull Biparamétrica

Nesta secção foram aplicados o MM para estimar os parâmetros da distribuição de Weibull biparamétrica.

Substituindo a f.d.p. de Weibull (1.5) na função (2.7), obteve-se

$$E(V^s) = \int_0^\infty v^s f(v) dv = \int_0^\infty v^s \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} e^{-\left(\frac{v}{c}\right)^k} dv = \int_0^\infty v^s \frac{k}{c^k} v^{k-1} e^{-\left(\frac{v}{c}\right)^k} dv$$

Assim fazendo uma substituição $t = \left(\frac{v}{c}\right)^k$, teve-se que

$$dt = \frac{k}{c^k} v^{k-1} dv \quad \Leftrightarrow dv = \frac{c^k}{k v^{k-1}} dt$$

$$\Rightarrow \frac{c^k}{k v^{k-1}} v^s \frac{k}{c^k} v^{k-1} e^{-\left(\frac{v}{c}\right)^k} = v^s e^{-\left(\frac{v}{c}\right)^k}$$

$$t = \left(\frac{v}{c}\right)^k \Leftrightarrow v^k = t c^k \Leftrightarrow v = (t c^k)^{1/k} = t^{1/k} c \Leftrightarrow v^s = t^{s/k} c^s$$

e, desta forma

$$E(V^s) = \int_0^\infty v^s \frac{k}{c^k} v^{k-1} e^{-\left(\frac{v}{c}\right)^k} dv = \int_0^\infty t^{s/k}\, c^s e^{-t} dt$$

Como a função gama é dado por

$$\Gamma(x) = \int_0^\infty t^{x-1}\, e^{-t} \tag{2.8}$$

Logo,

$$E(V^s) = \int_0^\infty t^{s/k}\, c^s e^{-t} dt = c^s \Gamma\left(1 + \frac{s}{k}\right)$$

Portanto, o valor médio é dado por

$$\overline{v} = E(V) = c\Gamma\left(1 + \frac{1}{k}\right) \tag{2.9}$$

e a variância é dada por

$$\sigma^2 = Var(V) \quad = E(V^2) - E^2(V) = c^2\Gamma\left(1 + \frac{2}{k}\right) - \left[c\Gamma\left(1 + \frac{1}{k}\right)\right]^2$$

$$= c^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right] \tag{2.10}$$

Desse modo, o parâmetro de forma k do MM pode ser determinado pela equação seguinte:

$$\frac{\overline{v}^2}{\sigma^2 + \overline{v}^2} \quad = \frac{c^2\Gamma^2\left(1 + \frac{1}{k}\right)}{c^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right] + c^2\Gamma^2\left(1 + \frac{1}{k}\right)}$$

$$= \frac{\Gamma^2\left(1 + \frac{1}{k}\right)}{\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) + \Gamma^2\left(1 + \frac{1}{k}\right)}$$

$$\Leftrightarrow \frac{\overline{v}^2}{\sigma^2 + \overline{v}^2} = \frac{\Gamma^2\left(1 + \frac{1}{k}\right)}{\Gamma\left(1 + \frac{2}{k}\right)}$$

(2.11)

e o parâmetro de escala c do MM pode ser descrito como

$$c = \frac{\overline{v}}{\Gamma\left(1+\frac{1}{k}\right)}$$

(2.12)

2.2.3 Estimação dos Parâmetros da Mistura de Duas Weibulls Triparamétricas

Nesta secção foram aplicados o MM para estimar os parâmetros para uma mistura de Weibull com duas componentes Weibulls triparamétricas ($m = 2$).

Considerou-se o ponto mínimo do gráfico como v_{divide}, e dividiu-se o conjunto da amostra da velocidade do vento V em dois subconjuntos, V_1 e V_2. Entre eles, V_1 é um subconjunto de observações cuja velocidade do vento é menor ou igual que v_{divide}, contendo n_1 observações, V_2 é um subconjunto de observações cuja velocidade do vento é maior que v_{divide}, contendo n_2 observações. Os parâmetros da mistura de Weibull podem ser estimados a partir dos dois subconjuntos amostrais V_1 e V_2 [21].

Considerou-se a velocidade mínima do vento de cada subconjunto de observações como o parâmetro de localização do MM $u_i (i = 1{,}2)$ [21]:

$$\begin{cases} u_1 = min\ \ V_1 \\ u_2 = min\ \ V_2 \end{cases}$$

(2.13)

Substituindo a f.d.p. da distribuição Weibull triparamétrica (1.7) na função (2.7), obtém-se

$$E(V^s) = \int_u^{\infty} v^s \frac{k}{c}\left(\frac{v-u}{c}\right)^{k-1} e^{-\left(\frac{v-u}{c}\right)^k} dv = \int_u^{\infty} v^s \frac{k}{c^k}(v-u)^{k-1} e^{-\left(\frac{v-u}{c}\right)^k} dv$$

Assim, fazendo uma substituição $t = \left(\frac{v-u}{c}\right)^k$, tem-se que

$$dt = \frac{k}{c^k}(v-u)^{k-1}dv \quad \Leftrightarrow dv = \frac{c^k}{k(v-u)^{k-1}}dt$$
$$\Rightarrow \frac{c^k}{k(v-u)^{k-1}} v^s \frac{k}{c^k}(v-u)^{k-1} e^{-\left(\frac{v-u}{c}\right)^k} = v^s e^{-\left(\frac{v-u}{c}\right)^k}$$

$$t = \frac{(v-u)^k}{c^k} \quad \Leftrightarrow (v-u)^k = tc^k \Leftrightarrow v-u = (tc^k)^{1/k} = t^{1/k}c$$
$$\Leftrightarrow v = t^{1/k}c + u \Rightarrow v^s = \left(t^{1/k}c + u\right)^s$$

e, desta forma,

$$E(V) = \int_u^\infty v \frac{k}{c^k}(v-u)^{k-1} e^{-\left(\frac{v-u}{c}\right)^k} dv = \int_0^\infty \left(t^{1/k}c + u\right) e^{-t} dt$$

$$\Leftrightarrow E(V) = c\int_o^\infty t^{1/k} e^{-t} dt + u\int_0^\infty e^{-t}\, dt = c\Gamma\left(1+\frac{1}{k}\right) - u[e^{-t}]_0^\infty = c\Gamma\left(1+\frac{1}{k}\right) + u$$

$$E(V^2) = \int_u^\infty v^2 \frac{k}{c^k}(v-u)^{k-1} e^{-\left(\frac{v-u}{c}\right)^k} dv = \int_0^\infty \left(t^{1/k}c + u\right)^2 e^{-t} dt$$

$$\Leftrightarrow E(V^2) = \int_0^\infty \left(c^2 t^{2/k} + 2cut^{1/k} + u^2\right) e^{-t} dt$$

$$= c^2 \int_0^\infty t^{2/k} e^{-t} dt + 2cu\int_0^\infty t^{1/k} e^{-t} dt + u^2 \int_0^\infty e^{-t}\, dt$$

$$= c^2\Gamma\left(1+\frac{2}{k}\right) + 2cu\Gamma\left(1+\frac{1}{k}\right) - u^2[e^{-t}]_0^\infty$$

$$= c^2\Gamma\left(1+\frac{2}{k}\right) + 2cu\Gamma\left(1+\frac{1}{k}\right) + u^2$$

Logo, a variância é dada por

$$Var(V) = E(V^2) - E^2(V) = c^2\Gamma\left(1+\frac{2}{k}\right) + 2cu\Gamma\left(1+\frac{1}{k}\right) + u^2 - \left[c\Gamma\left(1+\frac{1}{k}\right) + u\right]^2$$

$$= c^2\Gamma\left(1+\frac{2}{k}\right) + 2cu\Gamma\left(1+\frac{1}{k}\right) + u^2 - c^2\Gamma^2\left(1+\frac{1}{k}\right) - 2cu\Gamma\left(1+\frac{1}{k}\right) - u^2$$

$$= c^2\Gamma\left(1+\frac{2}{k}\right) - c^2\Gamma^2\left(1+\frac{1}{k}\right) = c^2\left[\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right]$$

Conclui-se a velocidade média e a variância são dadas por

$$\overline{v} = E(V) = c\Gamma\left(1 + \frac{1}{k}\right) + u \tag{2.14}$$

e

$$\sigma^2 = Var(V) = c^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right] \tag{2.15}$$

