UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Departamento de Ciências Exatas

PROFMAT - Mestrado Profissional em Matemática em Rede Nacional

Dissertação de Mestrado

A DISTRIBUIÇÃO ZETA-WEIBULL APLICADA À ANÁLISE DE SOBREVIVÊNCIA

Mateus Nascimento Ferreira

Orientador: Prof. Dra. Ana Carla Percontini da Paixão

Feira de Santana
Setembro de 2017

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Departamento de Ciências Exatas

PROFMAT - Mestrado Profissional em Matemática em Rede Nacional

A DISTRIBUIÇÃO ZETA-WEIBULL APLICADA À ANÁLISE DE SOBREVIVÊNCIA

Mateus Nascimento Ferreira

Dissertação apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT do Departamento de Ciências Exatas, UEFS, como requisito parcial para a obtenção do título de **Mestre**.

Orientador: Prof. Dra. Ana Carla Percontini da Paixão

Feira de Santana
18 de Setembro de 2017

Ficha Catalográfica – Biblioteca Central Julieta Carteado

Ferreira, Mateus Nascimento
F442d A distribuição Zeta-Weibull aplicada à análise de sobrevivência./ Mateus Nascimento Ferreira. / Feira de Santana, 2017.
74f.

Orientadora: Ana Carla Percontini da Paixão
Dissertação (Mestrado em Matemática) – Universidade Estadual de Feira de Santana. Departamento de Ciências Exatas, 2017.

1.Análise de sobrevivência. 2.Distribuição de probabilidade. 3.Distribuição de Weibull. 4. Distribuição de Zeta-Weibull. 5.Função Zeta I.Paixão, Ana Carla Percontini da, (Orient.) II.Universidade Estadual de Feira de Santana. III. Título.

CDU : 519.2

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
Departamento de Ciências Exatas
Mestrado Profissional em Matemática em Rede Nacional

PROFMAT

Ata da Sessão pública de defesa de dissertação do discente Mateus Nascimento Ferreira do Programa de Mestrado Profissional em Matemática em Rede Nacional da Universidade Estadual de Feira de Santana

Aos dezoito dias dias do mês de setembro de dois mil e dezessete às 9:00 horas na sala MT 55, UEFS , ocorreu a Sessão pública de defesa de dissertação apresentada sob o título **"A Distribuição Zeta-Weibull Aplicada à Análise de Sobrevivência"**, do discente **Mateus Nascimento Ferreira**, do Mestrado Profissional em Matemática em Rede Nacional - PROFMAT da Universidade Estadual de Feira de Santana, para obtenção do título de MESTRE. A Banca Examinadora foi composta pelos professores: Ana Carla Percontini da Paixão (Orientador, UEFS), Jailson Araujo Rodrigues (IFBA) e Aloisio Machado Da Siva Filho (UEFS). A sessão de defesa constou da apresentação do trabalho pelo discente e das arguições dos examinadores.

Em seguida, a Banca Examinadora se reuniu em sessão secreta para julgamento final do trabalho e atribuiu o conceito: APROVADO.

Sem mais a tratar, foi lavrada a presente ata, que segue assinada pelos membros da Banca Examinadora e pelo Coordenador Acadêmico Institucional do PROFMAT. Feira de Santana, 18 de setembro de 2017.

Ana Carla Percontini da Paixão

Profa. Dra. Ana Carla Percontini da Paixão (UEFS)
Orientador

Jailson de A. Rodrigues

Prof. Dr. Jailson Araujo Rodrigues (IFBA)

Aloisio Machado da S. Filho

Prof. Dr. Aloisio Machado Da Siva Filho (UEFS)

Visto do Coordenador:

Haroldo Benatti

Prof. Dr. Haroldo Gonçalves Benatti
Coordenador do PROFMAT/UEFS

Agradecimentos

Gostaria de agradecer primeiramente ao meu Senhor, pois sem ele nada seríamos e nada conseguiríamos.

Aos meus familiares, irmão, tios, primos e em especial aos meus pais, José e Zanira, pelo apoio incondicional, pelo carinho, pela paciência, pelos conselhos, pela vida.

A minha tia/madrinha Rita, pela acolhida e pelos bons papos.

A minha companheira, amiga, confidente e amor, Telma, pela paciência e incentivo.

Aos docentes do PROFMAT/UEFS pelos ensinamentos. Muito Obrigado!

A minha orientadora, professora Dra. Ana Carla Percontini da Paixão, pela confiança e seriedade com que conduziu esse processo.

Aos meus queridos colegas do PROFMAT, com os quais passei momentos maravilhosos nos últimos dois anos e que levarei para o resto da vida.

A CAPES pelo apoio financeiro.

Resumo

O objetivo deste trabalho é a criação de uma nova distribuição de probabilidade, chamada distribuição Zeta-Weibull, a partir da família de distribuições Zeta-G dada por Paixão em [15]. Para tal, temas relevantes sobre análise de sobrevivência e distribuições de probabilidades serão abordados, explorando também suas propriedades matemáticas. Serão introduzidos ainda os principais conceitos sobre a teoria da probabilidade, com ênfase aos resultados que serão utilizados ao longo do texto. Daremos destaque inicial para a função Zeta de Riemann e seu importante papel em várias áreas da pesquisa moderna, além de uma abordagem sobre a distribuição Weibull de Waloddi Weibull(1887-1979), matemático sueco reconhecido pelo seu trabalho na área de fadiga de materiais. Demonstraremos o resultado principal proposto aqui, a distribuição Zeta-Weibull, composição da classe geradora de distribuições proposta por [15] e a distribuição de Weibull. Algumas propriedades matemáticas dessa nova distribuição serão determinadas, entre elas, momentos, entropia de Rényi e confiabilidade. Mostraremos que a densidade Zeta-Weibull é uma combinação linear de densidades Weibull. Apresentaremos uma aplicação da distribuição Zeta-Weibull a um conjunto de dados reais para mostrar que ela é competitiva em relação a outras distribuições da literatura. Por fim, apresentaremos algumas sugestões de atividades para o ensino de Estatística e Probabilidade no Ensino Médio.

Palavras-chave: Análise de sobrevivência, distribuições de probabilidade, distribuição Weibull, função Zeta, Distribuição Zeta-Weibull.

Abstract

The objective of this work is the creation of a new probability distribution, called the Zeta-Weibull distribution, from the distribution family Zeta-G given by Paixão in [15]. To this end, relevant topics on survival analysis and probability distributions will be addressed, exploring their mathematical properties as well. The main concepts on probability theory will be introduced, with an emphasis on Results that will be used throughout the text. We will give initial prominence to Riemann's Zeta function and its important role in several areas of modern research, as well as an approach to the Weibull distribution of Waloddi Weibull (1887-1979), Swedish mathematician recognized for his work in the field of material fatigue. We will demonstrate the main result proposed here, the Zeta-Weibull distribution, the distribution-generating class composition proposed by [15] and the Weibull distribution. Some mathematical properties of this new distribution will be determined, among them, moments, Rényi entropy and reliability. We will show that the Zeta-Weibull density is a linear combination of Weibull densities. We will present an application of the Zeta-Weibull distribution to a set of real data to show that it is competitive in relation to other distributions in the literature. Finally, we will present some suggestions of activities for the teaching of Statistics and Probability in High School.

Keywords: Survival analysis, probability distributions, Weibull distribution, Zeta function, Zeta-Weibull distribution.

Lista de Figuras

Lista de Tabelas

Sumário

Introdução

Sempre que nos interessamos por determinado tema procuramos coletar o máximo de informação possível sobre ele, analisando em seguida essas informações para tirarmos nossas conclusões a respeito. Coletar dados, analisá-los e, posteriormente, tirar conclusões confiáveis sobre eles, é uma das atribuições da estatística, um ramo da matemática aplicada. "Estatística é uma palavra usada em uma variedade de sentidos, e muitas vezes invocada para emprestar credibilidade a outras opiniões que, de outra forma, seriam duvidosas. Essa ciência tem raízes históricas profundas, mas floresceu realmente no começo do século XX" [3].

As dúvidas sobre eventos futuros também é área de interesse da matemática, como prever fenômenos climáticos ou o percentual de nascimentos de pessoas em determinada cidade em um período específico. Nesses casos, não há como responder com certeza absoluta o que vai acontecer, pela simples razão de que o evento ainda não ocorreu. Em situações desse tipo o que é possível fazer é identificar os resultados prováveis e observar se há certa regularidade desses eventos em períodos passados, tentando obter um modelo o mais fidedigno possível para prever o que ocorrerá no futuro. A regularidade observada nesses eventos nos permite calcular o grau de certeza das previsões feitas, ou seja, a confiabilidade de que o evento irá ocorrer. A essa confiança damos o nome de probabilidade. Ainda segundo [3], na primeira parte do século XVIII a estatística e a probabilidade se desenvolveram juntas como duas áreas intimamente relacionadas com a matemática da incerteza.

Os resultados possíveis em um experimento com mesmas chances de ocorrer podem ser analisados através de uma distribuição de probabilidade, que podemos entender como uma função que associa a cada fenômeno aleatório possível sua chance de ocorrência. Se cada fenômeno aleatório tem a mesma chance de ocorrer, então a probabilidade de cada um deles deve ser a mesma, e neste caso, teremos um modelo equiprovável. Há vários modelos de distribuições de probabilidade na literatura, equiprováveis ou não, que são comumente utilizados por profissionais de áreas como saúde, educação, psicologia, engenharias, indústria, entre outras.

Uma das áreas da estatística que mais tem crescido nos últimos anos é a análise de sobrevivência. Também conhecida como confiabilidade, ela estuda basicamente o conjunto de medidas entre os tempos de ocorrência de um determinado evento, desde um tempo inicial até seu tempo final, também conhecido como tempo de falha. Desse modo, essa variável é não negativa e pode representar o tempo de falha do evento em questão. Temos, como exemplo, o estudo entre o tempo do diagnóstico de uma doença e a morte do paciente ou o tempo entre a compra de um aparelho celular até seu primeiro defeito ou ainda o tempo entre o ingresso de um aluno em um curso superior e sua diplomação, desligamento ou evasão. Em cada um desses casos, faz-se necessário especificar sempre os eventos que definem os intervalos de tempo. Uma outra característica importante da análise de sobrevivência é a presença de censura nos dados, ou seja, de objetos de estudo que não foram concluídos pelo evento de interesse ou falha. Por exemplo, numa produção de celulares, separamos um lote com 500 aparelhos, e colocando-os em condições normais e iguais de uso, observamos o tempo de falha de um determinado componente, ou seja, o tempo até que ele pare de funcionar. Percebe-se que o tempo de sobrevivência destes componentes está sujeito a variações aleatórias. Não é necessário esperar que todos os celulares tenham apresentado defeito no referido componente, e neste caso, um tipo de censura acontece, sendo formada exatamente por esses aparelhos não observados. Desse modo, vale ressaltar que a censura ocorre, pois nem sempre é possível esperar até que o evento em questão aconteça com todos os indivíduos da pesquisa.

A análise de sobrevivência é bastante utilizada atualmente em diversas áreas como nas ciências da saúde, na indústria, educação, engenharias e muitas outras. Do ponto de vista estatístico, qualquer das aplicações possíveis para a análise de sobrevivência podem ser tratados com as mesmas ferramentas.

Segundo [4], "a análise estatística de sobrevivência é um método estatístico usado para análise de dados de sobrevivência derivados de estudos de laboratórios, ou seja, ela estuda, por exemplo, o tempo em que um indivíduo sobrevive a um determinado tratamento". Esses métodos também podem ser utilizados na indústria, onde recebe o nome de Teoria da Confiabilidade. O que importa na análise de sobrevivência é indicar o tempo entre falhas. Aqui, chamaremos de *tempo de sobrevivência* ao período até a ocorrência de um determinado evento, que pode ser a morte de um paciente após o início de um determindo tratamento. *Falha* será a ocorrência do evento, como por exemplo, a quebra do equipamento eletrônico.

Nesse contexto, a obtenção de uma nova distribuição de probabilidade se faz, entre outros motivos, pela necessidade de se ter uma maneira de especificar como é a distribuição do tempo de vida em uma população de medidas. Dessa forma, segue um panorama de como será a organização deste trabalho.

O Capítulo 1 tráz os conceitos principais sobre probabilidade, como fenômenos determinísticos e aleatórios, variáveis discretas e contínuas, espaço amostral, eventos e o modelo equiprovável para um experimento aleatório, onde mostramos que cada variável possui a mesma chance de ocorrer. Ainda no Capítulo 1 introduziremos os conceitos de distribuição de probabilidade, função densidade de probabilidade, função de distribuição acumulada, função de sobrevivência, função taxa de falha, esperança, mediana, moda, variância, desvio padrão, momentos e função geratriz de momentos. No Capítulo 2 apresentamos a definição da função Zeta de Riemann, da distribuição discreta Zeta e da distribuição contínua de Weibull, além de apresentar a classe de distribuições Zeta-G proposta em [15].

No Capítulo 3, temos a nova distribuição Zeta-Weibull. Determinaremos a função densidade de probabilidade, a função de distribuição acumuldada, a função de sobrevivência e a função taxa de falha. Mostraremos que a função densidade de probabilidade da distribuição Zeta-Weibull pode ser escrita como uma combinação linear de densidades Weibull. Também encontraremos a função quantílica, além de expressões para os momentos, a função geratriz de momentos, a entropia de Rényi e para a confiabilidade. Utilizaremos o método da máxima verossimilhança para estimar os parâmetros da distribuição Zeta-Weibull.

No Capítulo 4 traremos um pouco da história e do ensino de Estatística e Probabilidade no Ensino Fundamental e Médio. Abordaremos algumas atividades comentadas com o intuito de ajudar os professores a dispertar o interesse dos alunos, instigando-os na busca do conhecimento.

Capítulo 1

Noções de Probabilidade

1.1 Conceitos Básicos

Nesta seção vamos apresentar alguns conceitos básicos sobre probabilidade que nos serão úteis nos próximos capítulos. Inicialmente, vale frisar que na natureza existem dois tipos de fenômenos: determinísticos e aleatórios. Chamaremos de experimento a qualquer ensaio ou experiência que tem por objetivo entender o comportamento de um fenômeno. Os fenômenos determinísticos são aqueles que apresentam sempre os mesmos resultados independente da quantidade de vezes que o experimento é repetido, ou seja, se o experimento não se altera o resultado também não se altera. Um exemplo de um fenômeno determinístico é a passagem do estado sólido para o estado líquido de uma pedra de gelo deixada a uma temperatura necessária para que isso ocorra. Não importa quantas vezes o fenômeno seja repetido, o resultado sempre será o mesmo.

Chamaremos de aleatórios os experimentos que quando repetidos sob as mesmas condições produzirem resultados diferentes, por exemplo, escolher uma lâmpada entre várias, de um mesmo lote de fabricação, e observar se ela queima ou não antes de 1000 horas de uso. Arremessar um dado não viciado e identificar o número que cai para cima, lançar um moeda não viciada e verificar se sai cara ou coroa, retirar uma carta do baralho e anotar a figura, ou ainda registrar a produção, em quantidade de frutos, de cada pé de laranja em uma lavoura, com as mesmas condições de temperatura, pressão, umidade e solo, são outros exemplos de experimentos aleatórios.

Os resultados dos experimentos aleatórios mencionados acima, embora pareçam acidentais, tedem a um resultado com pouca variação à medida em que o experimento é repetido um número grande de vezes. Essa afirmação pode ser facilmente traduzido pela chamada Lei dos Grandes Números, onde diz que a média aritmética de todos os valores observados em um experimento converge, em certo sentido, para um valor esperado dessas observações quando a quantidade de observações tende ao infinito. O valor esperado mencionado aqui diz respeito à Esperança matemática que será definida mais adiante. Uma leitura mais aprofundada sobre a Lei dos Grandes Números pode ser encontrada em [7]. O que se pode destacar aqui é que essa regularidade da qual nos referimos é muito importante na construção de modelos matemáticos eficientes que traduzam o mais fidedignamente possível, o comportamento do fenômeno, possibilitando assim, que previsões sejam feitas.

Chamamos de variável um elemento representante do conjunto de todos os resultados possíveis de um fenômeno, podendo indicar uma qualidade ou uma quantidade atribuída aos elementos deste conjunto. Se um dos pés de laranja do exemplo anterior tem 32 frutas, a quantidade de frutos é a variável atribuída a esse pé de laranja. Se a variável representa um atributo do objeto de estudo chamaremos ela de variável qualitativa, e se ela indica uma característica que pode ser indicada através de uma medida, então a chamaremos de variável quantitativa. As variáveis qualitativas podem ser divididas em nominal e ordinal, que podem ser caracterizadas por não existir ordenação entre as categorias de dados e por existir ordenação, respectivamente. Já as variáveis quantitativas são divididas em cotínuas, quando os dados são caracterizados por assumirem quaisquer valores em um escala contínua, ou seja, na reta real, e discretas, que podem assumir apenas um número finito ou infinito contável de valores, ou seja, assumir apenas valores inteiros. Neste trabalho, iremos nos ater às variáveis quantitativas contínuas ou simplesmente contínuas. Vejamos então outras definições necessárias.