Deste modo, o parâmetro de forma do MM $k_i(i = 1{,}2)$ pode ser determinado pela equação seguinte:

$$\frac{(\overline{v}_i - u_i)^2}{\sigma_i^2 + (\overline{v}_i - u_i)^2} = \frac{\left[c_i\Gamma\left(1 + \frac{1}{k_i}\right)\right]^2}{c_i^2\left[\Gamma\left(1 + \frac{2}{k_i}\right) - \Gamma^2\left(1 + \frac{1}{k_i}\right)\right] + \left[c_i\Gamma\left(1 + \frac{1}{k_i}\right)\right]^2}$$

$$= \frac{c_i^2\Gamma^2\left(1 + \frac{1}{k_i}\right)}{c_i^2\left[\Gamma\left(1 + \frac{2}{k_i}\right) - \Gamma^2\left(1 + \frac{1}{k_i}\right) + \Gamma^2\left(1 + \frac{1}{k_i}\right)\right]}$$

$$\Rightarrow \frac{(\overline{v}_i - u_i)^2}{\sigma_i^2 + (\overline{v}_i - u_i)^2} = \frac{\Gamma^2\left(1 + \frac{1}{k_i}\right)}{\Gamma\left(1 + \frac{2}{k_i}\right)} \tag{2.16}$$

onde $\overline{v}_i(i = 1{,}2)$ é a velocidade média das observações $V_1 \ \ e \ \ V_2$, $c_i(i = 1{,}2)$ é o parâmetro de escala do MM, $\sigma_i^2(i = 1{,}2)$ é a variância das observações $V_1 \ \ e \ \ V_2$ e Γ é a função Gama.

$$\overline{v}_i = c_i\Gamma\left(1 + \frac{1}{k_i}\right) + u_i \Leftrightarrow c_i\Gamma\left(1 + \frac{1}{k_i}\right) = \overline{v}_i - u_i$$

Portanto, obtém-se o parâmetro de escala do MM $c_i(i = 1{,}2)$:

$$c_i = \frac{\overline{v}_i - u_i}{\Gamma\left(1 + \frac{1}{k_i}\right)} \tag{2.17}$$

E o peso do MM $\lambda_i (i = 1,2)$ pode ser obtido pela proporção da amostra:

$$\lambda_i = \frac{n_i}{n_1 + n_2} \tag{2.18}$$

2.3 Algoritmo EM

O algoritmo EM [22] do inglês "*Expectation Maximization*" é um algoritmo clássico da estatística usado para determinar estimativas de parâmetros em modelos de mistura, em modelos com dados não observados.

Basicamente, se uma variável foi algumas vezes observada e outras não, pode-se utilizar os casos para os quais foi observada para "aprender" a predizer seus valores quando não foram observados. O algoritmo EM realiza esta tarefa, sempre que seja conhecida a forma geral da distribuição de probabilidade das variáveis [23].

O algoritmo EM utiliza o estimador de máxima verosimilhança dos parâmetros, ou seja, estima parâmetros que sejam os mais consistentes com os dados da amostra no sentido de maximizar a função de verosimilhança.

Antes de introduzir o algoritmo EM, será apresenta o estimador de máxima verosimilhança.

2.3.1 Estimador de Máxima Verosimilhança

A máxima verosimilhança (MV) [20] é um método para estimar os parâmetros de um modelo estatístico. Assim, a partir de um conjunto de dados, e dado uma modelo estatística, o EMV estima valores para os diferentes parâmetros do modelo.

A função de verosimilhança de n variáveis aleatórias $X_1, X_2, \ldots, X_n$ é definida como a densidade conjunta de n variáveis aleatórias, digamos $f_{X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n|\theta)$, considerada como função de θ. Em particular, se $X_1, \ldots, X_n$ é uma amostra aleatória da densidade $f(x|\theta)$, então a função de verosimilhança é

$$\mathcal{L}(x_1,\ldots,x_n|\theta) = \prod_{j=1}^{n} f\,(x_j|\theta)$$

(2.19)

Tomando o logaritmo da função de verosimilhança, temos a função de log-verosimilhança, dada por

$$\hat{\ell}(x_1,\ldots,x_n|\theta) = ln\ \mathcal{L} = \sum_{j=1}^{n} ln\ f(x_j|\theta)$$

(2.20)

A seguir, aplicados a EMV para estimar os parâmetros da distribuição de Weibull triparamétrica.

Substituindo a f.d.p. de Weibull triparamétrica (1.7) na função (2.20), obteve-se

$$\begin{aligned}
\hat{\ell}(v_1,\ldots,v_n|k,c,u) &= \hat{\ell}(v_1,\ldots,v_n|\theta) = \sum_{j=1}^{n} ln\ f(v_j|k,c,u) \\
&= \sum_{j=1}^{n} ln\left(\frac{k}{c}\left(\frac{v_j-u}{c}\right)^{k-1} e^{-\left(\frac{v_j-u}{c}\right)^{k}}\right) \\
&= \sum_{j=1}^{n} ln\left(\frac{k}{c^k}\left(v_j-u\right)^{k-1} e^{-\left(\frac{v_j-u}{c}\right)^{k}}\right) \\
&= \sum_{j=1}^{n} ln(k) - \sum_{j=1}^{n} ln(c^k) + \sum_{j=1}^{n} ln\left((v_j-u)^{k-1}\right) - \sum_{j=1}^{n}\left(\frac{v_j-u}{c}\right)^{k} \\
&= n(ln\ k) - nk(ln\ c) + (k-1)\sum_{j=1}^{n} ln(v_j-u) - c^{-k}\sum_{j=1}^{n}(v_j-u)^{k}
\end{aligned}$$

(2.21)

Encontrando a primeira derivada de $\hat{\ell}(v_1,\ldots,v_n|k,c,u)$ e igualando a zero tem-se:

$$\frac{\partial \hat{\ell}}{\partial c} = -\frac{nk}{c} + \frac{k}{c^{k+1}} \sum_{j=1}^{n} (v_j - u)^k = 0$$

$$\Rightarrow c = \sqrt[k]{\frac{1}{n} \sum_{j=1}^{n} (v_j - u)^k}$$

(2.22)

E

$$\frac{\partial \hat{\ell}}{\partial k} = \frac{n}{k} - n(ln\ c) + \sum_{j=1}^{n} ln(v_j - u) + \frac{ln\ c}{c^k} \sum_{j=1}^{n} (v_j - u)^k - c^{-k} \sum_{j=1}^{n} (v_j - u)^k ln(v_j - u) = 0$$

Substituindo a equação (2.22), pode-se escrever como

$$g_1(k, u) = \frac{1}{k} + \frac{1}{n} \sum_{j=1}^{n} ln(v_j - u) - \frac{\sum_{j=1}^{n} (v_j - u)^k ln(v_j - u)}{\sum_{j=1}^{n} (v_j - u)^k} = 0$$

(2.23)

Da mesma forma,

$$\frac{\partial \hat{\ell}}{\partial u} = (k-1) \sum_{j=1}^{n} \frac{1}{u - v_j} + kc^{-k} \sum_{j=1}^{n} (v_j - u)^k - 1$$

Então, pode escrever-se como

$$g_2(k, u) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{v_j - u} \cdot \frac{\sum_{j=1}^{n} (v_j - u)^k}{\sum_{j=1}^{n} (v_j - u)^{k-1}} + \frac{k}{1-k} = 0$$

(2.24)

Pode-se estimar os parâmetros usando um software estatístico especificado ou resolvendo os dois simultaneamente. Por exemplo, o método de Newton-Raphson com uma boa estimativa inicial de $k^{(0)}$ e $u^{(0)}$, e atualização pela seguinte relação iterativa [24]

$$\theta^{(t+1)} = \theta^{(t)} - \frac{g(\theta^{(t)})}{g'(\theta^{(t)})}$$

(2.25)

até que a mudança em duas etapas sucessivas seja insignificante, onde t é o número da iteração,

$$\theta = \begin{bmatrix} k \\ u \end{bmatrix},\ g(\theta) = \begin{bmatrix} g_1(k,u) \\ g_2(k,u \end{bmatrix},\ g'(\theta) = \begin{bmatrix} \frac{\partial g_1}{\partial k} & \frac{\partial g_1}{\partial u} \\ \frac{\partial g_2}{\partial k} & \frac{\partial g_2}{\partial u} \end{bmatrix}$$

2.3.2 Especificação do Algoritmo EM

O algoritmo EM pode ser aplicado nos casos onde se deseja estimar algum conjunto de parâmetros que descreve uma certa distribuição de probabilidade conjunta, dada somente uma parte observada dos dados produzidos por esta distribuição. Para ser mais preciso, considere que $X = \{x_1, \ldots, x_n\}$ denota os dados observados em num conjunto de n instâncias ocorridas independentemente, seja $Z = \{z_1, \ldots, z_n\}$ os dados não observados nestas mesmas instâncias, e seja $Y = (X \cup Z)$ o total de dados. Z pode ser tratada como uma variável aleatória cuja distribuição de probabilidade depende do conjunto de parâmetros desconhecido θ e dos dados observados X. Analogamente, Y é uma variável aleatória porque é definida em termos da variável aleatória Z [25].

O algoritmo EM [26-28] visa, iterativamente, maximizar a função de verosimilhança $\mathcal{L}(X|\theta)$ com respeito a θ.

- Seja $\theta^{(t)}$ a estimativa de máximo na iteração t, para $t = 0,1, \ldots$
- Defina $Q(\theta|\theta^{(t)})$ a esperança da função de log-verosimilhança para dados completos, condicional aos dados observados X. Isto é,

$$Q\left(\theta\middle|\theta^{(t)}\right) = E\left[ln\,\mathcal{L}(\mathrm{Y}|\theta)\middle|X, \theta^{(t)}\right] \tag{2.26}$$

Na sua forma geral, o algoritmo EM é inicializado com $\theta^{(0)}$ e então altera entre passos de esperança e maximização. O algoritmo pode ser descrito como:

1. Iniciar $\theta^{(0)}$ e fazer $t = 0$;
2. **Passo E**: *Expectation(E):* Calcular $Q(\theta|\theta^{(t)})$;
3. **Passo M**: *Maximization(M)*: Obter $\theta^{(t+1)}$ ao maximizar $Q(\theta|\theta^{(t)})$ com respeito a θ a fazer $t = t + 1$;

4. Repetir passos E e M até obter convergência para a sequência de estimativas $\theta^{(t)}$, segundo algum critério de paragem.

Pode-se considerar os critérios citados anteriormente nos métodos de maximização, usualmente como

$$(\theta^{(t+1)} - \theta^{(t)})^T (\theta^{(t+1)-\theta(t)}) \tag{2.27}$$

ou

$$\left|Q(\theta^{(t+1)}|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)})\right| \tag{2.28}$$

2.3.3 Função fitmixture do R

No *package 'ForestFit'* do R existe a função *fitmixture* que estima os parâmetros do modelo de mistura usando o algoritmo de EM [29].

Forma:

fitmixture(data, family, K, initial=FALSE, starts)

Argumentos:

data	Vetor de observações.
family	Nome da família incluindo: "birnbaum-saunders", "burrxii", "chen", "f", "Frechet", "gamma", "gompetrz", "log-normal", "log-logistic", "lomax", "skew-normal" e "weibull".
K	Número de componentes.
initial	A sequência de valores iniciais dos parâmetros. Por padrão, os valores iniciais são determinados automaticamente por k-means método de agrupamento.

starts	Se initial=TRUE, então a sequência dos valores iniciais deve ser fornecida.

2.3.3.1 Exemplo da aplicação da função fitmixture

Um exemplo de uma simulação de mistura de Weibull com duas componentes Weibull pode ser visto na Figura 2.1, onde $k_1 = 3, c_1 = 1, k_2 = 6, c_2 = 3, \lambda_1 = 0.5$.

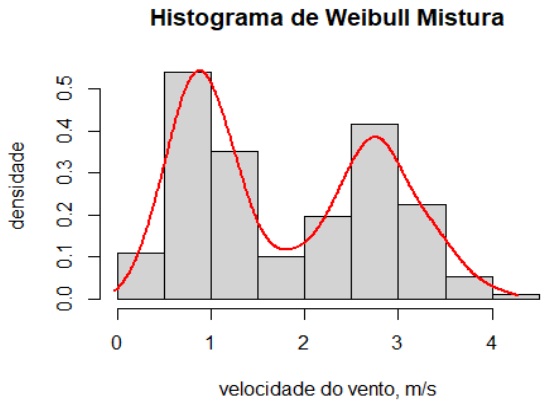

Figura 2. 1 - Exemplo do histograma da Weibull mistura

Aplicado a função fitmixture, obteve-se

```
library(ForestFit)
K <- 2
est_exe <- fitmixture(exemplo,"weibull",K,initial = FALSE)
est_exe$estimate

##         weight    alpha      beta
## [1,] 0.4979645 3.349708 0.9908677
## [2,] 0.5020355 5.842069 2.9593797
```

2.4 Análise Estatística

Nesta secção apresenta-se os conceitos básicos do erro absoluto e análise do melhor ajustamento.

2.4.1 Erro Absoluto

Ao utilizar dispositivos de medição ou fórmulas matemáticas para estimar dados de magnitudes físicas, ocorrem erros numéricos devido aos problemas inerentes ao uso de cada método. De seguir, são descritos o erro absoluto.

O valor da diferença entre o valor estimado e o valor observado na realidade é conhecido como erro absoluto (Equação (2.29)). Ele tem as mesmas unidades da magnitude analisada, e o erro absoluto quantifica o desvio do valor real em termos absolutos [30].

$$e_a = |V_e - V_r| \tag{2.29}$$

Sendo V_e é o valor estimado e V_r é o valor real.

2.4.2 Análise do Melhor Ajustamento

Depois de ajustar os dados de diferentes modelos como linear, polinomial, exponencial, distribuição, ou outros, é necessário avaliar o ajuste usando análises gráfica e numérica. A análise numérica é feita usando o cálculo dos valores estatísticos como:

- Soma dos Quadrados do Erro (SSE): Esta é uma medida da diferença entre os dados e um modelo de ajuste. Na Equação (2.30) , y_i é o valor de dados observados, $\hat{y}_i$ é o valor previsto do ajuste.

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.30}$$

- Coeficiente de Determinação (R-quadrado): mede o quão bem-sucedido é o ajuste para explicar a variação dos dados. Na Equação (2.31), SST é soma total dos quadrados (total sum of squares) e $\bar{y}_i$ é a média dos dados observados.

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{2.31}$$

- Raiz do Erro Quadrático Médio (RMSE): É uma estimativa do desvio padrão da componente aleatória nos dados. O RMSE é definido pela Equação (2.32), onde n é o número de observações.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{2.32}$$

Considera-se o melhor ajuste quando o RMSE é próximo a zero, e o R^2 tem um valor perto de um [30].

Capítulo 3

3. Estudo de Simulação

Wang Songyan e Yu Jilai escolheram calcular os parâmetros da mistura de duas componentes Weibull triparamétrica. Primeiro utilizaram o método de momento para obter o valor inicial dos parâmetros, em seguida, fizeram a estimativa de máxima verosimilhança para calcular os parâmetros, depois disso, usaram o método da teoria dos números (number thery method) para otimizar os parâmetros [21].

Mahdi Teimouri adotou só a mistura de Weibull biparamétricas, e utilizou o algoritmo EM para estimar os parâmetros [29].

Optou-se por adotar a mistura de Weibull biparamétrica neste trabalho, neste caso, a f.d.p. de Weibull mistura pode descrita como:

$$f_{mix}(v) = \sum_{i=1}^{m} \lambda_i \, f_i(v|c_i, k_i) \tag{3.1}$$

Neste capítulo, usou-se o RStudio na simulação dos diferentes casos de estudo a tipificar.

3.1 Simulação da Mistura de Distribuições Weibull

Criou-se uma função *misturaWeibull* para misturar duas componentes de distribuição de Weibull, onde n é o número da amostra aleatória, $k1$ e $c1$ são os parâmetros da distribuição Weibull 1, $k2$ e $c2$ são os parâmetros da distribuição Weibull 2, e lam1 é o peso da distribuição Weibull 1.

```
misturaWeibull <- function(n,k1,c1,k2,c2,lam1){
  set.seed(127)
  v1 <- rweibull(n,k1,c1) #Weibull 1
  v2 <- rweibull(n,k2,c2) #Weibull 2

  x <- vector(mode="numeric",length= n)
```

```
  aleatorio <- runif(n)
  for ( i in 1:n ){
    if(aleatorio[i] < lam1 ) {
      x[i] <- v1[i]
    }
    else{
      x[i] <- v2[i]
    }
  }
  return (x)
}
```

Considerando um número de observações igual a 500, foram feitas várias simulações com diferentes valores para os parâmetros da mistura de Weibull. Por fim, foi feita uma tipificação e escolheram-se 4 casos para a análise.