Definição 1.1. *Chamaremos de espaço amostral Ω ao conjunto de todos os resultados possíveis em uma experiência aleatória.*

Vejamos alguns exemplos:

a) No lançamento de uma moeda honesta, $\Omega = \{cara, coroa\}$;

b) No lançamento de um dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$;

c) Na contagem de laranjas, $\Omega = \{n \in \mathbb{Z}/n \geq 0\}$;

d) Na retirada de uma carta do baralho, $\Omega = \{$as 52 cartas do baralho$\}$.

Definição 1.2. *Um subconjunto qualquer do espaço amostral será chamado de evento.*

Se um evento possui apenas um elemento será chamado de evento simples ou evento unitário. Caso possua mais de um elemento será chamado evento composto. Usaremos as letras maiúsculas do alfabeto para representar um evento. Note que no lançamento de uma moeda, os eventos são quatro: {cara }, {coroa}, {cara,coroa}$= \Omega$ e $\{\varnothing\}$. É perfeitamente aceitável imaginar que no lançamento de uma moeda saia *cara*, por exemplo. Mas é difícil imaginar que em um único lançamento saia *cara* e *coroa* ao mesmo tempo. Chamaremos de *evento certo* ao evento que contêm todos os resultados possíveis do experimento. No lançamento de um dado, o evento $A = \{1, 2, 3, 4, 5, 6\}$ é um evento certo. Se um evento é impossível de acontecer em um experimento aleatório, como por exemplo, sair *cara* e *coroa* ao mesmo tempo em um único lançamento da moeda, o chamaremos de *evento impossível.*

Definição 1.3. *Uma variável aleatória é uma função* $X : \Omega \to \mathbb{R}$ *que associa a cada evento do espaço amostral* Ω *um número real.*

Vamos relacionar a cada variável aleatória um número que chamaremos de *probabilidade* e que indicará a confiabilidade na ocorrência da referida variável.

Definição 1.4. *Uma probabilidade é uma função que associa a cada variável aleatória* X *um número* $P(X)$ *de modo que* $0 \leq P(X) \leq 1$ *e* $P(\Omega) = 1$.

Observação 1.5. *Se* A *e* B *são eventos mutuamente excludentes, ou seja, que não ocorrem simultaneamente, então* $P(A \cup B) = P(A) + P(B)$.

Exemplo 1.6. *Lança-se um dado, com seis faces numeradas de um a seis, e ao observar a face voltada para cima, temos que o espaço amostral é dado por* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *e que há* 64 *eventos dos quais podemos destacar:* $\varnothing, A_1 = \{1\}, A_2 = \{2\}, A_3 = \{3\}, A_4 = \{4\}, A_5 = \{5\}, A_6 = \{6\}$ *e* Ω. *É fato que todos os outros eventos, com duas ou mais faces ocorrendo em um único lançamento, são impossíveis de acontecer, assim como o próprio* $\varnothing$. *Nessas condições, uma probabilidade que pode ser definida é:*

$P(\varnothing) = 0,\ P(A_1) = 0,1,\ P(A_2) = 0,1,\ P(A_3) = 0,2,\ P(A_4) = 0,3,\ P(A_5) = 0,1,$ $P(A_6) = 0,2$ *e* $P(\Omega) = 1$

Observe que o modelo atende a Definição 1.4 e a Observação 1.5.

O que procuramos em um modelo probabilístico é que ele traduza nossa confiança na ocorrência de um evento, o que nos parece lógico quando em um experimento todos os eventos possíveis tenham a mesma chance de ocorrer. No exemplo 1.6, seria mais adequado se o dado tivesse sido lançado um grande número de vezes e obtido 10% dos resultados para $P(A_1), P(A_2), P(A_5)$, 20% para $P(A_3), P(A_6)$ e 30% para $P(A_4)$.

Se, em um experimento aleatório, os n elementos no espaço amostral tem a mesma chance de ocorrer, dizemos que o modelo é *equiprovável.* No exemplo 1.6, para que o modelo seja equiprovável, a probabilidade de cada face acontecer deve ser exatamente igual a $\dfrac{1}{6}$. Isso ocorrendo, nosso modelo satisfaz a Definição 1.4 e a Observação 1.5.

Em um modelo equiprovável com n elementos no espaço amostral, a probabilidade de cada um deles é dada por $\frac{1}{n}$. De fato, atende a Definição 1.4 e a Observação 1.5, uma vez que, se $\Omega = \{A_1, A_2, ..., A_n\}$ e $P(A_1) = P(A_2) = ... = P(A_n) = E$, da Observação 1.5 temos que

$$\begin{aligned}
1 &= P(\Omega) = P(\{A_1, A_2, ..., A_n\}) \\
&= P(A_1 \cup A_2 \cup ... \cup A_n) \\
&= P(A_1) + P(A_2) + ... + P(A_n) \\
&= E + E + ... + E = nE
\end{aligned}$$

e portanto,

$$E = \frac{1}{n}.$$

Caso um evento X qualquer, seja formado por k elementos, onde $k \leq n$, então

$$P(X) = \frac{k}{n}.$$

Assim, a probabilidade de um evento é a razão entre o número de casos favoráveis ao evento e o número de casos possíves. Foi esse o modelo adotado por vários matemáticos como o italiano Jerônimo Cardano (1501-1576), e os francêses Blaise Pascal (1623-1662) e Pierre Simon Laplace (1749-1827), segundo [11].

O que acontece na prática é que nem sempre é possível determinar, pela definição acima, a probabilidade da ocorrência de um evento. Qual a probabilidade de uma mulher de 30 anos fumante desenvolver câncer de pulmão, sabendo que ela reside em Camaçari/BA? Para responder a esta pergunta o que pode ser feito é observar com que frequência esse fato ocorre. A partir de um grande número de observações podemos obter uma boa estimativa da probabilidade de ocorrência desse tipo de evento. Assim, a *frequência relativa* de um evento X, denotada aqui por $F_r(X)$, é definida como a razão entre o número de vezes que o evento X ocorre pela quantidade de observações feitas do experimento.

A seguir, temos a definição frequencista de probabilidade.

Definição 1.7. *Seja A um experimento qualquer e X um evento do espaço amostral associado ao experimento A. Se o experimento é repetido n vezes, então a probabilidade do evento X é definida como*

$$P(X) = \lim_{n \to \infty} F_r(X)$$

1.2 Distribuições de Probabilidade

Uma distribuição de probabilidade é um modelo matemático que relaciona um certo valor da variável em estudo com a sua probabilidade de ocorrência. Há, logicamente, dois tipos de distribuição de probabilidade: As distribuições discretas, quando a variável que está sendo medida é discreta, e as distribuições contínuas, quando a variável que está sendo medida é expressa em uma escala contínua. Algumas das distribuições discretas mais conhecidas são a de Bernoulli, a Binomial, a Geométrica e a Poisson. As principais distribuições contínuas são a Uniforme, a Exponencial, a Normal, a Gama e a Weibull.

Dada uma variável aleatória discreta X, que pode assumir qualquer valor no conjunto $\{X_1, X_2, \ldots, X_k\}$, com $k \in \mathbb{N}$, e suas respectivas probabilidades $P_1, P_2, \ldots, P_k$, satisfazendo a Definição 1.4 e a Observação 1.5, em que $P_1 + P_2 + \ldots + P_k = 1$, então, dizemos que está definida uma distribuição de probabilidade discreta para a variável X. O conceito de função de probabilidade ou função massa de probabilidade para uma variável aleatória discreta é análogo ao conceito de função densidade de probabilidade para variáveis contínuas. No caso da função massa de probabilidade, ela associa a cada possível ocorrência de uma variável aleatória discreta uma probabilidade compreendida entre 0 e 1 inclusive.

Definição 1.8. *Se Ω é o espaço amostral de uma variável aleatória discreta, então $f : \Omega \to \mathbb{R}$ é sua função massa de probabilidade, em que*

- $0 \leq f(x) \leq 1$
- $f(x) = P(X = x)$

De maneira análoga podemos definir uma distribuição de probabilidades para o caso em que a variável aleatória X é contínua, ou seja, quando os valores assumidos pela variável estão dentro de um intervalo real. Vamos abordar melhor essa situação no próximo exemplo.

Considere o comprimento de um tubo circular utilizado na fabricação de descargas de um determinado modelo de veículo, de maneira que uma análise anterior feita no equipamento que corta o tubo identificou que os tamanhos deles variam entre 60 e 65 centímetros. Suponha que escolhemos, ao acaso, um dos tubos fabricados em um determinado dia, e que, com o auxílio de um fita métrica, possamos medir o seu comprimento. Vamos denotar por X a variável aleatória que representa o comprimento da peça. É fato que, apesar da variável X poder assumir qualquer valor entre 60 e 65 centímetros, o equipamento que escolhemos para medi-lo pode não ser tão eficaz, fazendo facilmente com que um tubo com $62,2576$ cm seja medido por $62,26$ cm. Suponha então que dispomos de um instrumento capaz de medir os comprimentos dos tubos sem aproximações. Dessa maneira, é bastante coerente pensar que o comprimento dos tubos podem assumir qualquer valor real entre 60 e 65 centímetros. Nessa perpectiva, podemos pensar em como definir uma distribuição de probabilidades para essa variável contínua, sendo que, atribuir uma probabilidade a cada ponto do intervalo se faz impossível, uma vez que estamos tratando de infinitos números reais dentro do intervalo.

Observe que, na situação descrita acima, se atribuirmos a cada ponto do intervalo uma probabilidade maior que zero, a soma dessas probabilidades será igual a infinito, o que contradiz a Definição 1.4. De uma maneira geral, em casos como esse, atribui-se valores para a variável aleatória dentro de um intervalo ao invés de considerar um único valor na atribuição de probabilidades. Dessa forma, supondo que cada comprimento dentro do intervalo $[60; 65]$ tem igual probabilidade de ocorrer, se dividirmos o espaço amostral em cinco intervalos de comprimento 1, por exemplo, é bastante razoável que cada um desses intervalos tenha uma probabilidade de ocorrer de $\frac{1}{5}$, o que satisfaz claramente a Definição 1.4. Se dividirmos o intervalo $[60; 65]$ em n intervalos de amplitudes iguais, igualmente espaçados, e com probabilidades iguais a $\frac{1}{n}$, também continuaremos satisfazendo a Definição 1.4. Aumentando cada vez mais o número de intervalos, fazendo-os tender a infinito, uma distribuição de probabilidades contínua será definida pela área abaixo de uma função que chamaremos de densidade de probabilidade.

Definição 1.9. *Dizemos que $f(x)$ é uma função densidade de probabilidade (fdp) para uma variável aleatória continua X, se satisfaz duas condições:*

i) $f(x) \geq 0$, para todo $x \in (-\infty, +\infty)$;

ii) A área definida por $f(x)$ é igual a 1, ou seja, $\displaystyle\int_{-\infty}^{+\infty} f(x)\, dx = 1$.

Dada uma função $f(x)$ que satisfaça as condições da definição acima, então, para o caso em que $a \leq b$, temos

$$P(a \leq x \leq b) = \int_a^b f(x)\, dx.$$

Conforme Farias em [2], uma observação importante que resulta da interpretação geométrica de probabilidade como área sob a curva de densidade de probabilidade é que, sendo X uma variável aleatória contínua, então a probabilidade do evento $\{X = a\}$ é igual a zero. Isso por que o evento $\{X = a\}$ corresponde a um segmento de reta, o qual tem área nula.

Definição 1.10. *A funcão de distribuição acumulada (fda) para uma variável aleatória contínua X é definida por*

$$F(x) = P(X \leq x), \qquad x \in \mathbb{R}.$$

Proposição 1.11. *Se X é uma variável aleatória, sua função de distribuição acumulada $F(x)$ satisfaz as seguintes propriedades:*

i) $0 \leq F(x) \leq 1$;

ii) $\lim_{x \to +\infty} F(x) = 1$;

iii) $\lim_{x \to -\infty} F(x) = 0$;

iv) $x < y \Rightarrow F(x) \leq F(y)$;

v) $a < b \Rightarrow F(b) - F(a) = P(a \leq x \leq b)$.

Existe uma estreita relação entre a função densidade de probabilidade e a função de distribuição acumulada. Geometricamente, para qualquer real x, a área sob a curva da fdp entre $-\infty$ e x, é exatamente igual ao valor da função fda no ponto x. Utilizando a linguagem do cálculo, podemos dizer que essa relação é resultante do Teorema Fundamental do Cálculo (TFC), onde, por definição

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du.$$

Portanto, pelo TFC

$$f(x) = \frac{d}{dx}F(x). \tag{1.1}$$

Dessa forma, é importante ressaltar que, para especificar a função de distribuição acumulada de uma variável aleatória contínua é suficiente conhecer sua função de densidade, e vice-versa.

Definição 1.12. *A função quantílica de uma variável aleatória contínua X é definida por*

$$Q(x) = F^{-1}(x), \tag{1.2}$$

onde F é sua função de distribuição acumulada.

Definição 1.13. *Dada uma variável aleatória contínua X, definimos a esperança de X, ou média de X, ou ainda o valor esperado de X, por*

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx.$$

De uma maneira sucinta, podemos dizer que o valor esperado de uma variável aleatória contínua nada mais é do que a média que se espera obter de um experimento que é repetido várias vezes. A esperança de uma variável aleatória X, também denomindada de a média de X, é um valor único, representativo dos valores de X, e é conhecida em estatística como uma medida de tendência central, ou seja, uma medida que tende a estar no meio dos valores observados.Vejamos um exemplo.

Exemplo 1.14. *Seja a função densidade de probabilidade de uma variável aleatória contínua X dada por*

$$f(x) = \begin{cases} \frac{1}{2}x, & \textit{para } 0 < x < 2; \\ 0, & \textit{caso contrário.} \end{cases}$$

Então,

$$E[X] = \int_0^2 x \cdot \frac{1}{2}x\,dx = \int_0^2 \frac{1}{2}x^2\,dx = \frac{1}{6}x^3\Big|_0^2 = \frac{8}{6} - \frac{0}{6} = \frac{4}{3}.$$

Proposição 1.15. *Seja X uma variável aleatória e k uma constante, tem-se*

(i) $E[k] = k$;

(ii) $E[X \pm k] = E[X] \pm k$;

(iii) $E[k \cdot X] = k \cdot E[X]$;

(iv) *Se $X_1, X_2, \ldots, X_n$ são n variáveis aleatórias tais que $E[X_i]$ existe para $i = 1, 2, \ldots, n$, então*

$$E\left[X_1 + X_2 + \ldots + X_n\right] = E[X_1] + E[X_2] + \ldots + E[X_n].$$

Demonstração. Utilizando propriedades de integral e a Definição 1.9, temos que

(i)

$$E[k] = \int_{-\infty}^{+\infty} kf(x)\,dx = k\int_{-\infty}^{+\infty} f(x)\,dx = k.$$

(ii)

$$\begin{aligned} E[X \pm k] &= \int_{-\infty}^{+\infty} (x \pm k)f(x)\,dx \\ &= \int_{-\infty}^{+\infty} xf(x)\,dx \pm \int_{-\infty}^{+\infty} kf(x)\,dx \\ &= E[X] \pm k\int_{-\infty}^{+\infty} f(x)\,dx \\ &= E[X] \pm k. \end{aligned}$$

(iii)

$$E[k \cdot X] = \int_{-\infty}^{+\infty} k\,xf(x)\,dx = k \cdot \int_{-\infty}^{+\infty} xf(x)\,dx = k \cdot E[X].$$

(iv)

$$\begin{aligned} E\left[X_1 + X_2 + \ldots + X_n\right] &= \int_{-\infty}^{+\infty} \left(x_1 + x_2 + \ldots + x_n\right) f(x)\,dx \\ &= \int_{-\infty}^{+\infty} \left[x_1 f(x) + x_2 f(x) + \ldots + x_n f(x)\right]\,dx \\ &= \int_{-\infty}^{+\infty} x_1 f(x)\,dx + \int_{-\infty}^{+\infty} x_2 f(x)\,dx + \ldots + \int_{-\infty}^{+\infty} x_n f(x)\,dx \\ &= E[X_1] + E[X_2] + \ldots + E[X_n]. \end{aligned}$$

□

Embora a média ou esperança de uma variável aleatória X seja a mais usada, outras duas medidas importantes em estatística, conhecidas também como medidas de tendência central, são a Mediana e a Moda.

A *Mediana*(Md) é o valor que "divide"o conjunto de dados em dois grupos satisfazendo a seguinte propriedade

$$P(x \geq \text{Md}) \geq 0,5 \text{ e } P(x \leq \text{Md}) \geq 0,5.$$

Para o caso de uma distribuição contínua, a Mediana corresponde a uma ordenada que separa a curva de densidade em duas partes, cada uma delas com área igual a $\frac{1}{2}$.

A *Moda*(Mo) é o valor máximo registrado, ou seja, o valor que tem maior probabilidade de ocorrência. Em linguagem matemática temos que $f(\text{Mo}) = \textit{máx}\, f(x)$. Pode acontecer de dois, três ou mais valores terem probabilidades máximas de ocorrer em determinado evento. Nesses casos, dizemos que a distribuição é bimodal, trimodal ou multimodal, respectivamente.