Caso 1:

$k_1 = 2, c_1 = 1, k_2 = 5, c_2 = 3, \lambda_1 = 0.6$

```
num=500
x1 <- misturaWeibull(num,2,1,5,3,0.6)
fdp1 <- density(x1,bw=0.2,cut=0.22)

#histograma e grafico da linha
hist(x1,freq=F,xlab="velocidade do vento, m/s",ylab="densidade",main = "H
istograma de Weibull Mistura")
lines(fdp1,col="red", lwd=2)
legend('topright',
       legend = "f.d.p. de Weibull Mistura",lty=c(1,5),
       col= "red")
```

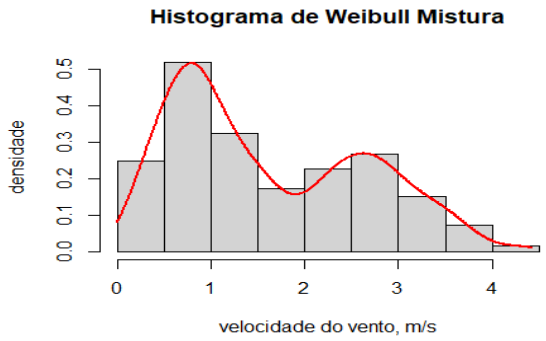

Figura 3. 1 - histograma de Weibull mistura do caso 1

Neste caso, o gráfico parece bimodal, ou seja, há uma mistura tem duas modas bem vincadas, logo é mais fácil de perceber que há uma mistura das duas componentes de distribuição de Weibull.

Caso 2:

$k_1 = 2, c_1 = 1, k_2 = 4, c_2 = 2, \lambda_1 = 0.5$

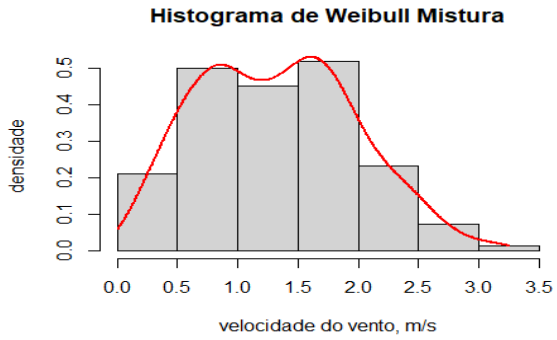

Figura 3. 2 - histograma de Weibull mistura do caso 2

Neste caso, o gráfico parece bimodal, mas os dois picos estão muito próximos. Não é tão fácil de perceber que há uma mistura de duas componentes de distribuição de Weibull.

Caso 3:

$k_1 = 2, c_1 = 1, k_2 = 3, c_2 = 2, \lambda_1 = 0.6$

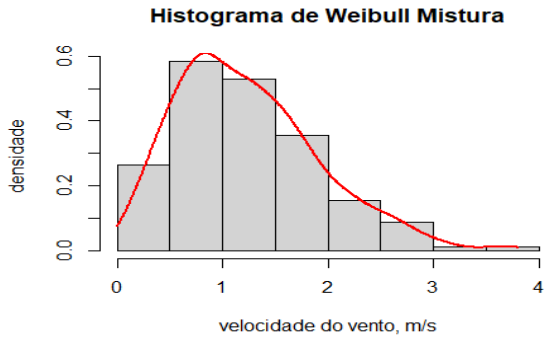

Figura 3. 3 - histograma de Weibull mistura do caso 3

Neste caso há uma primeira moda, mas depois decai sem voltar a subir, como se fosse uma Weibull pura, não sendo tão evidente a presença da outra Weibull.

Caso 4:

$k_1 = 2, c_1 = 1, k_2 = 6, c_2 = 1, \lambda_1 = 0.5$

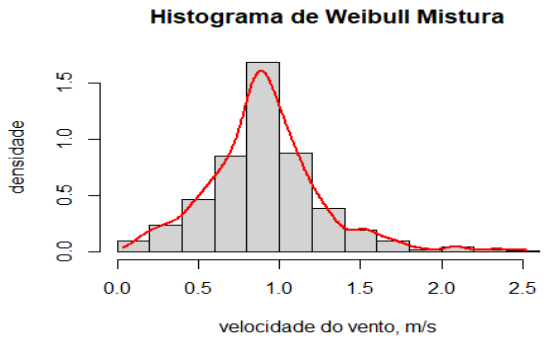

Figura 3. 4 - histograma de Weibull mistura do caso 4

Neste último caso, como os picos são mais ou menos sobreposto, parece uma unimodal. Em que é difícil identificar uma mistura, isto porque se utilizou o mesmo parâmetro escala.

3.2 Aplicando o MM nos 4 casos

3.2.1 Códigos das Funções que usar no MM

Primeiramente, criou-se uma função *MM* para obter os parâmetros da distribuição de Weibull biparamétrica, note-se $dado$ é o vetor de observações.

```
MM <- function(dado){

  v_media <- mean(dado)  #velocidade médias
  variancia <- var(dado) #variancia das duas observacoes

  #parametro de forma
  f_km <- function(x,aux){
    return((gamma(1+1/x))**2/(gamma(1+2/x))-aux)
  }
  aux <- (v_media)^2/(variancia+(v_media)^2)
  paramk <- uniroot(f_km,c(0.1,10),aux=aux,tol=0.1)
  km <- paramk$root

  #parametro de escala
  cm <- v_media/(gamma(1+1/km))

  output <- list(k=km, c=cm)
  return(output)
}
```

Seguidamente, criou-se uma função *Pminimo* para obter o ponto mínimo do gráfico mistura, onde x é o vetor das abscissas do gráfico mistura e y é o vetor das ordenadas do gráfico mistura.

```
Pminimo <- function(x,y){
  n <- length(x)
  d <- vector(mode="numeric",length= n) #derivadas
  for ( i in 2:n ){
    d[i] <- (y[i]-y[i-1])/(x[i]-x[i-1])
  }
  i <- 2
  while ( (d[i]*d[i+1]>0) || (d[i]>0 && d[i+1]<0) ) {
    i = i+1
  }
  dim <- i
```

```
  return (x[dim]) #o ponto mínimo
}
```

De seguinte, criou-se uma função *MM_mistura* para obter os parâmetros da mistura de Weibull biparamétrica, note-se $dado$ é o vetor de observações.

```
MM_mistura <- function(dado){
  den <- density(dado)
  den_x <- den$x
  den_y <- den$y
  v_divide <- Pminimo(den_x,den_y)
  V1 <- subset(dado, dado <= v_divide)
  V2 <- subset(dado, dado > v_divide)

  MM1 <- MM(V1) #obter k1 e c1 da amostra V1
  MM2 <- MM(V2) #obter k2 e c2 da amostra V2

  #peso
  n1 <- length(V1)
  n2 <- length(V2)
  lambda1 <- n1/(n1+n2)

  output <- list(k1 = MM1$k, c1 = MM1$c, k2 = MM2$k, c2=MM2$c, lambda1 =
lambda1)
  return(output)
}
```

3.2.2 Resultados dos parâmetros pelo MM

Nos casos (1) e (2) usou-se a função *MM_mistura*, pois os seus gráficos são claramente bimodais. Já nos casos (3) e (4), adotou-se a função *MM* porque nesses casos a bimodalidade não aparece de forma clara, levando a crer que se trata de uma Weibull pura.

Caso 1:

```
param1 <- MM_mistura(x1)
param1

## $k1
## [1] 2.21995
##
## $c1
## [1] 1.015485
##
## $k2
## [1] 6.207623
##
## $c2
## [1] 3.010897
##
## $lambda1
## [1] 0.62
```

Parâmetros	k_1	$c_1(m/s)$	k_2	$c_2(m/s)$	λ_1	λ_2
Real	2	1	5	3	0.6	0.4
Estimado	2.220	1.015	6.208	3.011	0.62	0.38
Erro Absoluto	0.22	0.015	1.208	0.011	0.02	0.02

Ao analisar os resultados, concluiu-se que os resultados de $k_1, c_1, c_2, \lambda_1 \; e \; \lambda_2$ são aproximados aos valores reais, enquanto que o resultado do k_2 não está de acordo com o valor real.