As medidas de tendência central ou de posição (ou ainda de localização) nos omitem as informações sobre a homogeneidade do conjunto de dados. Essas medidas tem por objetivo trazer um valor que represente todo o grupo sem, no entanto, levar em consideração o quanto esses dados estão ou não próximos um do outro ou da própria média. As medidas que nos dão uma noção de o quanto o nosso grupo de dados se distancia da média, ou seja, o quanto o conjunto de dados é homogêneo, são chamadas medidas de dispersão. São elas: Variância (σ^2) e Desvio Padrão (σ).

Para entendermos melhor como essas medidas são importantes, vamos supor que a variável aleatória contínua X representa o tempo de duração de pilhas de relógios que estão sendo entregues por um fabricante e que $E(X) = 2000$ horas. Pode ser que a maioria dessas pilhas tenha um tempo de vida compreendido entre 1950 e 2050 horas. Mas pode ser também que essas pilhas tenham sido fabricadas em lotes diferentes e que a qualidade em um desses lotes tenha sido superior a do outro, de maneira que metade destas pilhas tenha duração de 1200 horas e a outra metade duração de 2800 horas. Nesse sentido é que há a necessidade de uma medida que traduza essa disparidade entre os dados da variável.

Definição 1.16. *Se X é uma variável aleatória contínua, $f(x)$ é sua função densidade de probabilidade e fazendo $\mu = E(X)$, então a variância de X é definida por*

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx.$$

Definição 1.17. *O desvio padrão de uma variável aleatória contínua X é dado por*

$$\sigma = \sqrt{\int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx}.$$

Uma propriedade importante será mostrada a seguir. Ela diz que a variância pode ser calculada como a diferença entre a esperança do quadrado de X pelo quadrado da esperança de X.

Proposição 1.18. *A variância de uma variável aleatória contínua X é dada por*

$$\sigma^2 = E(X^2) - [E(X)]^2.$$

Demonstração.

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{+\infty} (x^2 - 2\mu x + \mu^2) f(x)\,dx.$$

Aplicando propriedades da integral, temos

$$\int_{-\infty}^{+\infty} (x^2 - 2\mu x + \mu^2) f(x)\,dx = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - 2\mu \int_{-\infty}^{+\infty} x f(x)\,dx + \mu^2 \int_{-\infty}^{+\infty} f(x)\,dx.$$

Logo, usando a Definição 1.13, temos

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{+\infty} x^2 f(x)\,dx - 2\mu \int_{-\infty}^{+\infty} x f(x)\,dx + \mu^2 \int_{-\infty}^{+\infty} f(x)\,dx \\
&= E(X^2) - 2\mu \cdot \mu + \mu^2 \cdot 1 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2.
\end{aligned}$$

Portanto

$$\sigma^2 = E(X^2) - [E(X)]^2. \tag{1.3}$$

□

1.3 Função de Sobrevivência

O tempo entre uma observação inicial e a ocorrência de um evento (ou falha) é chamado de *sobrevivência*. Chamaremos de função de sobrevivência $S(x)$ a probabilidade de que o indivíduo ou componente possua tempo de vida maior do que x, ou seja, a probabilidade de que, após passado um tempo x, ele tenha sobrevivido ao evento de interesse. Importante ressaltar que a sobrevivência tem um amplo campo de aplicações, não só na área da saúde ou indústria, como também na educação, comércio entre outras. Podemos utilizá-la por exemplo para estudar o tempo entre a soltura de um preso e sua reincidência no mundo do crime. Como vimos na Definição 1.10, a função de distribuição acumulada é dada por

$F(x) = P(X \leq x)$, ou seja, ela é a probabilidade de que o evento aconteça até o tempo x, o que nos leva a concluir que

$$S(x) = P(X > x) = 1 - P(X \leq x) = 1 - F(x). \tag{1.4}$$

Importante lembrar que o tempo de sobrevivência é uma variável aleatória positiva e, em geral, contínua

1.4 Função Taxa de Falha

A função taxa de falha $h(x)$ é definida como a probabilidade de que um evento de interesse aconteça em um intervalo curto de tempo, dado que ele não ocorreu até o início do intervalo.

Definição 1.19. *A taxa de falha no intervalo $[x, x + \Delta x]$ é definida como a probabilidade de que a falha ocorra neste intervalo, dado que não ocorreu antes de x, dividido pelo comprimento do intervalo.*

Teorema 1.20. *A função taxa de falha $h(x)$ de uma variável aleatória continua X é dada por*

$$h(x) = \frac{f(x)}{S(x)}. \tag{1.5}$$

Demonstração. A probabilidade de que uma falha ocorra em um intervalo de tempo $[x, x + \Delta x]$ pode ser expressa em termos da função sobrevivência como

$$\begin{aligned} P[x \leq X \leq x + \Delta x] &= P[X \leq x + \Delta x] - P[X \leq x] \\ &= [1 - P(X > x + \Delta x)] - [1 - P(X > x)] \\ &= P(X > x) - P(X > x + \Delta x) \\ &= S(x) - S(x + \Delta x). \end{aligned}$$

Utilizando a definição de probabilidade condicional, encontrada em [11], página 158, e a Definição 1.19, temos

$$\frac{P[x \leq X \leq x + \Delta x \mid X > x]}{x + \Delta x - x} = \frac{P[x \leq X \leq x + \Delta x]}{(x + \Delta x - x)\, P[X > x]} = \frac{S(x) - S(x + \Delta x)}{(\Delta x)\, S(x)}.$$

Assim, a taxa instantânea de falha em um tempo x será dada por

$$h(x) = \lim_{\Delta x \to 0^+} \frac{P[x \leq X \leq x + \Delta x \mid X > x]}{\Delta x} = \lim_{\Delta x \to 0^+} \frac{S(x) - S(x + \Delta x)}{(\Delta x)\, S(x)}$$

quando Δx tender a zero. Usando a Equação (1.4), temos que

$$h(x) = \lim_{\Delta x \to 0^+} \frac{1 - F(x) - (1 - F(x + \Delta x))}{\Delta x\, S(x)} = \frac{1}{S(x)} \lim_{\Delta x \to 0^+} \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

Pela definição de derivada, o limite na equação acima é a derivada da função de distribuição acumulada no ponto x. Portanto, usando (1.1), obtemos

$$h(x) = \frac{f(x)}{S(x)}.$$

□

1.5 Momentos

De suma importância em estatística, os momentos caracterizam as distribuições de probabilidade. Como veremos, cada momento fornece uma informação diferente sobre os dados em estudo. Estes momentos diferem pela ordem, podendo ser de 1ª, 2ª, 3ª, etc. A seguir, definimos momentos de ordem n.

Definição 1.21. *Para toda variável aleatória contínua X e qualquer inteiro positivo n, a esperança $E(X^n)$ é denominado n-ésimo momento de X, ou momento de ordem n. Portanto, usando a Definição 1.13, temos que*

$$\mu_n^{'} = E(X^n) = \int_{-\infty}^{+\infty} x^n f(x)\, dx, \tag{1.6}$$

onde $f(x)$ é uma função densidade de probabilidade.

O momento de 1ª ordem será dado por

$$\mu_1^{'} = E(X) = \int_{-\infty}^{+\infty} x f(x) \mathrm{dx},$$

ou seja, o primeiro momento mede o valor médio dos dados.

1.6 Função Geratriz de Momentos

A função que definiremos a seguir recebe o nome de função geratriz de momentos, pois utilizando-a, podemos encontrar, caso exista, qualquer momento da variável aleatória X.

Definição 1.22. *Seja X uma variável aleatória contínua. A função geratriz de momentos $M_x(t)$, com $t \in \mathbb{R}$ é definida como*

$$M_X(t) = E\left[\mathrm{e}^{tX}\right]$$

onde, pela Definição 1.13, obtemos

$$M_X(t) = \int_{-\infty}^{+\infty} \mathrm{e}^{tX} f(x)\, dx \tag{1.7}$$

Para calcularmos o n-ésimo momento de uma variável aleatória X, é necessário que a função esteja definida em uma vizinhança do ponto zero, pois os momentos serão obtidos através de sucessivas diferenciações no ponto zero, ou seja, a partir da função geratriz de momentos de uma variável aleatória X, podemos encontrar o n-ésimo momento calculando, caso exista, a n-ésima derivada da função no ponto $t = 0$.

De fato, seja $M_X^n(t)$ a n-ésima derivada de $M_X(t)$,

$$M_X^n(t) = \left[\frac{\mathrm{d^n}}{\mathrm{dt^n}} E\left(\mathrm{e^{tX}}\right)\right].$$

Utilizando propriedades de Cálculo e a Definição 1.13, a expressão acima reduz-se a

$$\begin{aligned}
M_X^n(t) &= \left[\frac{\mathrm{d^n}}{\mathrm{dt^n}} \int_{-\infty}^{+\infty} \mathrm{e}^{tX} f(x) \mathrm{d}x\right] \\
&= \int_{-\infty}^{+\infty} \frac{\mathrm{d^n}}{\mathrm{dt^n}} \mathrm{e}^{tX} f(x) \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} X^n \mathrm{e}^{tX} f(x) \mathrm{d}x \\
&= E\left[X^n \cdot e^{tX}\right].
\end{aligned}$$

Fazendo $t = 0$ na equação acima, obtemos

$$M_X^n(0) = E\left[X^n \cdot \mathrm{e}^{0 \cdot X}\right] = E\left[X^n\right],$$

e portanto

$$\begin{aligned}
&M_X^1(0) = E\left[X\right]; \\
&M_X^2(0) = E\left[X^2\right]; \\
&M_X^3(0) = E\left[X^3\right]; \\
&... \\
&M_X^n(0) = E\left[X^n\right],
\end{aligned}$$

caso existam as derivadas.

Uma outra maneira de deduzir os n-ésimos momentos de uma variável aleatória X, é aplicar a Definição 1.22.

Usando a expansão em série da função exponencial e^{tX}, temos

$$e^{\mathrm{tX}} = 1 + \mathrm{tX} + \frac{(\mathrm{tX})^2}{2!} + \frac{(\mathrm{tX})^3}{3!} + \ldots \frac{(\mathrm{tX})^{\mathrm{n}}}{\mathrm{n}!} + \ldots = \sum_{\mathrm{n}=0}^{\infty} \frac{\mathrm{t^n X^n}}{\mathrm{n}!}.$$

Aplicando a esperança em ambos os lados da igualdade acima, e usando a Definição 1.22, chegamos a

$$\begin{aligned}
M_X(t) = E\left[\mathrm{e}^{\mathrm{tX}}\right] &= E\left[\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}\right] \\
&= E\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \ldots \frac{(tX)^n}{n!} + \ldots\right] \\
&= 1 + E\left[tX\right] + E\left[\frac{(tX)^2}{2!}\right] + E\left[\frac{(tX)^3}{3!}\right] + \ldots + E\left[\frac{(tX)^n}{n!}\right] + \ldots \\
&= 1 + tE\left[X\right] + t^2\frac{E[X^2]}{2!} + t^3\frac{E[X^3]}{3!} + \ldots + t^n\frac{E[X^n]}{n!} + \ldots
\end{aligned}$$

Como $M_X(t)$ é uma função de variável t, podemos tomar a derivada de $M_X(t)$ em relação a t, supondo que o lado direito possa ser escrito como a soma infinita das respectivas derivadas.

$$\begin{aligned}
M_X^1(t) &= 0 + E\left[X\right] + 2t\frac{E[X^2]}{2!} + 3t^2\frac{E[X^3]}{3\cdot 2!} + \ldots + nt^{n-1}\frac{E[X^n]}{n\cdot(n-1)!} + \ldots \\
&= E\left[X\right] + tE[X^2] + t^2\frac{E[X^3]}{2!} + \ldots + t^{n-1}\frac{E[X^n]}{(n-1)!} + \ldots
\end{aligned}$$

Fazendo $t = 0$, temos

$$M_X^1(0) = E\left[X\right] + 0E[X^2] + 0^2\frac{E[X^3]}{2!} + \ldots + 0^{n-1}\frac{E[X^n]}{(n-1)!} + \ldots,$$

portanto

$$M_X^1(0) = E\left[X\right].$$

Derivando mais uma vez a função geratriz de momentos, temos

$$M_X^2(t) = 0 + E[X^2] + 2t\frac{E[X^3]}{2!} + \ldots + (n-1)t^{n-2}\frac{E[X^n]}{(n-1)!} + \ldots,$$

portanto

$$M_X^2(t) = E[X^2] + tE[X^3] + \ldots + t^{n-2}E[X^n] + \ldots.$$

Fazendo $t = 0$,

$$M_X^2(0) = E[X^2] + 0E[X^3] + \ldots + 0^{n-2}E[X^n] + \ldots,$$

portanto

$$M_X^2(0) = E[X^2].$$

Seguindo com as derivações e raciocínio análogo, obtemos

$$M_X^3(0) = E[X^3], \;\; M_X^4(0) = E[X^4], \;\; \ldots \;\; M_X^n(0) = E[X^n].$$

1.7 Entropia de Rényi

A entropia é a medida de variação ou incerteza de uma variável aleatória X. A ideia associada a entropia é a de que, quanto mais incerto é o resultado de um experimento aleatório, maior será a informação obtida ao observar a sua ocorrência. Uma medida de entropia muito conhecida é a entropia de Rényi(1921-1970). Alfréd Rényi foi um matemático húngaro que trabalhou nas áreas de combinatória, teoria dos grafos, teoria dos números e teoria das probabilidades.

A entropia de Rényi de uma variável aleatória X com função densidade de probabilidade dada por $f(x)$, é definida como

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \left(\int_0^\infty f^\gamma(x)\, dx \right), \tag{1.8}$$

com $\gamma > 0$ e $\gamma \neq 1$.

Capítulo 2

Distribuição Weibull e a Classe Zeta-G

Como dito anteriormente, o presente trabalho propõe uma nova distribuição contínua, composição da distribuição discreta Zeta com a distribuição contínua Weibull. Distribuições generalizadas têm sido amplamente estudadas ao longo dos últimos anos, buscando métodos para modelagem do tempo de duração de componentes ou tempo de vida de indivíduos. Esse fato é justificado, muitas vezes, em função das distribuições existentes não se ajustarem de modo satisfatório ao conjunto de dados reais em estudo. Não se encontra na literatura composições de distribuições Zeta, porém, com a nova classe Zeta-G proposta por Paixão(2014), abre-se a possibilidade de estudos de novas distribuições geradas a partir da Zeta, que origina-se da função Zeta de Riemann.

Georg Friedrich Bernhard Riemann, nasceu em 1826, numa aldeia de Hanover, Alemanha. Filho de um pastor luterano, estudou na Universidade de Berlim e depois na de Gottingen, obtendo seu doutorado nesta última instituição com uma brilhante tese no campo da teoria das funções complexas. Nessa tese, encontram-se as chamadas *equações diferenciais de Cauchy-Riemann* que garantem a analiticidade de uma função de variável complexa, e o produtivo conceito de *superfície de Riemann*, que introduz considerações topológicas na análise. Riemann tornou claro o conceito de integrabilidade pela definição do que chamamos agora *integral de Riemann*[5], página 614. Riemann faleceu em 1866, no norte da Itália, vítima da tuberculose.

Definição 2.1. *A Função Zeta de Riemann é definida pela soma*

$$\zeta(s) = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \dots,$$

ou seja,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{2.1}$$

onde $s > 1$ é um número complexo da forma $s = a + bi$, com $a, b \in \mathbb{R}$ e i é a unidade imaginária dos números complexos tal que $i^2 = -1$.

Por volta de 1859, Riemann conjecturou que todos os zeros da função Zeta tem sua parte real $a = \frac{1}{2}$. Em 1914, Sir Godfrey Harold Hardy (1877-1947), um inglês especialista em teoria dos números, conseguiu provar que $\zeta(s)$ tem uma infinidade de zeros para $a = \frac{1}{2}$ [5], página 614. A Figura 2.1 mostra o gráfico da função Zeta de Riemann.

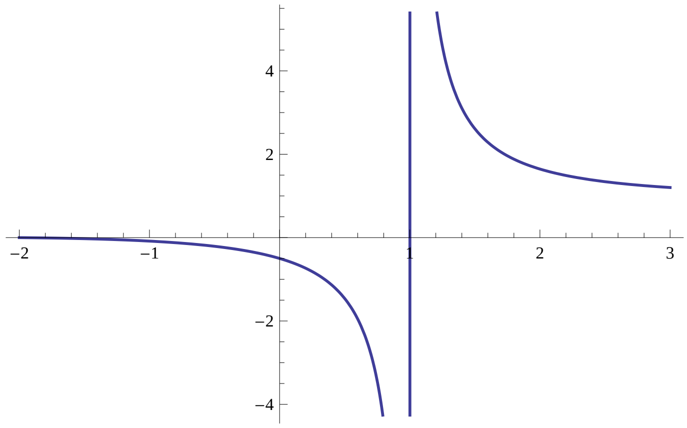

Figura 2.1: Gráfico da função Zeta de Riemann

Conforme [13] "a função Zeta de Riemann desempenha papel importante em várias áreas da pesquisa moderna. Na Física Teórica aparece em problemas de regularização de determinantes infinitos que surgem em Teoria de Campos; e, também em alguns trabalhos teóricos sobre o importante fenômeno da supercondutividade. Muito importante para o estudo dos números primos, a função Zeta ocupa amplo espaço em qualquer texto sobre a Teoria dos Números. Sua relação com outras funções especiais também lhe reserva uma posição importante na Teoria das Funções".