De seguida, criou-se uma função *Analise_numerica* para determinar o RMSE e R-quadrado, onde obs é o vetor de valores observados da variável que está sendo prevista, e $prev$ é o vetor de valores previstos.

A função *fdpWeibullmis2* desenha os gráficos das f.d.p. da mistura de Weibull bimodal com os parâmetros do estimado e dos dados observados. Para se ter uma avaliação da qualidade do ajustamento, também se imprime o RMSE e R-quadrado. Nessa função, $k1, c1, k2, c2\ e\ lam$ são os parâmetros estimados, e $fdpmis$ é a f.d.p. dos dados observados.

```
Analise_numerica <- function(obs,prev){

  #Soma dos Quadrados do Erro
  SSE <- sum((obs-prev)^2)

  #Erro Quadrático Médio
  MSE <- SSE/length(obs)

  #Raiz do Erro Quadrático Médio
  RMSE <- sqrt(MSE)

  #Soma total dos Quadrados
  SST <- sum((obs-mean(obs))^2)

  #Coeficiente de Determinação
  R_Quadrado <- 1-SSE/SST

  output <- list(RMSE = RMSE, R_Quadrado = R_Quadrado)
  return(output)
}



fdpWeibullmis2 <- function(k1,c1,k2,c2,lam,fdpmis){

  x <- fdpmis$x
  f <- lam*k1/(c1^k1)*x^(k1-1)*exp(-(x/c1)^k1)+(1-lam)*k2/(c2^k2)*x^(k2-1
)*exp(-(x/c2)^k2)

  plot(x,f,type="l",col="red",xlab="velocidade do vento, m/s",ylab="densi
dade de probabilidade", main='f.d.p. de Weibull Mistura',ylim=c(0,max(fdp
mis$y)+0.1))

  lines(fdpmis,col = "blue")

  legend('topright',legend = c('parametros do estimado','real'),text.col
= c('red','blue'),lty=c(1), col= c('red','blue'))

  real <- fdpmis$y
```

```
  #Analise numerica
  print(Analise_numerica(real,f))
}

fdp1_MM <- fdpWeibullmis2(param1$k1,param1$c1,param1$k2,param1$c2,param1$
lambda1,fdp1)
```

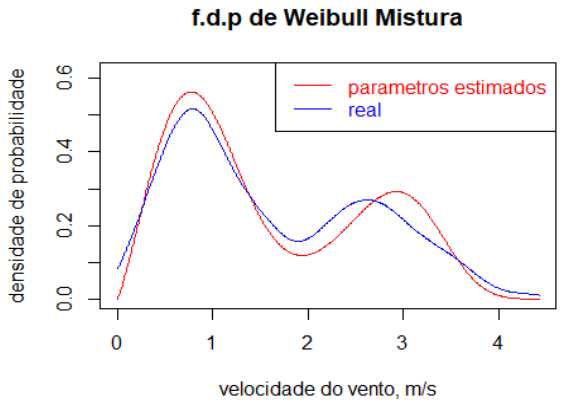

Figura 3. 5 - gráfico da f.d.p. Weibull com parâmetros MM do caso 1

```
## $RMSE
## [1] 0.04243558
##
## $R_Quadrado
## [1] 0.9048485
```

Ao observar a Figura 3.5, concluiu-se que os gráficos são semelhantes, embora se verifique um ligeiro desfasamento na localização da segunda moda. E, como o $RMSE \approx 0.042 < 0.05$ é próximo de zero e o $R^2 \approx 0.905 > 0.9$ tem um valor elevado, logo os resultados do MM aproximam-se dos valores reais.

Caso 2:

```
param2 <- MM_mistura(x2)
param2

## $k1
## [1] 2.793592
##
## $c1
## [1] 0.8142821
##
```

```
## $k2
## [1] 4.930177
##
## $c2
## [1] 1.972431
##
## $lambda1
## [1] 0.44
```

Parâmetros	k_1	$c_1(m/s)$	k_2	$c_2(m/s)$	λ_1	λ_2
Real	2	1	4	2	0.5	0.5
Estimado	2.794	0.814	4.930	1.972	0.44	0.56
Erro Absoluto	0.794	0.186	0.93	0.028	0.06	0.06

Ao analisar os resultados, concluiu-se os resultados de $c_2, \lambda_1 \, e \, \lambda_2$ são aproximados aos valores reais, enquanto que os resultados dos outros parâmetros não estão de acordo com o real.

De seguida, apresentam-se os gráficos das f.d.p. de Weibull mistura bimodal com os parâmetros estimados e com os parâmetros utilizados na simulação. Mais uma vez, para efeitos de avaliação do modelo, também se apresenta o RMSE e R-quadrado.

```
fdp2_MM <- fdpWeibullmis2(param2$k1,param2$c1,param2$k2,param2$c2,param2$
lambda1,fdp2)
```

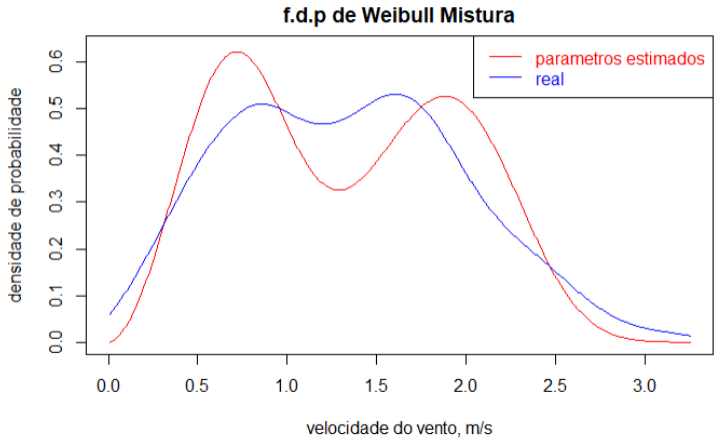

Figura 3. 6 - gráfico da f.d.p. Weibull com parâmetros MM do caso 2

```
## $RMSE
## [1] 0.08993491
##
## $R_Quadrado
## [1] 0.75547
```

Pela observação da Figura 3.6, verificou-se que os gráficos não são semelhantes. E, como o $RMSE \approx 0.090 > 0.05$ não é próximo de zero e o $R^2 \approx 0.755 < 0.9$ não é perto de um, logo os resultados do MM não são muito satisfatórios para este caso de estudo.

Caso 3:

```
param3 <- MM(x3)
param3

## $k
## [1] 1.908232
##
## $c
## [1] 1.392719
```

A função *fdpWeibull* desenha os gráficos das f.d.p. Weibull unimodal com os parâmetros estimados e com os dados observados. Também se imprime o RMSE e R-quadrado, onde k *e* c são os parâmetros estimados, e fdp é a f.d.p. dos dados observados.

```
fdpWeibull <- function(k,c,fdp){
  x <- fdp$x
  f <- k/(c^k)*x^(k-1)*exp(-(x/c)^k)

  plot(x,f,type="l",col="red",xlab="velocidade do vento, m/s",ylab="densi
dade de probabilidade", main='f.d.p de Weibull',ylim=c(0,max(fdp$y)+0.01)
)

  lines(fdp,col = "blue")

  legend('topright',legend = c('parametros do estimado','real'),text.col
= c('red','blue'), lty=c(1),col= c('red','blue'))

  real <- fdp$y

  #Analise numerica
```

```
  print(Analise_numerica(real,f))
}

fdp3_MM <- fdpWeibull(param3$k,param3$c,fdp3)
```

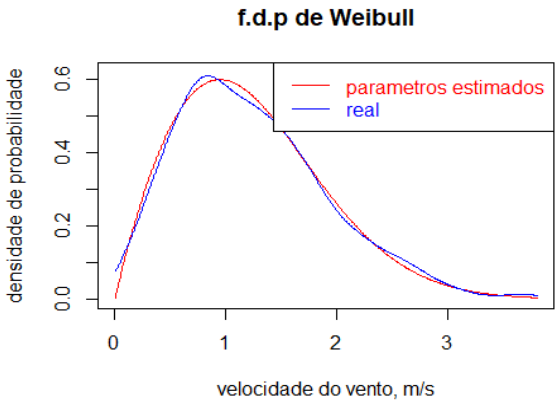

Figura 3. 7 - gráfico da f.d.p. Weibull com parâmetros MM do caso 3

```
## $RMSE
## [1] 0.01563063
##
## $R_Quadrado
## [1] 0.9943893
```

Ao observar a Figura 3.7, concluiu-se que os gráficos são semelhantes. E, como o $RMSE \approx 0.016 < 0.05$ é próximo de zero e o $R^2 \approx 0.994 > 0.9$ é perto de um, logo os resultados do MM aproximam-se dos valores reais.