Uma distribuição Zeta de uma variável aleatória discreta será representada a seguir por sua função massa de probabilidade, ou simplesmente função de probabilidade.

Definição 2.2. *Dizemos que uma variável aleatória discreta X tem uma distribuição Zeta com parâmetro α, se sua função massa de probabilidade é dada por*

$$P(k) = \frac{k^{-\alpha}}{\zeta(s)} \quad k=1,2,\ldots,$$

onde $\zeta(s)$ é dada na Equação (2.1), e $\alpha > 1$.

A distribuição Zeta deve seu nome à função Zeta de Riemann, porém, foi George Kingsley Zipf (1902-1950), quem a popularizou. Conforme [18], embora a distribuição Zeta tenha sido primeiramente usada por Vilfredo Pareto (1848-1923), cientísta político, sociólogo e economista italiano, para descrever a distribuição dos rendimentos das famílias de um determinado país, foi Zipf quem aplicou a distribuição Zeta em uma ampla variedade de problemas em diferentes áreas.

A função de probabilidade Zeta ou Zipf é um exemplo de uma distribuição de cauda pesada cuja importância cresceu bastante desde meados dos anos 1990. Distribuições de cauda pesada são caracterizadas por ter uma probabilidade de ocorrência muito baixa no fim da cauda, ou seja, uma distribuição onde os dados são classificados de forma decrescente. As aplicações desta função de probabilidade incluem: número de consumidores afetados por um *blackout*, tamanhos de arquivos solicitados em transferência via Web e atraso de pacotes na internet.

2.1 Distribuição de Weibull

Ernst Hjalmar Waloddi Weibull (1887-1979) ficou famoso pelo seu brilhante trabalho na área de fadiga de materiais, sendo seu modelo estatístico um dos mais populares no mundo. Seu modelo também é bastante utilizado em áreas de saúde para obtenção de respostas sobre dados de vida e em previsões do tempo por ter representações gráficas bastante simples.

Definição 2.3. *A função de distribuição acumulada da distribuição Weibull é dada por*

$$G(x) = 1 - e^{-\beta x^{\alpha}}, \tag{2.2}$$

onde $\alpha > 0$ é o parâmetro de forma e $\beta > 0$ é o parâmetro de escala.

Um parâmetro de forma indica a forma da curva e, consequentemente, a característica das falhas. O parâmetro de escala define como seus dados estão propagados. Um valor maior da escala amplia a sua distribuição, enquanto um valor de escala menor reduz a sua distribuição de dados.

Definição 2.4. *A função densidade de probabilidade da distribuição Weibull é dada por*

$$g(x) = \alpha\beta x^{\alpha-1}\, e^{-\beta x^{\alpha}}. \tag{2.3}$$

De fato, derivando (2.2) temos

$$g(x) = \frac{d}{dx}G(x) = \frac{d}{dx}(1 - e^{-\beta x^{\alpha}}) = 0 - e^{-\beta x^{\alpha}}(-\alpha\beta x^{\alpha-1}) = \alpha\beta x^{\alpha-1}\, e^{-\beta x^{\alpha}}.$$

De acordo com o levantamento feito por Santos(2012) em [16], muitas classes de distribuições baseadas na distribuição Weibull foram propostas nos últimos anos, tais como as distribuições Weibull exponencializada, Weibull estendida, Weibull modificada, Weibull estendida flexível, beta Weibull, Weibull modificada generalizada, entre outras. A distribuição Weibull é muito utilizada para a modelagem de dados de confiabilidade, podendo modelar dados assimétricos ou simétricos. Dizemos que uma distribuição é assimétrica quando ela não se apresenta simétrica em relação ao seu máximo, ou seja, quando uma das caudas é mais longa que a outra. Se a cauda for mais longa à direita, dizemos que a distribuição é assimétrica à direita ou positiva. Caso contrário, dizemos que a distribuição é assimétrica à esqueda ou negativa.

Uma outra característica da distribuição Weibull diz respeito a sua função de falha. A curva da função de falha da distribuição Weibull não apresenta a forma de banheira, o que a torna deficiente para caracterizar alguns conjuntos de dados. Na esperança de obtermos a forma de banheira é que a compomos com outras distribuições. As curvas da função de falha da distribuição Weibull podem ser crescente, decrescente ou constante.

2.2 Classe de Distribuições Zeta-G

Paixão[15] propõe a nova família de distribuições Zeta-G com um parâmetro adicional de forma. Assim, podemos dizer que a família de distribuições Zeta-G pode gerar novas distribuições que, segundo Paixão, têm o papel de governar a assimetria e gerar densidades com caudas mais pesadas ou mais leves, além de serem úteis em fornecer uma distribuição mais flexível para modelar a função taxa de falha.

Nesta seção, definiremos as principais funções da distribuição Zeta-G proposta por Paixão, entre elas, a função densidade de probablilidade, a função de distribuição acumulada, a função quantílica e a função geratriz de momentos.

Definição 2.5. *A função densidade de probabilidade da família Zeta-G é dada por*

$$f(x) = \frac{Li_{s-1}[1-G(x)]g(x)}{\zeta(s)[1-G(x)]}, \tag{2.4}$$

em que $g(x) = \frac{d}{dx}G(x)$ *e* $L_i(x)$ *é a função polilogarítmica definida pela série de potências*

$$Li_s(z) = \sum_{k=1}^{+\infty} \frac{z^k}{k^s}, \tag{2.5}$$

com $|z| < 1$ *e* $\zeta(s)$ *é a função Zeta de Riemann. Se uma variável aleatória* X *tem uma densidade (2.4), escrevemos* $X \sim Zeta-G(\tau, s)$*, onde* τ *é o vetor de parâmetros associado à distribuição base* $G(x)$*.*

Definição 2.6. *A função de distribuição acumulada da família Zeta-G é dada por*

$$F(x) = \frac{\zeta(s) - Li_s[1-G(x)]}{\zeta(s)} \tag{2.6}$$

Definição 2.7. *A função quantílica da clase Zeta-G dada por* $q = Q(w, s) = F^{-1}(w, s)$, *onde F é a função de distribuição acumulada, é definida como*

$$q = Q(w, s) = G^{-1}\left\{ \sum_{n=1}^{k} \frac{(-1)^n f_{n-1,n}}{n c_1^n} (w - w_0)^n \right\}, \tag{2.7}$$

em que $c_j = \sum_{k=1}^{\infty} \frac{(-1)^j \binom{k}{j}}{\zeta(s) k^s}$ *$(j \geq 0)$ e os coeficientes $f_{j,n}$ são dados recursivamente por*

$$f_{j,n} = j^{-1} \sum_{m=1}^{j} \big[m(n+1) - j\big] d_m f_{j-m,n},$$

com $d_0 = 1$ e $d_j = -c_1^{-1} \sum_{i=1}^{j} d_{j-i} c_{i+1}$ *$(j \geq 1)$.*

Paixão também deriva uma expressão de forma fechada para a entropia de Rényi, quando X é uma variável aleatória para a classe Zeta-G. Neste caso, a entropia de Rényi reduz-se a

$$\begin{aligned} I_R(\gamma) &= \frac{\gamma}{\gamma - 1} \log[\zeta(s)] + \frac{1}{1-\gamma} \log\left\{ \sum_{j,i,k=0}^{\infty} \sum_{v=0}^{k+i} \frac{(-1)^{k+i} g_i \, e_{i,k} \binom{k+i}{v} \Gamma(\gamma + j)}{\Gamma(\gamma) j!} \right. \\ &\times \left. \int_0^{\infty} g(x)^{\gamma} G(x)^{j+v} \, dx \right\}, \end{aligned} \tag{2.8}$$

com $g_i = \sum_{t=i}^{\infty} \frac{(-1)^{t-i} (\gamma)_t}{t!} \binom{t}{i}$, onde $(\gamma)_t = \gamma(\gamma-1)\ldots(\gamma-t+1)$ é o fatorial descendente, e $e_{i,k} = (k \, a_0^{-1} \sum_{m=1}^{k} [m(i+1) - k] a_{m \, e_{i,k-m}})$ com $k \geq 1$, $e_{i,0} = a_0^i$ e $a_k = (k+1)^{1-s}$. O farorial descendente de γ a t é definido como $(\gamma)_t = \frac{\gamma!}{(\gamma - t)!} = \prod_{i=\gamma-t+1}^{\gamma} i$.

Capítulo 3

A distribuição Zeta-Weibull

Neste Capítulo, apresentaremos a distribuição Zeta-Weibull, discutindo algumas de suas propriedades e determinando a função densidade de probabilidade, a função de distribuição acumulada, expansões para a fdp e fda, função quantílica, momentos, função geratriz de momentos, entropia de Rényi, confiabilidade e estimação. Apresentaremos também uma aplicação da distribuição Zeta-Weibull ao conjunto de dados reais obtido em [17].

Seja X uma variável aleatória contínua com uma densidade Zeta-Weibull. Então, escrevemos $X \sim ZW(s, \alpha, \beta)$, em que s denota o parâmetro da distribuição Zeta e α, β são os parâmetros da distribuição Weibull.

Teorema 3.1. *Se $X \sim ZW(s, \alpha, \beta)$, então a função densidade de probabilidade da distribuição Zeta-Weibull é dada por*

$$f(x) = \frac{\alpha\beta x^{\alpha-1} Li_{s-1}[\, e^{-\beta x^{\alpha}}]}{\zeta(s)}, \tag{3.1}$$

com $\alpha > 0$, $\beta > 0$, $s > 1$ e $x \in [0, +\infty)$.

Demonstração. De fato, substituindo as Equações (2.2) e (2.3) em (2.4), temos

$$f(x) \;=\; \frac{Li_{s-1}[1 - (1 -\, e^{-\beta x^{\alpha}})]\, \alpha\beta x^{\alpha-1}\, e^{-\beta x^{\alpha}}}{\zeta(s)[1 - (1 -\, e^{-\beta x^{\alpha}})]}. \tag{3.2}$$

Simplificando,

$$\begin{aligned} f(x) &= \frac{Li_{s-1}[\, e^{-\beta x^{\alpha}}]\, \alpha\beta x^{\alpha-1}\, e^{-\beta x^{\alpha}}}{\zeta(s)\, e^{-\beta x^{\alpha}}} \\ &= \frac{\alpha\beta x^{\alpha-1} Li_{s-1}[\, e^{-\beta x^{\alpha}}]}{\zeta(s)}. \end{aligned}$$

□

Usando o software **Mathematica** (versão 9.0.1.0), verificamos que a função densidade da distribuição Zeta-Weibull dada acima, satisfaz a Definição1.9.

Teorema 3.2. *Se $X \sim ZW(s, \alpha, \beta)$, então a função de distribuição acumulada da distribuição Zeta-Weibull é dada por*

$$F(x) = \frac{\zeta(s) - Li_s[e^{-\beta x^\alpha}]}{\zeta(s)}. \tag{3.3}$$

Demonstração. De fato, substituindo (2.2) e (2.3) na Equação (2.6), temos

$$\begin{aligned} F(x) &= \frac{\zeta(s) - Li_s[1 - (1 - e^{-\beta x^\alpha})]}{\zeta(s)} \\ &= \frac{\zeta(s) - Li_s[e^{-\beta x^\alpha}]}{\zeta(s)}. \end{aligned}$$

□

Os gráficos da função densidade de probabilidade da distribuição Zeta-Weibull que serão apresentados aqui foram construídos no software **Mathematica**(versão 9.0.1.0). Um gráfico de distribuição de probabilidade pode ser utilizado para visualizar e comparar as formas das curvas de distribuição, podendo, entre outas coisas, determinar como a mudança de um valor de parâmetro pode afetar a curva. O gráfico na Figura 3.1 representa a função densidade para alguns valores do parâmetro de forma α, de modo que foram fixados os valores $\beta = 0,8$ e $s = 3$.

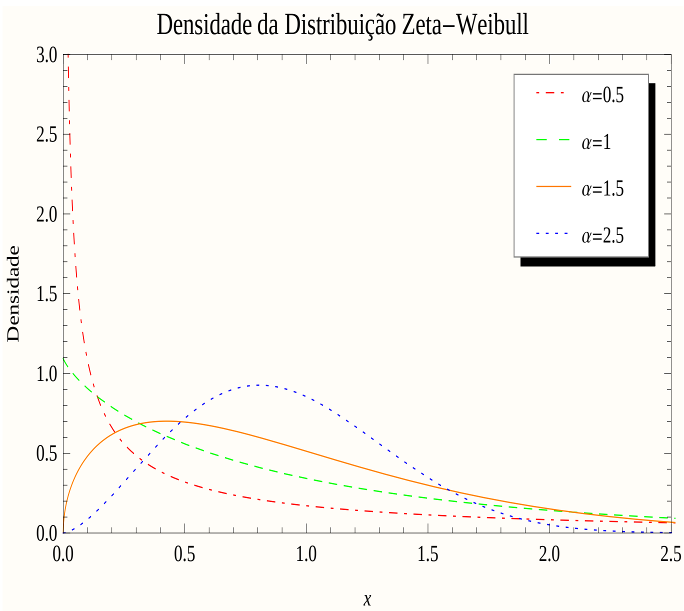

Figura 3.1: fdp Zeta-Weibull variando o parâmetro α

O parâmetro α é adimensional e como o próprio nome sugere, tal parâmetro interfere no formato da função densidade de probabilidade. Quando $\alpha < 1$ a fdp apresenta frequências elevadas na parte inicial da vida, sendo estritamente decrescente. Quando $\alpha = 1$, a curva da fdp ainda é estritamente decrescente, entretanto, as frequências na parte inicial da vida são menos elevadas. Para valores de α maiores do que um, notamos que grande parte da densidade concentra-se ao redor de um determinado tempo de vida.

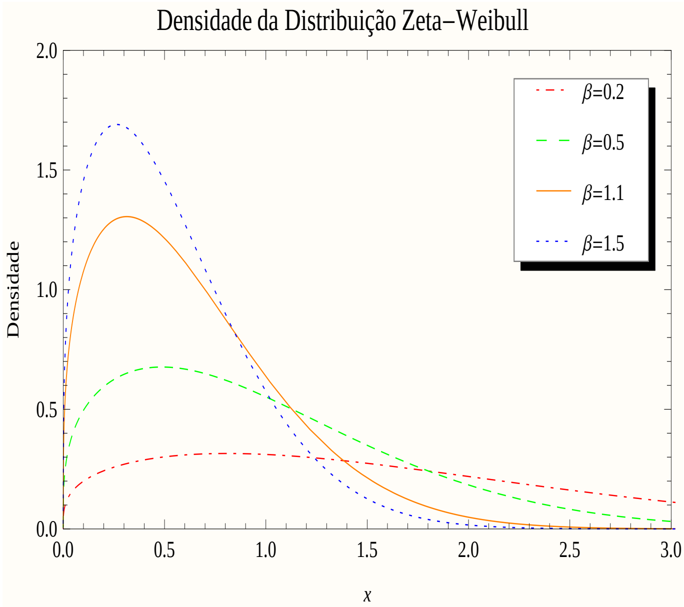

Figura 3.2: fdp Zeta-Weibull variando o parâmetro β

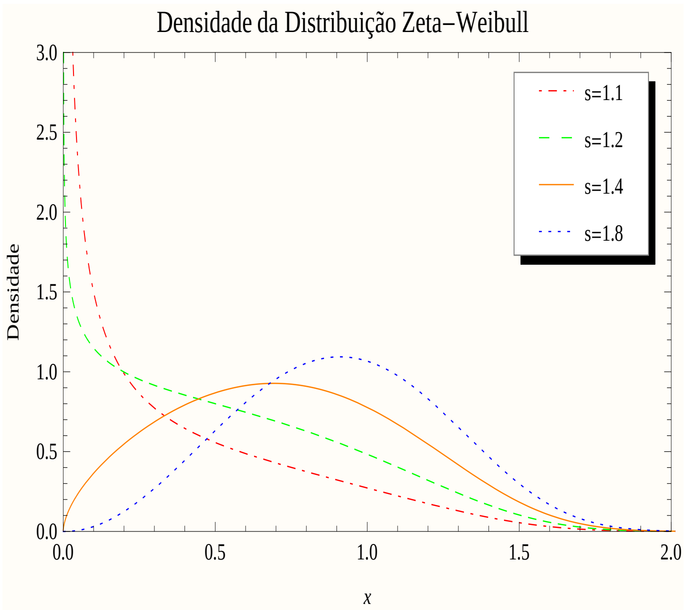

Figura 3.3: fdp Zeta-Weibull variando o parâmetro s

Na Figura 3.2 foram fixados os valores $\alpha = 1,8$ e $s = 4$, enquanto que o parâmetro de escala β varia. O parâmetro de escala β determina a dispersão (ou escala) da curva. Para o valor de α escolhido, a distribuição apresenta uma concentração ao redor de um determinado tempo de vida, e, neste caso, o parâmetro β estende ou estreita a curva de densidade. Para valores de β mais próximos de zero, a curva de densidade se estica para a direita, "diminuindo"sua curtose, que pode ser definida como o grau de "achatamento"de uma distribuição de frequências, medido em relação a uma distribuição normal, ou seja, o grau de concentração dos valores dessa distribuição em torno do seu centro. Já para valores de β cada vez maiores, a curva de densidade se estreita, "aumentando"sua curtose, ou seja, aumentando a probabilidade ao redor de um tempo de vida..