Caso 4:

```
param4 <- MM(x4)
param4

## $k
## [1] 2.798647
##
## $c
## [1] 1.020667
```

De seguida, apresentam-se os gráficos das f.d.p. Weibull unimodal com os parâmetros estimados e com os dados da simulação, também se mostra o RMSE e R-quadrado.

```
fdp4_MM <- fdpWeibull(param4$k,param4$c,fdp4)
```

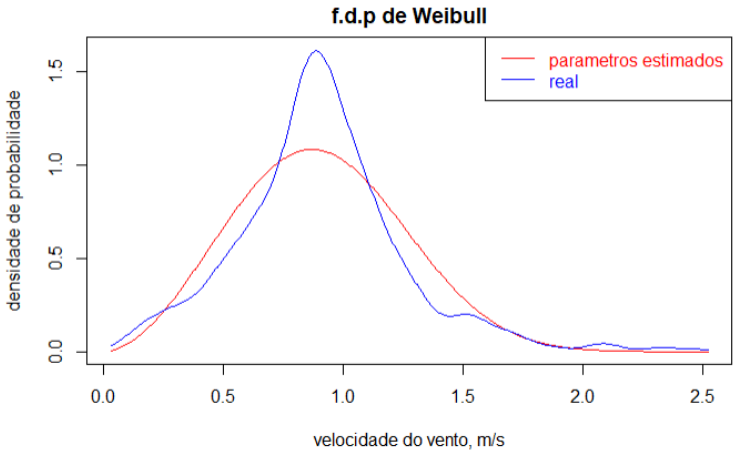

Figura 3. 8 - gráfico da f.d.p. Weibull com parâmetros MM do caso 4

```
## $RMSE
## [1] 0.1615804
##
## $R_Quadrado
## [1] 0.8776091
```

Pela observação da Figura 3.8, verificou-se que os gráficos não são semelhantes. E, como o $RMSE \approx 0.162 > 0.05$ não é próximo de zero e o $R^2 \approx 0.878 < 0.9$ não é perto de um, logo os resultados do MM não são muito bons neste caso.

Pelo estudo dos 4 casos tipificados, concluiu-se que o MM tem melhores resultados nos casos (1) e (3).

3.3 Aplicando o Algoritmo EM nos 4 Casos

Usou-se a função *fitmixture* para obter os parâmetros. Nos casos (1) e (2) considerou-se o número de componentes K igual a 2, pois os gráficos apontam claramente para uma distribuição bimodal. Nos casos (3) e (4) adotou-se um número de componentes K igual a 1, porque nesses casos o gráfico aponta para uma distribuição unimodal.

Caso 1:

```
library(ForestFit)

## Package 'ForestFit' version 0.7.1
## Type 'citation("ForestFit")' for citing this R package in publications
.

K <- 2
sepa1 <- fitmixture(x1,"weibull",K,initial = FALSE)
sepa1$estimate

##         weight    alpha      beta
## [1,] 0.4080005 5.304524 2.9907731
## [2,] 0.5919995 2.270710 0.9888055
```

Parâmetros	k_1	$c_1(m/s)$	k_2	$c_2(m/s)$	λ_1	λ_2
Real	2	1	5	3	0.6	0.4
Estimado	2.271	0.989	5.305	2.991	0.592	0.408
Erro Absoluto	0.271	0.011	0.305	0.009	0.008	0.008

Ao analisar os resultados, concluiu-se os resultados de $c_1, k_2, c_2, \lambda_1 \; e \; \lambda_2$ são aproximados ao real, enquanto que o resultado do k_1 não está de acordo ao real.

De seguida, apresentam-se os gráficos das f.d.p. de Weibull mistura bimodal com os parâmetros estimados e com os dados da simulação, também se mostra o RMSE e R-quadrado.

```
fdp1_EM <- fdpWeibullmis2(2.270710,0.9888055,5.304524,2.9907731,0.5919995
,fdp1)
```

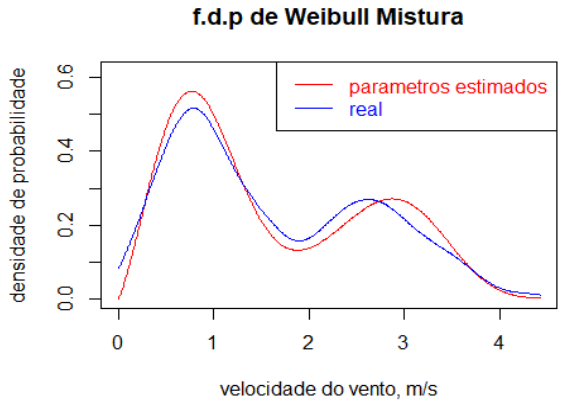

Figura 3. 9 - gráfico da f.d.p. Weibull com parâmetros EM do caso 1

```
## $RMSE
## [1] 0.03491624
##
## $R_Quadrado
## [1] 0.9355816
```

Ao observar a Figura 3.9, concluiu-se que os gráficos são semelhantes, mais uma vez com o ligeiro desfasamento na localização da segunda moda. E, como o $RMSE \approx 0.035 < 0.05$ é próximo de zero e o $R^2 \approx 0.936 > 0.9$ é perto de um, logo os resultados do algoritmo EM aproximam-se dos valores reais.

Caso 2:

```
##         weight    alpha      beta
## [1,] 0.2990056 3.075833 0.6848172
## [2,] 0.7009944 3.775008 1.8679812
```

Parâmetros	k_1	$c_1(m/s)$	k_2	$c_2(m/s)$	λ_1	λ_2
Real	2	1	4	2	0.5	0.5
Estimado	3.076	0.685	3.775	1.868	0.299	0.701
Erro Absoluto	1.076	0.315	0.225	0.132	0.201	0.201

Ao analisar os resultados, concluiu-se os resultados não estão de acordo ao real.

De seguinte, apresentam-se os gráficos das f.d.p. de Weibull mistura bimodal com os parâmetros estimados e com os dados da simulação, também se apresenta o RMSE e R-quadrado.

```
fdp2_EM <- fdpWeibullmis2(3.075833,0.6848172,3.775008,1.8679812,0.2990056
,fdp2)
```

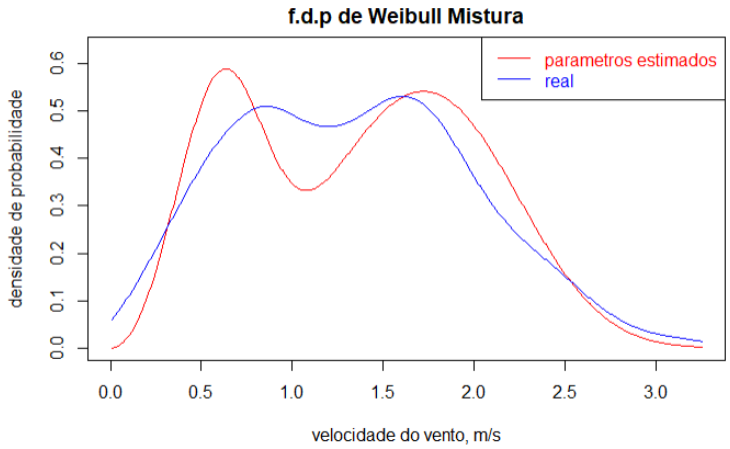

Figura 3. 10 - gráfico da f.d.p. Weibull com parâmetros EM do caso 2

```
## $RMSE
## [1] 0.07521525
##
## $R_Quadrado
## [1] 0.8289641
```

Pela observação da Figura 3.10, verificou-se que os gráficos não são semelhantes. E, como o $RMSE \approx 0.075 > 0.05$ não é próximo de zero e o $R^2 \approx 0.829 < 0.9$ não é perto de um, logo os resultados do algoritmo EM não são muito bons neste caso.