Na Figura 3.3 foram fixados os valores $\alpha = 4$ e $\beta = 0,5$, enquanto que o parâmetro de forma s varia. Por s ser também um parâmetro de forma, mais uma vez temos curvas com alta frequência na parte inicial da vida, casos em que s está se aproximando de 1 pela direita, e curvas com grande densidade ao redor de um determinado tempo de vida, quando s cresce.

3.1 Função de Sobrevivência e Função Taxa de Falha

No próximo Teorema determinamos a expressão para a função de sobrevivência da distribuição Zeta-Weibull.

Teorema 3.3. *Se $X \sim ZW(s, \alpha, \beta)$, então a função de sobrevivência da distribuição Zeta-Weibull é dada por*

$$S(x) = \frac{Li_s[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}{\zeta(s)}. \tag{3.4}$$

Demonstração. De fato, substituindo (3.3) em (1.4), temos

$$\begin{aligned} S(x) &= 1 - \frac{\zeta(s) - Li_s[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}{\zeta(s)} \\ &= \frac{\zeta(s) - \zeta(s) + Li_s[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}{\zeta(s)} \\ &= \frac{Li_s[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}{\zeta(s)}. \end{aligned}$$

□

O gráfico de uma função de sobrevivência exibe as probabilidades de sobrevivência para cada valor do tempo. Dessa forma, cada ponto no gráfico representa a proporção de sobreviventes em um tempo t e, como esse valor decresce a cada instante, temos que a função de sobrevivência é uma função não crescente no tempo, como pode ser observado nas Figuras 3.4, 3.5 e 3.6. O gráfico na Figura 3.4 têm dois parâmetros fixos, $s = 2$ e

$\beta = 0,5$, enquanto que o parâmetro de forma α varia. Na Figura 3.5 os parâmetros fixos são $s = 3$ e $\alpha = 1,3$, enquanto que o parâmetro de escala β varia. O gráfico da Figura 3.6 assume alguns valores para o parâmetro de forma s e têm parâmetros fixos $\alpha = 2$ e $\beta = 0,1$.

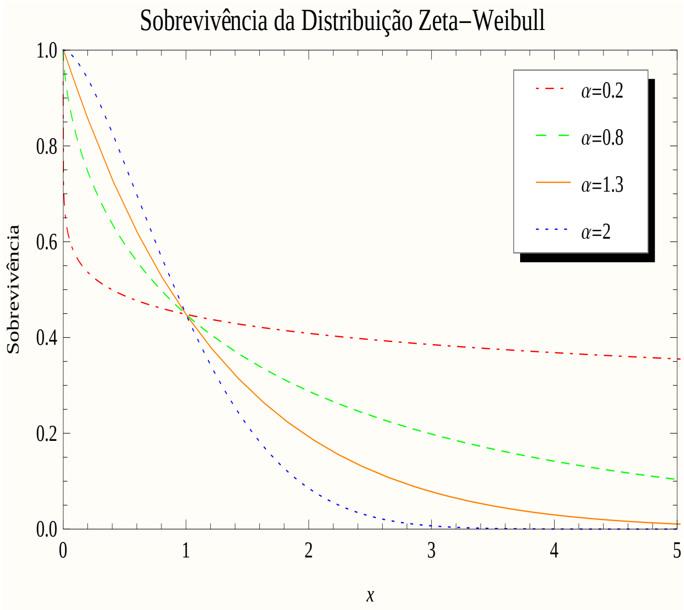

Figura 3.4: Sobrevivência Zeta-Weibull variando o parâmetro α

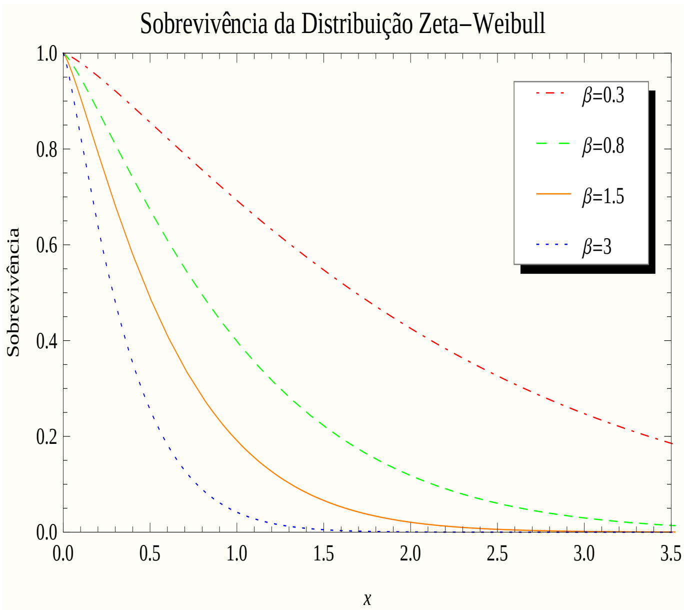

Figura 3.5: Sobrevivência Zeta-Weibull variando o parâmetro β

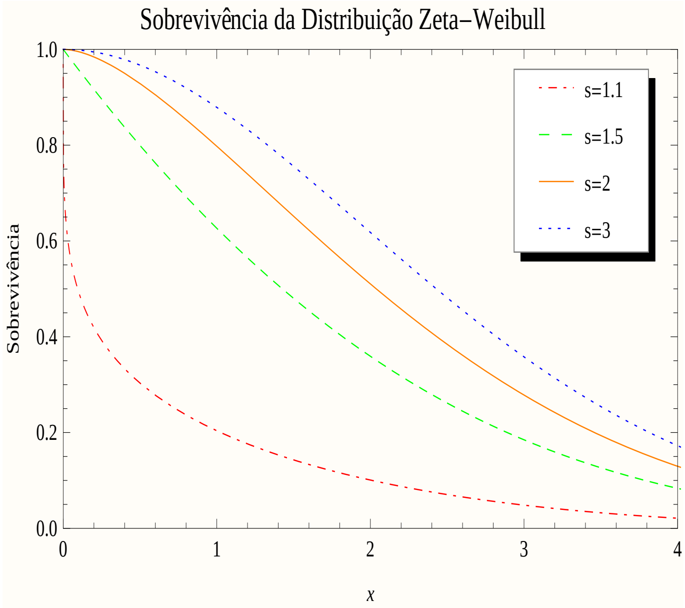

Figura 3.6: Sobrevivência Zeta-Weibull variando o parâmetro s

Teorema 3.4. *Se $X \sim ZW(s, \alpha, \beta)$, então a função taxa de falha da distribuição Zeta-Weibull é dada por*

$$h(x) = \frac{\alpha\beta x^{\alpha-1} Li_{s-1}[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}{Li_s[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}. \tag{3.5}$$

Demonstração. De fato, substituindo as Equações (3.1) e (3.4) em (1.5), temos

$$\begin{aligned} h(x) &= \frac{f(x)}{S(x)} \\ &= \frac{\dfrac{\alpha\beta x^{\alpha-1} Li_{s-1}[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}{\zeta(s)}}{\dfrac{Li_s[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}{\zeta(s)}} \\ &= \frac{\alpha\beta x^{\alpha-1} Li_{s-1}[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}{Li_s[\mathrm{e}^{-\beta \mathrm{x}^{\alpha}}]}. \end{aligned}$$

□

Por tratar da quantidade de risco associada a um evento em um determinado tempo, a função taxa de falha é uma das medidas mais importantes em análise de sobrevivência. Analisando os gráficos da função taxa de falha vemos duas situações:

(i) função crescente - quando a incidência de falha não decresce com o tempo.

(ii) função decrescente - quando a incidência de falha não cresce com o tempo.

Muitas situações apresentam função de risco dada pela combinação dessas categorias. Um modelo importante de gráfico de risco que existe são as curvas em forma de **banheira.**

Nessas curvas, a taxa de falha é frequentemente alta inicialmente, baixa no centro e novamente alta no final da vida. Observando uma curva em forma de **banheira** que descreve a vida útil de algum equipamento, podemos dividir sua vida operacional em três estágios:

(1) "Estágio de falha inicial" (ocorrência de falhas precoces),

(2) "Estágio de vida normal" (onde a incidência de falhas é relativamente estável no tempo), e

(3) "Estágio de desgaste" (quando o produto passa a apresentar desgaste acentuado e falhas passam a ocorrer com maior frequência).

O "estágio de falha inicial" aponta para o fato de que um equipamento tem usualmente, no início de sua vida útil, uma taxa elevada de falhas, devido a problemas de fabricação, componentes defeituosos, falha de montagem ou instalação inadequada. Com o passar do tempo essas falhas são corrigidas e os equipamentos entram em um patamar de estabilidade, ou seja, no "estágio de vida normal". Depois de certo tempo, as falhas começam a aumentar devido ao desgaste dos componentes. Este é o "estágio de desgaste".

A função taxa de falha da distribuição Weibull não apresenta curva em forma de banheira, podendo ser apenas crescente ou decrescente. Apresentaremos a seguir gráficos da função taxa de falha da distribuição Zeta-Weibull, fixando dois dos parâmetros e variando o terceiro. Na Figura 3.7 fixamos os parâmetros $s = 4$ e $\beta = 0,5$ variando α. Note que o gráfico passa a ser crescente quando $\alpha = 1,3$. Essa mudança é importante para que a distribuição possa ser aplicada a diferentes tipos de conjuntos de dados, de acordo com suas peculiaridades.

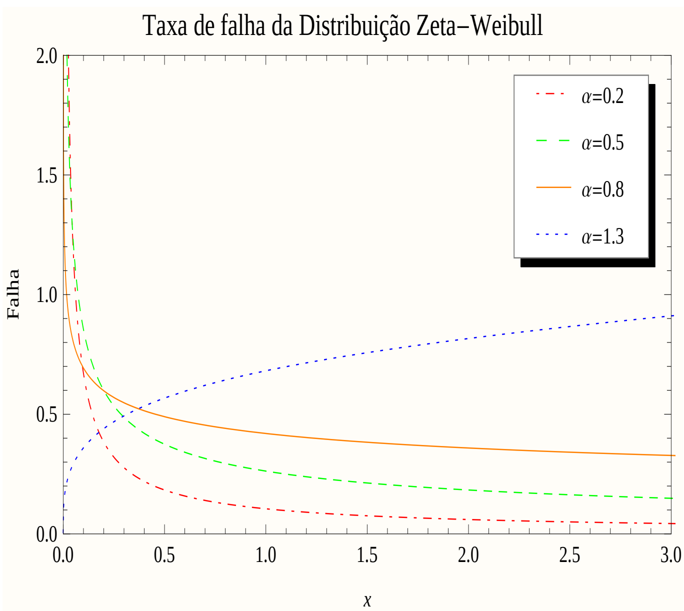

Figura 3.7: Taxa de Falha Zeta-Weibull variando o parâmetro α

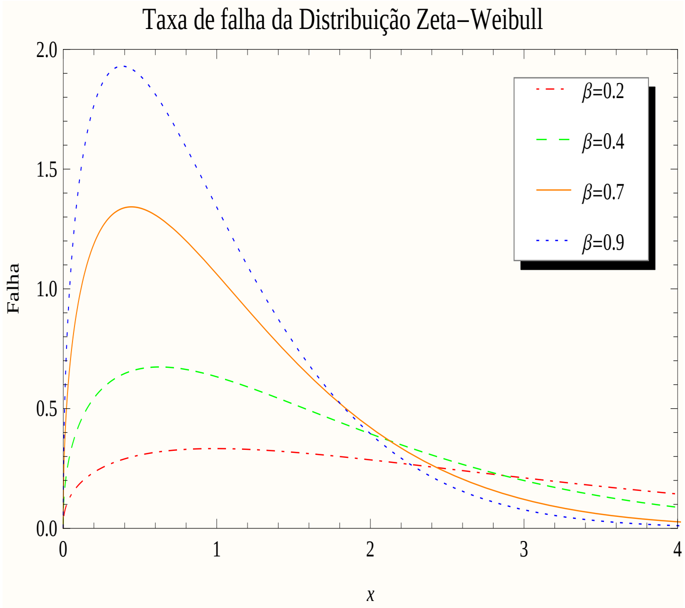

Figura 3.8: Taxa de Falha Zeta-Weibull variando o parâmetro β

Na Figura 3.8 fixamos os parâmetros $s = 3$ e $\alpha = 1,4$. Analisando o gráfico percebe-se que grande parte da densidade de falhas concentra-se ao redor de um determinado tempo de vida, o que pode ser caracterizado pelo desgaste natural da algum componente do objeto de estudo. Geralmente esse tipo de falha está relacionada ao tempo de operação de algum equipamento.

Na Figura 3.9 fixamos os parâmetros $\alpha = 1,5$ e $\beta = 0,7$, com s variando.

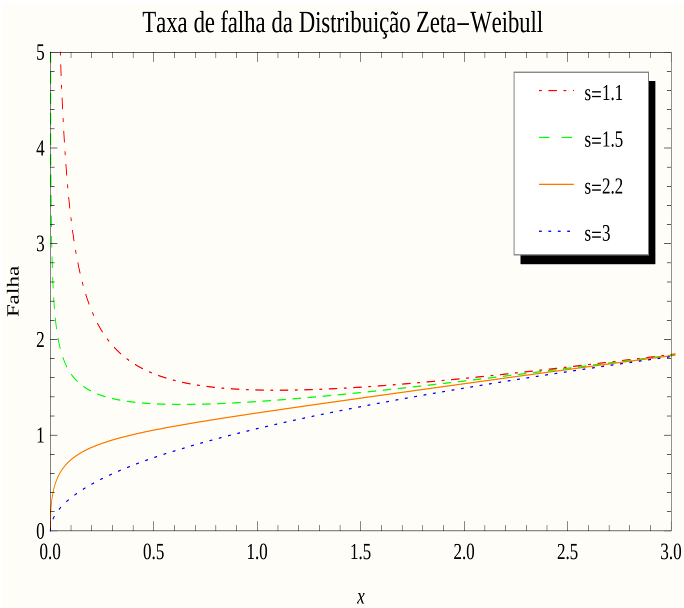

Figura 3.9: Taxa de Falha Zeta-Weibull variando o parâmetro s

3.2 Expansão da Função Densidade de Probabilidade

No teorema a seguir mostraremos que a função densidade de probabilidade da distribuição Zeta-Weibull pode ser escrita como combinação linear de densidades Weibull. Escrever a função densidade da Zeta-Weibull dessa forma é um resultado importante porque, assim, ela herdará as propriedades matemáticas da distribuição Weibull.

Teorema 3.5. *A função densidade de probabilidade da distribuição Zeta-Weibull é dada por* $f(x) = \sum_{k=1}^{\infty} \psi_k \, g(x, \alpha, k\beta)$*, onde* $g(x, \alpha, k\beta)$ *denota a Distribuição Weibull com parâmetros* α *e* $k\beta$ *e* $\psi_k = \dfrac{1}{\zeta(s)\, k^s}$ *.*

Demonstração. De fato, a fdp da distribuição Zeta-Weibull dada pela Equação (3.1) é

$$f(x) = \frac{\alpha\beta x^{\alpha-1} Li_{s-1}[\, e^{-\beta x^{\alpha}}]}{\zeta(s)}.$$

Substituindo a Equação (2.5) na expressão acima, obtemos

$$f(x) = \frac{\alpha\beta x^{\alpha-1} \displaystyle\sum_{k=1}^{\infty} \frac{(\, e^{-\beta x^{\alpha}})^k}{k^{s-1}}}{\zeta(s)}.$$

Donde,

$$\begin{aligned} f(x) &= \alpha\beta x^{\alpha-1} \sum_{k=1}^{\infty} \frac{(\, e^{-\beta x^{\alpha}})^k}{\zeta(s)\, k^{s-1}} \\ &= \sum_{k=1}^{\infty} \frac{\alpha\beta x^{\alpha-1}\, e^{-k\,\beta x^{\alpha}}}{\zeta(s)\, k^{s-1}}. \end{aligned}$$

Multiplicando e dividindo por k a fração acima, obtemos

$$f(x) = \sum_{k=1}^{\infty} \frac{\alpha\, k\beta x^{\alpha-1}\, e^{-k\,\beta x^{\alpha}}}{\zeta(s)\, k^{s-1}\, k}.$$

Portanto,

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{\zeta(s)\, k^s} \alpha k\beta x^{\alpha-1}\, e^{-k\beta x^{\alpha}}.$$

Comparando a última expressão com a função (2.3), trocando o parâmetro β por $k\beta$ e fazendo $\psi_k = \dfrac{1}{\zeta(s)\, k^s}$, a fdp da Zeta-Weibull pode ser escrita como

$$f(x) = \sum_{k=1}^{\infty} \psi_k \, g(x, \alpha, k\beta), \tag{3.6}$$

ou seja, como uma combinação linear de densidades Weibull, com parâmetros α e $k\beta$. $\square$

Integrando (3.6), escrevemos a fda da distribuição Zeta-Weibull como

$$F(x) = \sum_{k=1}^{\infty} \psi_k G(x, \alpha, k\beta). \tag{3.7}$$

É possível também encontrar a fda da distribuição Zeta-Weibull substituindo as Equações (2.1) e (2.5) na Equação (3.3), ou seja

$$\begin{aligned} F(x) &= \frac{\left[\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^s}\right] - \left[\displaystyle\sum_{k=1}^{\infty} \frac{\left[e^{-\beta x^{\alpha}}\right]^k}{k^s}\right]}{\zeta(s)} \\ &= \frac{\displaystyle\sum_{k=1}^{\infty} \frac{1 - \left[e^{-\beta x^{\alpha}}\right]^k}{k^s}}{\zeta(s)} \\ &= \sum_{k=1}^{\infty} \left[\frac{1 - (e^{-\beta x^{\alpha}})^k}{\zeta(s)\, k^s}\right], \end{aligned}$$

portanto

$$F(x) = \sum_{k=1}^{\infty} \psi_k \left[1 - e^{-k\beta x^{\alpha}}\right]. \tag{3.8}$$

3.3 Outra representação para a fdp e fda

Nesta seção encontraremos expansões para a fdp e fda da distribuição Zeta-Weibull utilizando o conceito de distribuições exponencializadas. Esta é uma representação muito útil, pois ao escrever uma densidade em termos de uma exponencializada pode-se utilizar suas propriedades.