Caso 3:

```
##      weight    alpha     beta
## [1,]      1 1.906597 1.393243
```

De seguida, apresentam-se os gráficos das f.d.p. Weibull unimodal com os parâmetros estimados e com os dados da simulação, também se mostra o RMSE e R-quadrado.

```
fdp3_EM <- fdpWeibull(1.906597,1.393243,fdp3)
```

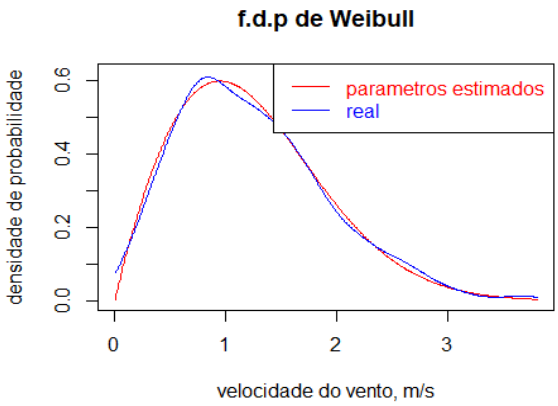

Figura 3. 11 - gráfico da f.d.p. Weibull com parâmetros EM do caso 3

```
## $RMSE
## [1] 0.01557755
##
## $R_Quadrado
## [1] 0.9944274
```

Ao observar a Figura 3.11, concluiu-se que os gráficos são semelhantes. E, como o $RMSE \approx 0.016 < 0.05$ é próximo de zero e o $R^2 \approx 0.994 > 0.9$ é perto de um, logo os resultados do algoritmo EM se aproximam dos valores reais utilizados na simulação.

Caso 4:

```
##      weight    alpha     beta
## [1,]      1 2.692295 1.018202
```

De seguida, apresentam-se os gráficos das f.d.p. Weibull unimodal com os parâmetros estimados e com os dados da simulação, também se apresenta o RMSE e R-quadrado.

```
fdp4_EM <- fdpWeibull(2.692295,1.018202,fdp4)
```

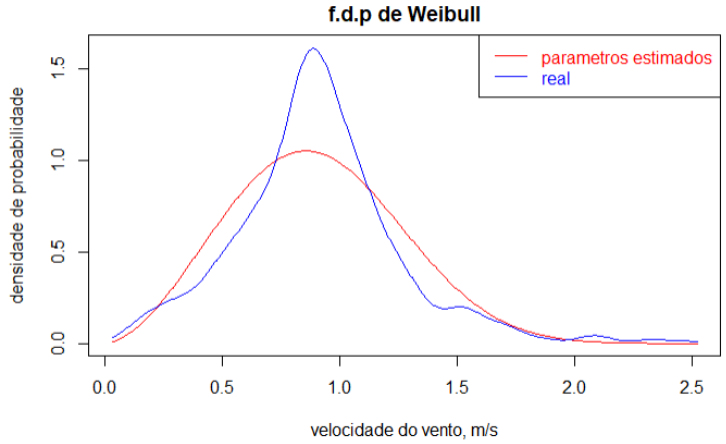

Figura 3. 12 - gráfico da f.d.p. Weibull com parâmetros EM do caso 4

```
## $RMSE
## [1] 0.1720594
##
## $R_Quadrado
## [1] 0.8612194
```

Pela observação da Figura 3.12, verificou-se que os gráficos não são semelhantes, embora tenha sido possível identificar a localização da moda. E, como o $RMSE \approx 0.172 > 0.05$ não é próximo de zero e o $R^2 \approx 0.861 < 0.9$ não é perto de um, logo os resultados do algoritmo EM não são muito bons na identificação dos parâmetros neste caso.

Pelo estudo dos 4 casos tipificados, concluiu-se que o algoritmo EM também é melhor para aplicar nos casos (1) e (3).

3.4 Comparação dos Métodos de Estimação

Nesta secção, comparou-se os resultados do MM e do algoritmo EM.

Resultados do MM:

Caso	k_1	$c_1(m/s)$	k_2	$c_2(m/s)$	λ_1	λ_2	RMSE	R^2
1	2.220	1.015	6.208	3.011	0.62	0.38	0.042	0.905
2	2.794	0.814	4.930	1.972	0.44	0.56	0.090	0.755
3	1.908	1.393	-	-	-	-	0.016	0.994
4	2.799	1.021	-	-	-	-	0.162	0.878

Resultados do algoritmo EM:

Caso	k_1	$c_1(m/s)$	k_2	$c_2(m/s)$	λ_1	λ_2	RMSE	R^2
1	2.271	0.989	5.305	2.991	0.592	0.408	0.035	0.936
2	3.076	0.685	3.775	1.868	0.299	0.701	0.075	0.829
3	1.907	1.393	-	-	-	-	0.016	0.994
4	2.692	1.018	-	-	-	-	0.172	0.861

Comparando os resultados, verificou-se que o MM e o algoritmo EM aplicado nos casos (1) e (3) são melhores do que aplicado nos casos (2) e (4). O algoritmo EM foi o que obteve os melhores resultados nos casos (1) e (2), apresentando o menor RMSE e maior R^2. No caso (3), o RMSE e R^2 são mesmo para dois métodos. Então conseguiu-se identificar que, nos casos de estudo utilizados, a precisão do algoritmo EM é melhor do que a do método dos momentos.

Capítulo 4

4. Caso de Estudo com Dados Reais

Neste capítulo usou-se os dados reais para a elaboração do trabalho proposto, bem como os resultados obtidos de acordo com a metodologia descrita anteriormente e a discussão dos mesmos.

4.1 Velocidade de Vento

Os dados utilizados neste trabalho dizem respeito à velocidade média de vento registada pelo sistema SCADA numa turbina do parque eólico da Freita (2010). E é uma extração aleatória de todos os dados.

```
#leitura dos dados
dados <- read.csv("vento_weibull.csv", sep=";", dec=",")
vel <- dados$avr_Wind.speed
fdpmis <- density(vel,bw=0.7,cut=0.59)

#histograma e grafico da linha
hist(vel,freq=F,xlab="velocidade do vento, m/s",ylab="densidade",main = "
Histograma dos dados reais")
lines(fdpmis,col="red", lwd=2)
```

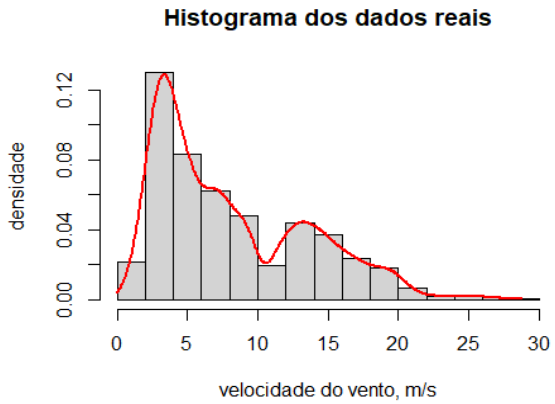

Figura 4. 1 - histograma dos dados reais

4.2 Aplicando o MM nos dados

Usou-se a função *MM_mistura*, pois o gráfico aponta para uma distribuição bimodal.

```
pardados <- MM_mistura(vel)
pardados

## $k1
## [1] 2.252574
##
## $c1
## [1] 5.585502
##
## $k2
## [1] 5.521476
##
## $c2
## [1] 16.59562
##
## $lambda1
## [1] 0.6915145
```

De seguinte, apresentam-se os gráficos das f.d.p. de Weibull mistura bimodal com os parâmetros estimados e com os dados reais, também se apresenta o RMSE e R-quadrado.

```
fdpdados_MM <- fdpWeibullmis2(pardados$k1,pardados$c1,pardados$k2,pardados$c2,pardados$lambda1,fdpmis)
```

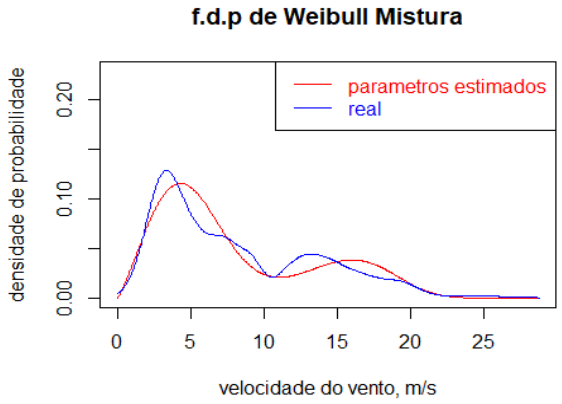

Figura 4. 2 - gráfico da f.d.p. Weibull com parâmetros MM dos dados reais

```
## $RMSE
## [1] 0.01100713
##
## $R_Quadrado
## [1] 0.8920346
```

Ao observar a Figura 4.2, concluiu-se que os gráficos são um pouco semelhantes em termos de forma, embora haja um desfasamento na localização das modas das duas componentes. Como o $RMSE \approx 0.011 < 0.05$ é próximo de zero e o $R^2 \approx 0.892 < 0.9$ tem um valor relativamente elevado, então os resultados do MM aproximam-se da distribuição observada neste caso.

4.2 Aplicando o EM nos dados

Mais uma vez, considerou-se o número de componentes K igual a 2, pois o gráfico aponta para uma distribuição bimodal.

```
sepa_dados <- fitmixture(vel,"weibull",2,initial = FALSE)
sepa_dados$estimate

##         weight    alpha      beta
## [1,] 0.3342845 4.641864 16.649334
## [2,] 0.6657155 2.283732  5.602903
```

De seguida, apresentam-se os gráficos das f.d.p. de Weibull mistura bimodal com os parâmetros estimados e com os dados reais, também mostra os o RMSE e R-quadrado.

```
fdpdados_EM <- fdpWeibullmis2(4.641864,16.649334,2.283732,5.602903,0.3342
845,fdpmis)
```

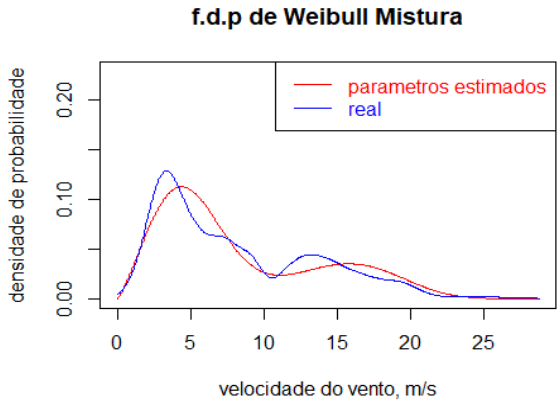

Figura 4. 3 - gráfico da f.d.p. Weibull com parâmetros EM dos dados reais

```
## $RMSE
## [1] 0.01080122
##
## $R_Quadrado
## [1] 0.8960361
```

Pela observação da Figura 4.3, verificou-se que os gráficos são um pouco semelhantes, embora continue a existir um desfasamento da localização das modas das duas componentes. E, como o $RMSE \approx 0.011 < 0.05$ é próximo de zero e o $R^2 \approx 0.896 < 0.9$ tem um valor relativamente elevado, então os resultados do algoritmo EM também se aproximam da distribuição observada na realidade.

Comparando os resultados do MM e do algoritmo EM, verificou-se o algoritmo EM foi o que obteve o melhor resultado, pois apresentou um RMSE menor e um maior R^2.

Capítulo 5

5. Conclusão

Este trabalho teve como objetivo principal a estimação dos parâmetros de uma mistura de distribuições Weibull.

A teoria associada a uma ferramenta computacional tornou possível a estimativa dos parâmetros de uma Weibull de mistura. Para isso, foram utilizados dois métodos distintos, o método dos momentos e o algoritmo EM.

Inicialmente, foram feitas várias simulações de Weibull de mistura com os valores dos parâmetros diferentes. Pelas simulações, foi feita uma análise e tipificação de diferentes possibilidades de ocorrência de distribuições bimodais. Para aplicação da metodologia, foram utilizados os 4 casos de estudo seguintes:

1) O gráfico da f.d.p. parece uma bimodal e os picos estão distantes.
2) O gráfico da f.d.p. parece uma bimodal e os picos estão próximos.
3) O gráfico da f.d.p. apresenta uma primeira moda, mas depois decai sem voltar a subir de forma muito evidente, como se fosse uma Weibull.
4) Os picos são mais ou menos sobrepostos, portanto o gráfico da f.d.p. parece uma distribuição unimodal. Neste caso, é difícil identificar uma mistura, porque estão a usar o mesmo parâmetro escala.

Depois, ao estudar os casos tipificados, verificou-se que o método dos momentos e o algoritmo EM aplicado nos casos (1) e (3) apresenta melhores resultados do que nos casos (2) e (4), também se verificou maior precisão no algoritmo EM do que método dos momentos. Para além disso, depois do caso de estudo, com dados reais recolhidos num parque eólico, também se conseguiu identificar que maior precisão do algoritmo EM do que no método dos momentos.

Em suma, no geral, os dois métodos não são aplicáveis em todas misturas de distribuições Weibull. Sendo assim, como trabalhos futuros são sugeridos o estudo e utilização de outros métodos com vista a obter melhores estimativas dos parâmetros de Weibull mistura.

Bibliografia

[1] J. Valinejad, M. Marzband, M. F. Akorede, I. D. Elliott, R. Godina, J. C. de O. Matias, E. Pouresmaeil, *Long-Term Decision on Wind Investment with Considering Different Load Ranges of Power Plant for Sustainable Electricity Energy Market*, 2018.

[2] J. Shi, W. Lee, X. Liu, *Generation Scheduling Optimization of Wind-Energy Storage System Based on Wind Power Output Fluctuation Features*, 2017.

[3] A. Baharvandi, J. Aghaei, T. Niknam, M. Shafie-Khah, R. Godina, J. P.S. Catalão, *Bundled Generation and Transmission Planning under Demand and Wind Generation Uncertainty based on a Combination of Robust and Stochastic Optimization*, 2018.

[4] C. Jichuan, D. Wei, Y. Fang, *Research on the methods estimating weibull distribution parameters*, 2011.

[5] L. Wang, F. Qian, *Shaping the PDF of the state variable based on piecewise linear control for non-linear stochastic systems*, 2015.

[6] M. Wang, L. Zeng, *Study of wind speed frequency distribution model*, 2011.

[7] J.A. Carta, P. Ramírez, *Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions*, 2007.

[8] P. R. Rodrigues, J. B. de O. Guerra, Y. A. Youssef, *Energia Eólica em Energias Renováveis*, 2011.

[9] D. C. Vaz, *Energia Eólica*, 2008.

[10] W. Weibull, *A Statistical Theory of the Strength of Materials*, 1939.

[11] R.L. Bailey, T.R. Dell, *Quantifying diameter distribution with the Weibull function*, 1973.

[12] C.G. Justus, W.R. Hargraves, A. Mikhail, D. Graber, *Methods for Estimating Wind Speed Frequency Distributions*, 1978.

[13] M. R. Patel, *Wind and Solar Power Systems*, 2006.

[14] J.F. Manwell, J.G. Mcgowan, A.L. Rogers, *Wind Energy Explained: Theory, Design and Application*, 2009.

[15] D.N. P. Murthy, M. Xie, R. Jiang, *Weibull Models*, 2004.

[16] R. A. de O. Filho, *Distribuição Espacial e Sazonal do Potencial Eólico em Minas Gerais a partir do Global Forecast System*, 2020.

[17] M. Teimouri, A. K. Gupta, *On the Three-Parameter Weibull Distribution Shape Parameter Estimation*, 2013.

[18] S. M. F. de S. Faria, *Modelo de Mistura: Aplicações em Análise de Regressão*, 2006.

[19] B.S. Everitt, D.J. Hand, *Finite Mixture Distributions*, 1981.

[20] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 2014.

[21] W. Songyan, Y. Jilai, *Aproximation of Two-peak Wind Speed Probability Density Function with Mixed Weibull Distribution*, 2010.
[22] G. Mclachlan, *Finite Mixture Models*, 2000.
[23] U. Britto, *Imputação de Dados Utilizando o Algoritmo EM e Regressão Linear no SPSS*, 2005.
[24] E. E. Elmahdy, A. W. Aboutahoun, *A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling*, 2012.
[25] T. Mitchell, *Machine Learning*, 1997.
[26] G. J. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*, 2008.
[27] A. P. Dempster, N. M. Laird, D.B. Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*, 1977.
[28] L. Bordes, D. Chauveau, *EM and Stochastic EM algorithms for reliability mixture models under random censoring*, 2012.
[29] M. Teimouri, *Package 'ForestFit'*, 2021.
[30] L. E. Q. Fernandez, *Método Genérico para Estimação e Modelagem do Erro RMS em Dados de Profundidade de Sensores para Visão 3D*, 2017.