Se $G(x)$ é uma distribuição qualquer, dizemos que uma variável aleatória X tem uma distribuição exponencializada, chamada de Exp-G, com parâmetro $t > 0$, se sua fdp e fda são dadas por

$$h_t(x) = t\, g(x)\, G(x)^{t-1} \tag{3.9}$$

e

$$H_t(x) = G(x)^t, \tag{3.10}$$

respectivamente. Note que, se $t = 1$, tem-se a própia distribuição base. Muitos resultados obtidos da Exp-G são utilizados em diversas classes generalizadas, uma vez que as expansões da fdp e fda dessas classes são combinações da própia Exp-G.

Aplicando a definição da função polilogarítmica dada na Equação (2.5) na Equação (3.2) abaixo

$$f(x)=\frac{Li_{s-1}[1-(1-\ e^{-\beta x^{\alpha}})]\,\alpha\beta x^{\alpha-1}\,e^{-\beta x^{\alpha}}}{\zeta(s)[1-(1-\ e^{-\beta x^{\alpha}})]},$$

temos que

$$f(x)=\frac{\displaystyle\sum_{r=1}^{\infty}\frac{[1-(1-\ e^{-\beta x^{\alpha}})]^{r}}{r^{s-1}}\,\alpha\beta x^{\alpha-1}\,e^{-\beta x^{\alpha}}}{\zeta(s)[1-(1-\ e^{-\beta x^{\alpha}})]},$$

ou seja,

$$f(x)=\sum_{r=1}^{\infty}\frac{[1-(1-\ e^{-\beta x^{\alpha}})]^{r-1}\,g(x)}{r^{s-1}\,\zeta(s)},\tag{3.11}$$

onde $g(x)$ é a fdp da distribuição Weibull dada em (2.3). Aplicando o binômio de Newton na expressão acima, temos

$$[1-(1-\ e^{-\beta x^{\alpha}})]^{r-1}=\sum_{k=0}^{r-1}(-1)^{k}(1-\ e^{-\beta x^{\alpha}})^{k}\binom{r-1}{k}.$$

Daí, reescrevendo (3.11)

$$f(x)=\sum_{r=1}^{\infty}\sum_{k=0}^{r-1}\frac{(-1)^{k}(1-\ e^{-\beta x^{\alpha}})^{k}\,g(x)}{r^{s-1}\,\zeta(s)}\binom{r-1}{k}.$$

Na Equação acmima, fazendo $k=t-1$, temos que $0\leq k\leq r-1\Rightarrow 0\leq t-1\leq r-1\Rightarrow 1\leq t\leq r$, ou seja,

$$f(x)=\sum_{r=1}^{\infty}\sum_{t=1}^{r}\frac{(-1)^{t-1}(1-\ e^{-\beta x^{\alpha}})^{t-1}\,g(x)}{r^{s-1}\,\zeta(s)}\binom{r-1}{t-1}$$

De $1\leq r\leq\infty$ e $1\leq t\leq r$, temos

$$1\leq t\leq r\leq\infty,$$

logo

$$1\leq t\leq\infty\quad\text{e}\quad t\leq r\leq\infty.$$

Portanto, $f(x)$ pode ser reescrita como

$$f(x)=\sum_{r=t}^{\infty}\sum_{t=1}^{\infty}\frac{(-1)^{t-1}(1-\ e^{-\beta x^{\alpha}})^{t-1}\,g(x)}{r^{s-1}\,\zeta(s)}\binom{r-1}{t-1}.$$

Substituindo a Equação (2.2) na expressão acima, temos

$$f(x)=\sum_{t=1}^{\infty}\sum_{r=t}^{\infty}\frac{(-1)^{t-1}G(x)^{t-1}\,g(x)}{r^{s-1}\,\zeta(s)}\binom{r-1}{t-1},$$

donde,

$$f(x) = \sum_{t=1}^{\infty} \sum_{r=t}^{\infty} \frac{(-1)^{t-1}}{t\, r^{s-1}\, \zeta(s)} \binom{r-1}{t-1} t\, g(x)\, G(x)^{t-1}.$$

Logo,

$$f(x) = \sum_{t=1}^{\infty} \omega_t\, h_t(x), \tag{3.12}$$

onde $h_t(x)$ é a função densidade de probabilidade da distribuição exponencializada definida em (3.9) e

$$\omega_t = \sum_{r=t}^{\infty} \frac{(-1)^{t-1}}{t\, r^{s-1}\, \zeta(s)} \binom{r-1}{t-1}$$

é uma constante.

Portanto, a densidade Zeta-Weibull pode ser escrita como uma combinação linear de densidades exponencializadas. Assim, podemos deduzir várias propriedades da distribuição Zeta-Weibull a partir das respectivas propriedades das distribuições exponencializadas que estão presentes, em grande número, na literatura.

Integrando (3.12), obtemos a seguinte expressão para a fda da distribuição Zeta-Weibull em função da distribuição exponencializada

$$F(x) = \sum_{t=1}^{\infty} \omega_t\, H_t(x), \tag{3.13}$$

em que $H_t(x)$ é a fda da distribuição exponencializada definida em (3.10).

3.4 Função Quantílica

Conforme Magalhães e Lima [10], página 27, os quantis são valores que limitam uma certa porcentagem de observações da variável. A notação para quantil será q_f, com f indicando a porcentagem definida pelo quantil. Por exemplo, q_{12} é o quantil que limita 12% dos valores inferiores do conjunto de observações ordenadas. Podemos dizer que f é uma proporção qualquer, com $0 < f < 1$, de maneira que 100% das observações sejam menores do que q_f.

Alguns quantis tem nomes particulares. Vejamos alguns deles.

- q_{25} é chamado de 1º quartil ou 25º percentil
- q_{50} é a mediana ou 2º quartil ou ainda 50º percentil
- q_{75} é chamado de 3º quartil ou 75º percentil
- q_{80} é chamado de 8º decil

Como vimos, Paixão [15] definiu a função quantílica da classe Zeta-G como em (2.7). Assim, para obtermos a função quantílica da distribuição Zeta-Weibull devemos substituir a função quantílica da distribuição Weibull em (2.7). Para isso, vamos encontrar a inversa da função de distribuição acumulada da distribuição Weibull dada em (2.2).

Fazendo $G(x) = u$, temos

$$u = 1 - \mathrm{e}^{-\beta \mathrm{x}^{\alpha}},$$

donde

$$\mathrm{e}^{-\beta \mathrm{x}^{\alpha}} = 1 - \mathrm{u}.$$

Aplicando a função logarítmica na base e a ambos os lados, obtemos

$$\log \mathrm{e}^{-\beta \mathrm{x}^{\alpha}} = \log(1 - \mathrm{u}).$$

Usando propriedades do logarítmo e isolando x, temos

$$-\beta x^{\alpha} = \log(1-u) \Rightarrow x^{\alpha} = \frac{-1}{\beta}\log(1-u) \Rightarrow x = \left[\frac{-1}{\beta}\log(1-u)\right]^{\frac{1}{\alpha}}.$$

Portanto, a função quantílica da distribuição Weibull é dada por

$$G^{-1}(u) = \left[\frac{-1}{\beta}\log(1-u)\right]^{\frac{1}{\alpha}}. \qquad (3.14)$$

Assim, aplicando a última expressão à Equação (2.7), obtemos a função quantílica da distribuição Zeta-Weibull.

Definição 3.6. *A função quantílica da distribuição Zeta-Weibull é dada por*

$$Q(w,s) = F^{-1}(w,s) = \left\{\frac{-1}{\beta}\log\left(1 - \left[\sum_{n=1}^{k}\frac{(-1)^n f_{n-1,n}}{n c_1^n}(w-w_0)^n\right]\right)\right\}^{\frac{1}{\alpha}}, \qquad (3.15)$$

onde $c_j = \sum_{k=1}^{\infty}\frac{(-1)^j\binom{k}{j}}{\zeta(s)k^s}$ *e os coeficientes* $f_{j,n}$ *são dados recursivamente por*

$$f_{j,n} = j^{-1}\sum_{m=1}^{j}\big[m(n+1)-j\big]d_m f_{j-m,n},$$

com $d_0 = 1$ *e* $d_j = -c_1^{-1}\sum_{i=1}^{j} d_{j-i}c_{i+1}$ *(*$j \geq 1$*).*

3.5 Momentos

Para calcularmos o r-ésimo momento da distribuição Zeta-Weibull, vamos substituir na expressão (1.6) a fdp da distribuição Zeta-Weibull que encontramos em (3.6), considerando que o suporte da distribuição Zeta-Weibull é $x \in [0, +\infty)$. Então

$$
\begin{aligned}
\mu_r^{'} &= \int_0^{+\infty} x^r f(x)\, dx \\
&= \int_0^{+\infty} x^r \sum_{k=1}^{\infty} \psi_k\, g(x, \alpha, k\beta)\, dx \\
&= \sum_{k=1}^{\infty} \int_0^{+\infty} x^r\, \psi_k\, \alpha k\beta x^{\alpha-1}\, \mathrm{e}^{-\mathrm{k}\beta\mathrm{x}^{\alpha}}\, \mathrm{dx} \\
&= \sum_{k=1}^{\infty} \psi_k\, \alpha k\beta \int_0^{+\infty} x^r x^{\alpha-1}\, e^{-k\beta x^{\alpha}}\, dx \\
&= \sum_{k=1}^{\infty} \psi_k\, \alpha k\beta \int_0^{+\infty} x^{r+\alpha-1}\, e^{-k\beta x^{\alpha}}\, dx.
\end{aligned}
$$

Usando o software **Mathematica**, calculamos a integral na expressão acima como

$$\int_0^{+\infty} x^{r+\alpha-1}\, e^{-k\beta x^{\alpha}}\, dx = \frac{(k\beta)^{-\frac{r+\alpha}{\alpha}}\Gamma(\frac{r+\alpha}{\alpha})}{\alpha}$$

onde Γ é a função Gama de Euler definida como $\Gamma(\alpha) = \displaystyle\int_0^{\infty} x^{\alpha-1} e^{-x}\, dx$.

Portanto,

$$\mu_r^{'} = \sum_{k=1}^{\infty} \psi_k\, \alpha k\beta \frac{(k\beta)^{-\frac{r+\alpha}{\alpha}}\Gamma(\frac{r+\alpha}{\alpha})}{\alpha},$$

Simplificando, obtemos

$$\mu_r^{'} = \beta^{-\frac{r}{\alpha}}\, \Gamma\left(\frac{r+\alpha}{\alpha}\right) \sum_{k=1}^{\infty} \psi_k\, k^{-\frac{r}{\alpha}}.$$

Como $\psi_k = \dfrac{1}{\zeta(s)\, k^s}$, então

$$\mu_r^{'} = \beta^{-\frac{r}{\alpha}}\, \Gamma\left(\frac{r+\alpha}{\alpha}\right) \sum_{k=1}^{\infty} \frac{1}{\zeta(s)\, k^s}\, k^{-\frac{r}{\alpha}},$$

ou seja,

$$\mu_r^{'} = \zeta(s)^{-1}\, \beta^{-\frac{r}{\alpha}}\, \Gamma\left(\frac{r+\alpha}{\alpha}\right) \sum_{k=1}^{\infty} \frac{k^{-\frac{r}{\alpha}}}{k^s}.$$

Portanto,

$$\mu_r' = \zeta(s)^{-1}\,\beta^{-\frac{r}{\alpha}}\,\Gamma\left(\frac{r+\alpha}{\alpha}\right)\sum_{k=1}^{\infty}\frac{1}{k^{\frac{r+\alpha\,s}{\alpha}}}.$$

Usando a Equação (2.1), escrevemos o somatório acima como

$$\sum_{k=1}^{\infty}\frac{1}{k^{\frac{r+\alpha\,s}{\alpha}}} = \zeta\left(\frac{r+\alpha\,s}{\alpha}\right).$$

Portanto, a expressão geral para os momentos da distribuição Zeta-Weibull é dada por

$$\mu_r' = \zeta(s)^{-1}\,\beta^{-\frac{r}{\alpha}}\,\zeta\left(\frac{r+\alpha\,s}{\alpha}\right)\Gamma\left(\frac{r+\alpha}{\alpha}\right). \tag{3.16}$$

É possível usar a última expressão para determinar os quatro primeiros momentos da distribuição Zeta-Weibull, ou seja

$$\begin{aligned}
\mu_1' &= \zeta(s)^{-1}\,\beta^{-\frac{1}{\alpha}}\,\zeta\left(\frac{1+\alpha\,s}{\alpha}\right)\Gamma\left(\frac{1+\alpha}{\alpha}\right).\\
\mu_2' &= \zeta(s)^{-1}\,\beta^{-\frac{2}{\alpha}}\,\zeta\left(\frac{2+\alpha\,s}{\alpha}\right)\Gamma\left(\frac{2+\alpha}{\alpha}\right).\\
\mu_3' &= \zeta(s)^{-1}\,\beta^{-\frac{3}{\alpha}}\,\zeta\left(\frac{3+\alpha\,s}{\alpha}\right)\Gamma\left(\frac{3+\alpha}{\alpha}\right).\\
\mu_4' &= \zeta(s)^{-1}\,\beta^{-\frac{4}{\alpha}}\,\zeta\left(\frac{4+\alpha\,s}{\alpha}\right)\Gamma\left(\frac{4+\alpha}{\alpha}\right).
\end{aligned}$$

3.6 Função Geratriz de Momentos

Substituindo a fdp da distribuição Zeta-Weibull dada em (3.6) na Equação (1.7), encontramos a função geratriz de momentos da distribuição Zeta-Weibull, ou seja

$$M_x(t) = \int_0^{+\infty} e^{tx}\sum_{k=1}^{\infty}\psi_k\,\alpha k\beta x^{\alpha-1}\cdot e^{-k\beta x^{\alpha}}\,dx.$$

Expandindo em série de potências a função exponencial acima, obtemos

$$M_x(t) = \sum_{k=1}^{\infty}\psi_k\,\alpha k\beta\int_0^{+\infty} e^{tx}x^{\alpha-1}\sum_{n=0}^{\infty}\frac{(-1)^n k^n\beta^n x^{n\alpha}}{n!}\,dx.$$

Utilizando propriedades de integração, temos

$$M_x(t) = \sum_{k=1}^{\infty}\sum_{n=0}^{\infty} \frac{(-1)^n\, \psi_k\, \alpha(k\beta)^{n+1}}{n!} \int_0^{+\infty} e^{tx} x^{\alpha(n+1)-1}\, dx.$$

Como $\psi_k = \dfrac{1}{\zeta(s)\, k^s}$, então

$$M_x(t) = \sum_{k=1}^{\infty}\sum_{n=0}^{\infty} \frac{(-1)^n\, \alpha\, k^{n-s+1}\beta^{n+1}}{\zeta(s)\, n!} \int_0^{+\infty} e^{tx} x^{\alpha(n+1)-1}\, dx.$$

A última integral pode ser resolvida usando o software **Mathematica**, reduzindo-se a

$$\int_0^{+\infty} e^{tx} x^{\alpha(n+1)-1}\, dx = (-1)^{-\alpha(n+1)} t^{-\alpha(n+1)} \Gamma(\alpha + n\alpha)$$

onde Γ é a função Gama de Euler definida como $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx$.

Portanto, podemos escrever a função geratriz de momentos (fgm) da distribuição Zeta-Weibull como

$$M_x(t) = \sum_{k=1}^{\infty}\sum_{n=0}^{\infty} \frac{(-1)^{n-\alpha(n+1)}\, \alpha\, k^{n-s+1}\beta^{n+1} t^{-\alpha(n+1)} \Gamma(\alpha + n\alpha)}{\zeta(s) n!}. \tag{3.17}$$

3.7 Entropia de Rényi

A entropia de Rényi da distribuição Zeta-Weibull pode ser determinada substituindo as Equações (2.2) e (2.3) na Equação (2.8). Então

$$\begin{aligned} I_R(\gamma) &= \frac{\gamma}{\gamma - 1} \log[\zeta(s)] + \frac{1}{1-\gamma} \log\Bigg\{ \sum_{j,i,k=0}^{\infty} \sum_{v=0}^{k+i} \frac{(-1)^{k+i} g_i\, e_{i,k} \binom{k+i}{v} \Gamma(\gamma + j)}{\Gamma(\gamma) j!} \\ &\times \int_0^{\infty} (\alpha\beta x^{\alpha-1}\, e^{-\beta x^{\alpha}})^{\gamma} (1 - e^{-\beta x^{\alpha}})^{-(-j-v)} dx \Bigg\}, \end{aligned} \tag{3.18}$$

onde as constantes g_i e $e_{i,k}$ são dadas na Seção 2.2. Utilizando a série de potência

$$(1-z)^{-a} = \sum_{w=0}^{\infty} \frac{\Gamma(a+w)}{\Gamma(a) w!} z^w, \tag{3.19}$$

para qualquer a real e $|z| < 1$, a integral em (3.18) reduz-se a

$$
\begin{aligned}
\int_0^\infty g(x)^\gamma G(x)^{j+v}\,dx &= \int_0^\infty (\alpha\beta x^{\alpha-1}\,e^{-\beta x^\alpha})^\gamma (1-\,e^{-\beta x^\alpha})^{-(-j-v)}\,dx \\
&= \int_0^\infty \alpha^\gamma\beta^\gamma x^{\gamma(\alpha-1)}\,e^{-\gamma\beta x^\alpha}\,\sum_{w=0}^\infty \frac{\Gamma(-j-v+w)}{\Gamma(-j-v)\,w!}\,e^{-w\beta x^\alpha}\,dx \\
&= \sum_{w=0}^\infty \frac{\alpha^\gamma\beta^\gamma\,\Gamma(-j-v+w)}{\Gamma(-j-v)\,w!}\int_0^\infty x^{\gamma(\alpha-1)}\,e^{-\gamma\beta x^\alpha}\,e^{-w\beta x^\alpha}\,dx \\
&= \sum_{w=0}^\infty \frac{\alpha^\gamma\beta^\gamma\,\Gamma(-j-v+w)}{\Gamma(-j-v)\,w!}\int_0^\infty x^{\gamma(\alpha-1)}\,e^{-\beta x^\alpha(\gamma+w)}\,dx \qquad (3.20)
\end{aligned}
$$

Resolvendo a integral na última equação, obtemos

$$
\begin{aligned}
\int_0^\infty g(x)^\gamma G(x)^{j+v}\,dx &= \sum_{w=0}^\infty \frac{\alpha^\gamma\beta^\gamma\,\Gamma(-j-v+w)}{\Gamma(-j-v)\,w!}\times\frac{[\beta(w+\gamma)]^{\frac{\gamma(1-\alpha)-1}{\alpha}}\,\Gamma\left(\frac{\gamma(\alpha-1)+1}{\alpha}\right)}{\alpha} \\
&= \sum_{w=0}^\infty \frac{\alpha^{\gamma-1}\beta^{\frac{\gamma-1}{\alpha}}(w+\gamma)^{\frac{\gamma(1-\alpha)-1}{\alpha}}\,\Gamma(-j-v+w)}{w!\,\Gamma(-j-v)}\Gamma\left(\frac{\gamma(\alpha-1)+1}{\alpha}\right).
\end{aligned}
$$

Substituindo o resultado acima em (3.18), obtemos a expressão para a entropia de Rényi da distribuição Zeta-Weibull como

$$
\begin{aligned}
I_R(\gamma) &= \frac{\gamma}{\gamma-1}\log[\zeta(s)]+\frac{1}{1-\gamma}\log\Bigg\{\sum_{j,i,k=0}^\infty\sum_{v=0}^{k+i}\frac{(-1)^{k+i}g_i\,e_{i,k}\binom{k+i}{v}\Gamma(\gamma+j)}{\Gamma(\gamma)j!} \\
&\times \sum_{w=0}^\infty \frac{\alpha^{\gamma-1}\beta^{\frac{\gamma-1}{\alpha}}(w+\gamma)^{\frac{\gamma(1-\alpha)-1}{\alpha}}\,\Gamma(-j-v+w)}{w!\,\Gamma(-j-v)}\Gamma\left(\frac{\gamma(\alpha-1)+1}{\alpha}\right)\Bigg\}.
\end{aligned}
$$

3.8 Confiabilidade

Confiabilidade é a probabilidade de que um sistema, equipamento ou componente desempenhe a função para a qual foi designado por um período de tempo especificado e sob um conjunto de condições pré estabelecidas. Em um sentido mais amplo, podemos dizer que a confiabilidade está associada à operação bem sucedida de um produto ou sistema, não ocorrendo portanto quebras ou falhas.

Se X_1 e X_2 são variáveis aleatórias independentes, a confiabilidade R é definida como $R = P(X_2 < X_1)$. Sabe-se que a expressão para a confiabilidade R pode ser representada por

$$
R = \int_0^\infty f_1(x)\,F_2(x)\,dx \qquad (3.21)
$$

onde $f(x)$ e $F(x)$ são a fdp e a fda, respectivamente.

Sejam X_m e X_n variáveis aleatórias independentes com $X_n \sim ZW(s,\alpha,\beta_1)$ e $X_m \sim ZW(s,\alpha,\beta_2)$. Substituindo (3.12) e (3.13) em (3.21), obtemos

$$R = \int_0^\infty \sum_{m,n=1}^{\infty} \omega_m \, h_m(x) \, \omega_n \, H_n(x) \, dx,$$

logo

$$R = \sum_{m,n=1}^{\infty} \omega_m \, \omega_n \int_0^\infty h_m(x) \, H_n(x) \, dx, \tag{3.22}$$

onde

$$\omega_m = \sum_{r=m}^{\infty} \frac{(-1)^{m-1}}{m \, r^{s-1} \, \zeta(s)} \binom{r}{m-1}$$

e

$$\omega_n = \sum_{r=n}^{\infty} \frac{(-1)^{n-1}}{n \, r^{s-1} \, \zeta(s)} \binom{r}{n-1}.$$

Substituindo (3.9) e (3.10) na Equação (3.22), obtemos

$$R = \sum_{m,n=1}^{\infty} \omega_m \, \omega_n \int_0^\infty m \, g_m(x) \, G_m(x)^{m-1} \, G_n(x)^n \, dx,$$

e portanto

$$R = \sum_{m,n=1}^{\infty} m \, \omega_m \, \omega_n \int_0^\infty (\alpha\beta_1 \, x^{\alpha-1} \, e^{-\beta_1 x^\alpha})(1-e^{-\beta_1 x^\alpha})^{m-1}(1-e^{-\beta_2 x^\alpha})^n \, dx.$$

Aplicando o binômio de Newton nos dois últimos termos, temos que

$$(1-e^{-\beta_1 x^\alpha})^{m-1} = \sum_{k=0}^{m-1} (-1)^k \, (e^{-\beta_1 x^\alpha})^k \binom{m-1}{k}$$

e

$$(1-e^{-\beta_2 x^\alpha})^{n} = \sum_{l=0}^{n} (-1)^l \, (e^{-\beta_2 x^\alpha})^l \binom{n}{l},$$

com $0 < k < m-1$ e $0 < l < n$. Daí

$$\begin{aligned} R &= \sum_{m,n=1}^{\infty} \sum_{k=0}^{m-1} \sum_{l=0}^{n} m \, \omega_m \, \omega_n \int_0^\infty \alpha\beta_1 \, x^{\alpha-1} \, e^{-\beta_1 x^\alpha} \, (-1)^k \, (e^{-\beta_1 x^\alpha})^k \binom{m-1}{k} \times \\ &\times \; (-1)^l \, (e^{-\beta_2 x^\alpha})^l \binom{n}{l} \, dx. \end{aligned}$$

Portanto,

$$R = \sum_{m,n=1}^{\infty} \sum_{k=0}^{m-1} \sum_{l=0}^{n} (-1)^{k+l}\, m\, \alpha\, \beta_1 \omega_m\, \omega_n \binom{m-1}{k} \binom{n}{l} \times \int_0^{\infty} x^{\alpha-1}\, e^{-x^{\alpha}[\beta_1(1+k)+\beta_2 l]}\, dx.$$

Resolvendo a integral na equação acima, obtemos

$$R = \sum_{m,n=1}^{\infty} \sum_{k=0}^{m-1} \sum_{l=0}^{n} (-1)^{k+l}\, m\, \alpha\, \beta_1 \omega_m\, \omega_n \binom{m-1}{k} \binom{n}{l} \times \frac{1}{\alpha\beta_1 + k\alpha\beta_1 + l\alpha\beta_2},$$

portanto, a expressão para a confiabilidade da distribuição Zeta-Weibull é dada por

$$R = \sum_{m,n=1}^{\infty} \sum_{k=0}^{m-1} \sum_{l=0}^{n} \frac{(-1)^{k+l}\, m\, \beta_1 \omega_m\, \omega_n}{\beta_1 + k\beta_1 + l\beta_2} \binom{m-1}{k} \binom{n}{l}, \tag{3.23}$$

onde ω_m e ω_n foram definidos anteriormente.

3.9 Estimação

Chamamos de inferência estatística a técnica que nos permite fazer afirmações e tirar conclusões, com base em dados parciais ou reduzidos de uma população, ou seja, a partir de uma amostra aleatória. Em outras palavras, podemos dizer que a inferência estatística é o processo de decisão que nos permite estimar características populacionais a partir da observação de uma parte reduzida dessa população. Estimativa é um valor atribuído a um parâmetro de uma população baseado na observação de uma amostra. Estimador é a estatística da amostra utilizada para estimar um parâmetro da população. Por exemplo, suponha que em uma fábrica de rodas para motocicletas tenha sido identificado um problema no espaçamento entre os raios das rodas. Analisar todas as rodas fabricadas é economicamente inviável, por isso pode-se coletar uma amostra aleatória, digamos com n rodas, e medir os espaçamentos entre os raios. Suponha que a média entre esses espaçamentos seja de $7,3cm$. Assim, $7,3cm$ é uma estimativa para o estimador média aritmética.

A estimação de parâmetros da distribuição Zeta-Weibull será feita utilizando-se o método da máxima verossimilhança. Esse método consiste no processo de obtenção de informação sobre um vetor de parâmetros θ, ou seja, consiste em obter estimadores para os parâmetros de interesse.

Definição 3.7. *Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de tamanho n da variável aleatória X. A função de verossimilhana L é definida por*

$$L(\theta; X_1, X_2, \ldots, X_n) = f(x_1;\theta) \times f(x_2;\theta) \times \ldots \times f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta),$$

em que θ denota o parâmetro desconhecido.

Para caracterizar melhor a amostra observada temos que encontrar o valor de θ que maximize a função $L(\theta)$, ou seja, o valor de θ que maximiza a probabilidade da amostra observada ocorrer. A função de verossimilhança mostra que a contribuição de cada observação não censurada é a sua função de densidade, enquanto que a contribuição de cada observação censurada não é a sua função de densidade. As observações censuradas deixam claro que o tempo de falha é maior que o tempo de censura observado e portanto, sua contribuição para a função de verossimilhança é a sua função de confiabilidade.

O logaritmo natural da função de verossimilhança de θ é denotado por

$$l(\theta) = \log\, L(\theta).$$

Assim, o valor de θ que maximiza a função de verossimilhança L(θ), também maximiza l(θ), a menos de uma constante.

Utilizando a fdp da distribuição Zeta-Weibull dada pela Equação (3.1), encontramos a função de log-verossimilhança para a distribuição Zeta-Weibull, com parâmetros (s, α, β), que pode ser expressa como

$$l(s,\alpha,\beta) = \log\left[\prod_{i=1}^{n} f(s,\alpha,\beta)\right] = \log\left[\prod_{i=1}^{n} \frac{\alpha\beta\, x_i^{\alpha-1} Li_{s-1}[\,e^{-\beta\, x_i^{\alpha}}]}{\zeta(s)}\right]$$

Aplicando propriedades dos logaritmos

$$\begin{aligned} l(s,\alpha,\beta) &= \log\left[\frac{\alpha\beta\, x_1^{\alpha-1} Li_{s-1}[\,e^{-\beta\, x_1^{\alpha}}]}{\zeta(s)} \times \ldots \times \frac{\alpha\beta\, x_n^{\alpha-1} Li_{s-1}[\,e^{-\beta\, x_n^{\alpha}}]}{\zeta(s)}\right] \\ &= \sum_{i=1}^{n} \log(\alpha\,\beta\, x_i^{\alpha-1}) + \sum_{i=1}^{n} \log\left[Li_{s-1}[\,e^{-\beta\, x_i^{\alpha}}]\right] - n\, \log\zeta(s). \end{aligned} \tag{3.24}$$

Os componentes do vetor escore $U = \left(\frac{\partial\, l}{\partial\, s}, \frac{\partial\, l}{\partial\alpha}, \frac{\partial\, l}{\partial\beta}\right)$ são obtidos por diferenciação. Derivando parcialmente a expressão (3.24) em relação a s, α e β, respectivamente, e igualando o resultado a zero, obtemos

$$\frac{\partial\, l(s,\alpha,\beta)}{\partial\, s} = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\, s} Li_{s-1}[\,e^{-\beta\, x_i^{\alpha}}]}{Li_{s-1}[\,e^{-\beta\, x_i^{\alpha}}]} + \frac{n \displaystyle\sum_{k=1}^{\infty} k^{-s}\, \log k}{\zeta(s)} = 0, \tag{3.25}$$

$$\frac{\partial\, l(s,\alpha,\beta)}{\partial\alpha} = -\beta \sum_{i=1}^{n} \frac{x_i^{\alpha}\, \log(x_i)\, Li_{s-2}[\,e^{-\beta\, x_i^{\alpha}}]}{Li_{s-1}[\,e^{-\beta\, x_i^{\alpha}}]} + \sum_{i=1}^{n} \frac{(1 + \alpha\, \log(x_i))}{\alpha} = 0 \tag{3.26}$$

e

$$\frac{\partial\, l(s,\alpha,\beta)}{\partial \beta} \;=\; \frac{n}{\beta} - \sum_{i=1}^{n} \frac{x_i^{\alpha} Li_{s-2}[\, e^{-\beta\, x_i^{\alpha}}]}{Li_{s-1}[\, e^{-\beta\, x_i^{\alpha}}]} = 0. \tag{3.27}$$

Utilizamos o software **Mathematica** para obter as derivadas acima. Os estimadores de máxima verossimilhança $\hat{s}$, $\hat{\alpha}$ e $\hat{\beta}$ para os parâmetros s, α e β, respectivamente, são obtidos pela solução interativa das expressõs (3.25), (3.26) e (3.27). As estimativas iniciais para s, α e β são obtidas ajustando a desidade Zeta-Weibull aos dados.

3.10 Aplicação

Nesta seção, analisaremos um conjunto de dados reais usando a distribuição Zeta-Weibull, comparando-a com as distribuições Zeta-Rayleigh(ZR), Zeta Gumbel(ZG), Zeta-Fréchet(ZF), Weibull Estendida(WE) e Beta Weibull Poisson(BWP). Para a estimação dos parâmetros, adotamos o máximo da função de verossimilhança e todos os cálculos foram realizados usando o pacote bbmle, de autoria de Benjamin Bolker, do software R(versão 3.3.3). Os critérios de informação utilizados foram os de Akaike(AIC), Akaike Corrigido(AICc), Bayesiano de Schwarz(BIC) e o critério de Hannan-Quinn(HQIC), critérios estes que dependem do número de observações e do número de parâmetros, sendo frequentemente utilizados para selecionar modelos em diversas áreas. Esses critérios, apesar de conceitualmente diferentes, utilizam o máximo da função de verossimilhança como medida de ajustamento, definindo, porém, valores diferentes. Em todos os casos o melhor modelo será aquele que apresentar menor valor de AIC, AICc, BIC e HQIC.

O critério de Akaike pode ser calculado como

$$AIC = -2l(\theta) + 2p,$$

em que p denota o número de parâmetros.

O critério de Akaike corrigido pode ser calculado como

$$AICc = AIC + \frac{2(p+1)(p+2)}{n-p-2},$$

p denotando o número de parâmetros e n o número de observações da amostra.

O critério de Schwarz pode ser calculado como

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

p denotando o número de parâmetros e n o número de observações da amostra.

O critério de Hannan-Quinn pode ser calculado como

$$HQIC = -2l(\theta) + 2p\log(\log(n)),$$

onde p é o número de parâmetros e n é o número de observações da amostra.

O conjunto de dados utilizados mostra o tempo entre falhas das bombas de reatores secundários do reator RSG-GAS, obtidos a partir de [17]. Os dados estão na Tabela 3.1.

Tabela 3.1: Tempo entre falhas das bombas de reactores secundários(milhares de horas)

2.160	0.746	0.402	0.954	0.491	6.560	4.992	0.347
0.150	0.358	0.101	1.359	3.465	1.060	0.614	1.921
4.082	0.199	0.605	0.273	0.070	0.062	5.320	

As funções densidades das distribuições Zeta-Rayleigh, Zeta-Gumbel, Zeta-Fréchet, Weibull Estendida e Beta Weibull Poisson são dadas, respectivamente, por

$$f_{ZR}(x) = \frac{Li_{s-1}\left[e^{\frac{-x^2}{2\sigma^2}}\right] \cdot x}{\zeta(s) \cdot \sigma^2},$$

$$f_{ZG}(x) = \frac{Li_{s-1}\left[e^{-e^{-\frac{x-\lambda}{\alpha}}}\right] \cdot e^{\frac{x-\lambda}{\alpha} - e^{\frac{x-\lambda}{\alpha}} + e^{-\frac{x-\lambda}{\alpha}}}}{\zeta(s) \cdot \alpha},$$

$$f_{ZF}(x) = \frac{\lambda\,\sigma^{\lambda}\,x^{-\lambda-1}\,Li_{s-1}\left[1 - e^{-\left(\frac{\sigma}{x}\right)^{\lambda}}\right]}{\zeta(s)\left[e^{\left(\frac{\sigma}{x}\right)^{\lambda}} - 1\right]},$$

$$f_{WE}(x) = \frac{\lambda\,\beta}{\alpha} \cdot \left(\frac{x}{\alpha}\right)^{\beta-1} \cdot e^{-\lambda\left(\frac{x}{\alpha}\right)^{\beta}},$$

e

$$f_{BWP}(x) = \frac{\alpha\,\beta\,\lambda\,e^{-\lambda}\,(e^{\lambda}-1)^{2-a-b}}{B(a,b)\,(1-e^{-\lambda})} e^{-\beta\,x^{\alpha}}\,x^{\alpha-1}\,e^{\lambda\,e^{-\beta\,x^{\alpha}}}$$

A Tabela 3.2 fornece as estimativas dos parâmetros e os valores que maximizam a função Log-Verossimilhança das distribuição ZW, ZR, ZG, ZF, WE e BWP. Os valores de AIC, BIC, HQIC e AICc das respectivas distribuições estão na Tabela 3.3.

Tabela 3.2: Estimativas dos parâmetros e Máximo da Log-Verossimilhança

Distribuição	Estimativas	Log
ZW(s,α,β)	2.3482 0.9315 0.5546	-36.6136
ZR(s,σ)	14.8759 0.9490	-40.5191
ZG(s,α,λ)	13.5467 2.2630 2.6282	-86.2777
ZF(s,σ,λ)	9.8556 0.3572 0.7829	-37.2495
WE(α,λ,β)	7.9501 10.0866 0.8077	-32.5139
BWP(α,β,λ,a,b)	27.7821 15.6510 0.1554 0.1311 8.7452	-31.8115

Tabela 3.3: Critérios de estimação

Distribuição	AIC	BIC	HQIC	AICc
ZW(s,α,β)	70.7060	80.7899	74.7813	70.8209
ZR(s,σ)	85.0383	91.7609	87.7551	85.1149
ZG(s,α,λ)	110.9246	121.0085	114.9998	111.0394
ZF(s,σ,λ)	71.8850	81.9688	75.9602	71.9998
WE(α,λ,β)	71.0278	81.1117	75.1030	71.1426
BWP(α,β,λ,a,b)	73.6231	90.4296	80.4152	73.8145

Como podemos ver na Tabela 3.3, a distribuição Zeta-Weibull tem os menores valores de AIC, BIC, HQIC e AICc entre os seis modelos ajustados e, portanto, se ajusta melhor aos dados. O histograma dos dados com as densidades das distribuições ZW, ZR, ZG, ZF, WE e BWP é apresentado na Figura 3.10. Nele, podemos notar que a curva da densidade Zeta-Weibull modela os dados com um ajuste melhor que as demais distribuições.

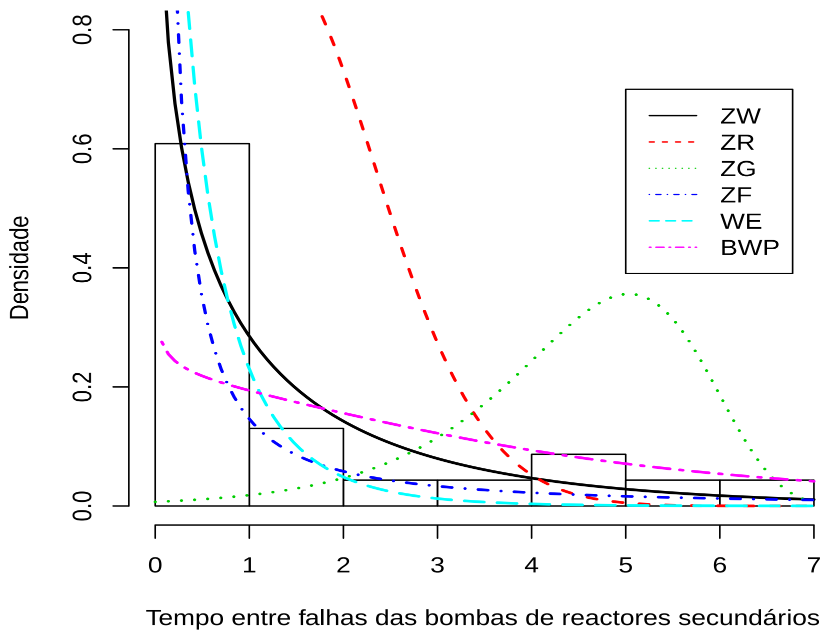

Figura 3.10: Histograma dos dados e as fdps das distribuições ZW, ZR, ZG, ZF, WE e BWP

Capítulo 4

O Ensino da Estatística

O presente capítulo tem como objetivo instrumentalizar o professor do ensino básico para o aprimoramento do exercício da docência nos conteúdos de Probabilidade e Estatística, tomando como referência suas experiências, dificuldades, expectativas e, fundamentalmente, sua busca constante em tornar o ato de ensinar mais natural e agradável, dotando-o de sentido e contextualização. Além disso, o docente poderá interagir com outros professores sobre as possibilidades de aplicação prática dessas técnicas sempre visando a otimização das oportunidades e melhorando a relação ensino/aprendizagem.

Apesar do desnivelamento aparente entre este capítulo e os anteriores, resaltamos que, este texto como um todo, serve de embasamento teórico para os docentes em matemática que trabalham em qualquer nível de ensino. Discutiremos sobre a importância do desenvolvimento dos conteúdos de estatística no Ensino Fundamental e Médio, assim como a preparação dos professores de matemática para o trabalho com estes conteúdos. Dessa forma, contribuiremos de forma didática para a difusão dos temas abordados entre os profissionais da área e, portanto, alcançaremos diretamente os discentes desde os primeiros anos de ensino.

É notório que a palavra *Estatística* permeia a vida dos cidadãos que acompanham diariamente os noticiários, sejam telejornais, internet, revistas ou outros. A necessidade de interpretar corretamente as informações disponibilizadas por estes meios se faz a cada dia mais necessária, o que leva à necessidade de se ensinar estatística a um número de pessoas cada vez maior. Segundo os Parâmetros Curriculares Nacionais do Ensino Médio PCNEM [1], a análise de dados tem sido essencial em problemas sociais e econômicos, como nas estatísticas relacionadas à saúde, populações, transportes, orçamentos e questões de mercado.

É importante frisar que boas técnicas aplicadas ao ensino da matemática no Ensino Fundamental e Médio, podem ser deteminantes para a compreensão na leitura de informações que circulam, na mídia em geral, na forma de gráficos, tabelas e informações de caráter estatístico. Nesta perspectiva, uma boa formação do professor do Ensino fundamental e Médio torna-se imprescindível. Alguma deficiência existente na formação desses professores pode acarretar em uma inadequada atenção aos tópicos de estatística nos currículos escolares, deixando-os em segundo plano. Como destaca Lopes [9], "ao ensino da Matemática fica o compromisso de não só ensinar o domínio dos números, mas também a organização de dados e leitura de gráficos."

Dentre muitos pontos importantes que devem ser avaliados no ensino da estatística no Ensino Médio, devemos levar em consideração as sugestões curriculares oficiais, que têm influência forte no planejamento e na construção do currículo, seja na escola pública ou privada. O livro didático nas escolas públicas, distribuidos gratuitamente através do Programa Nacional do Livro Didático (PNLD), determina, de maneira prática, como os conteúdos devem ser discutidos com os estudantes, ou seja, eles têm influência na prática dos professores em sala de aula e, portanto, é importante conhecer como eles abordam o ensino de Estatística.

Segundo os PCNEM, estatística e probabilidade lidam com dados e informações em conjuntos finitos e utilizam procedimentos que permitem controlar com certa segurança a incerteza e mobilidade desses dados, por isso, a contagem ou análise combinatória é parte instrumental desse tema. Os conteúdos e respectivas habilidades propostos para o ensino da estatística no Ensino Médio são:

1. Estatística: descrição de dados, representações gráficas, análise de dados, médias, moda e mediana, variância e desvio padrão.

- Identificar formas adequadas para descrever e representar dados numéricos e informações de natureza social, econômica, política, científico-tecnológica ou abstrata.
- Ler e interpretar dados e informações de caráter estatístico apresentados em diferentes linguagens e representações, na mídia ou em outros textos e meios de comunicação.
- Obter médias e avaliar desvios de conjuntos de dados ou informações de diferentes naturezas.
- Compreender e emitir juízos sobre informações estatísticas de natureza social, econômica, política ou científica apresentados em textos, notícias, propagandas, censos, pesquisas e outros.

2. Contagem: princípio multiplicativo; problemas de contagem.

- Decidir sobre a forma mais adequada de organizar números e informações com o objetivo de simplificar cálculos em situações reais envolvendo grande quantidade de dados ou de eventos.
- Identificar regularidades para estabelecer regras e propriedades em processos nos quais se fazem necessários os processos de contagem.

- Identificar dados e relações envolvidas numa situação-problema que envolva o raciocínio combinatório, utilizando os processos da contagem.

3. Probabilidade: possibilidades; cálculo de probabilidades.

- Reconhecer o caráter aleatório de fenômenos e eventos naturais, científico-tecnológicos ou sociais, compreendendo o significado e a importância da probabilidade como meio de prever resultados.

- Quantificar e fazer previsões em situações aplicadas a diferentes áreas de conhecimento e da vida cotidiana que envolvam o pensamento probabilístico.

- Identificar em diferentes áreas científicas e outras atividades práticas modelos e problemas que fazem uso de estatísticas e probabilidades.

A escolha de uma forma e sequência de distribuição dos temas nas três séries do ensino médio traz em si um projeto de formação dos alunos. Nesta perspectiva, os PCNEM propõem a seguinte distribuição:

1ª série Estatística: descrição de dados; gráficos.

2ª série Estatística: análise de dados. Contagem.

3ª série Probabilidade.

Dadas as bases para o ensino e a aprendizagem da Estatística no ensino médio, segue este capítulo que tem por objetivo desenvolver um conjunto de ações, em forma de atividades, para que o professor possa trabalhar em sala de aula, tentando assim, melhorar o estímulo ao aprendizado dos alunos.

4.1 Atividade 1 - Pesquisa hobby

Objetivo: coleta de dados, construção de tabelas e gráficos, comparando resultados.

Desenvolvimento: os alunos devem coletar os dados e posteriormente construir tabelas e gráficos para a discussão dos resultados.

Cada aluno deverá fazer uma sondagem na rua onde mora, entrevistando 25 pessoas para saber qual é o hobby preferido de cada uma delas. Suponha que os resultados da pesquisa de um determinado aluno estejam organizados na tabela abaixo:

Tabela 4.1: Atividade 1

Hobby	Número de entrevistados
Praticar esporte	8
Música	6
Patinação	3
Dança	7
Aeromodelismo	1
Total	25

A partir dessa tabela, os alunos deverão construir gráficos, calcular a moda e saber calcular a porcentagem de entrevistados que prefere praticar esporte, música, patinação, dança ou aeromodelismo. A complementação da tabela com as frequências das respectivas porcentagens deve ser feita. Nesse momento o professor deve instigar os alunos a descobrir qual deve ser a porcentagem de cada participante da pesquisa, fazendo-os entender que as 25 pessoas participantes correspondem à 100% dos entrevistados e que, portanto, cada um deles corresponde a $\frac{100}{25} = 4\%$. Dessa forma, espera-se que os alunos cheguem a seguinte tabela:

Tabela 4.2: Atividade 1

Hobby	Número de entrevistados	Porcentagem
Praticar esporte	8	32%
Música	6	24%
Patinação	3	12%
Dança	7	28%
Aeromodelismo	1	4%
Total	25	100%

O conceito de moda deve estar claro para os alunos, e eles não devem sentir dificuldade em enxergar em uma tabela qual a variável em que a porcentagem é maior. Assim, espera-se que eles saibam que "praticar esporte" é a atividade mais realizada entre os entrevistados. A moda também fica evidente ao olharmos para um gráfico do tipo barras. Nesse momento o professor deve orientar os alunos na construção desse tipo de gráfico, utilizando para tal, ou a quantidade de pessoas em cada categoria ou a porcentagem correspondente.

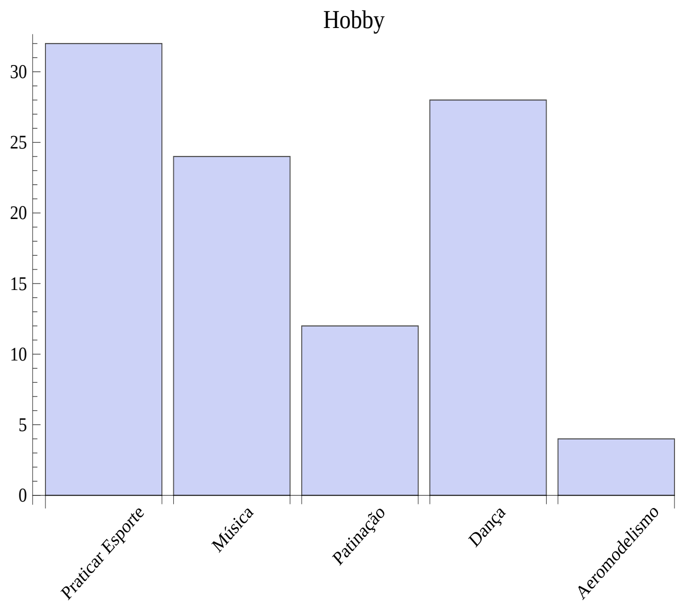

Figura 4.1: Atividade: hobby - porcentagem correspondente de cada atividade

4.2 Atividade 2 - Lançamento de moeda

Objetivo: organizar dados e informações, identificar regularidades e criar regras a partir destas observações.

Desenvolvimento: os alunos devem registrar os resultados obtidos no lançamento de uma moeda e posteriormente chegar a conclusões sobre a regularidade obtida quando, cada vez mais, aumenta-se a quantidade de lançamantos da moeda.

A atividade pode ser desenvolvida em grupo, com o intuito de que os alunos discutam os questionamentos e reconheçam juntos o caráter aleatório de fenômenos e eventos. Eles devem observar na prática que a probabilidade de ocorrência da face cara e da face coroa no lançamento de uma moeda é de $\frac{1}{2}$ para cada, ou seja, 50% de chances de ocorrer cara e 50% de chances de ocorrer coroa.

Com uma moeda em mãos, os alunos devem responder aos seguintes questionamentos:

(a) Sem lançar a moeda, qual o resultado que vocês esperam para 50 lançamentos dessa moeda?

(b) Lancem a moeda 50 vezes, registrem os resultados em uma tabela e em seguida calculem a soma dos resultados de cada face. Registrem esses valores percentualmente.

(c) Os resultados obtidos no item (b) foram os que vocês registraram no item (a)?

(d) Repitam a experiência jogando a mesma moeda, num mesmo local e com a mesma intensidade de força. Registrem os resultados e comparem com os resultados do item (b). Os resultados foram os mesmos?

(e) Se uma moeda for lançada 500 vezes e em todos os lançamentos sairem coroa, o que vocês podem imaginar a respeito dessa moeda?

Com o item (a), espera-se que os alunos tentem alguns resultados. Nesse momento é importante deixar claro que esses resultados não passam de meras especulações, uma vez que trata-se de um evento que ainda não ocorreu. Com o experimento feito no item (b), os alunos devem perceber que o resultado pode não ser exatamente 50% para cada face. Indague-os sobre esse fato. Faço-os perceber que uma probabilidade nada mais é do que uma previsão para um evento futuro. Neste hora o professor pode desafiá-los a repetir o experimento com 100, 200, 300 lançamentos. Espera-se com isso que os resultados se aproximem cada vez mais do esperado, ou seja, que o percentual de registros de cada face se aproxime cada vez mais de 50% à medida que a quantidade de lançamentos aumente. Os itens (c) e (d) devem reforçar a idéia da aleatoriedade de um evento. O item (e) deve induzí-los a entender que a aleatoriedade de um evento deve-se à honestidade dos materiais envolvidos no processo.

4.3 Atividade 3 - Lançamento de dado

Objetivo: observar a regularidade no lançamento de um dado a partir dos dados de uma tabela de frequências.

Desenvolvimento: a partir das informações de uma tabela sobre o lançamento de um dado, os alunos devem encontrar os percentuais relativos a cada uma das faces do dado e em seguida tentar identificar se o dado é ou não honesto.

Considere a seguinte tabela onde foram anotados os resultados obtidos após 1000 lançamentos de um dado:

Tabela 4.3: Atividade 3

Face	Número de vezes
1	158
2	170
3	163
4	172
5	166
6	171
Total	1000

De posse da tabela, o professor deve pedir para que seus alunos:

(a) Expressem os resultados de cada uma das faces em porcentagem.

(b) Decidam, em conjunto, se o referido dado é ou não honesto.

A partir do conhecimento já adquirido pelos alunos, o professor deve indagá-los sobre qual deve ser a probabilidade de ocorrência de cada uma das faces do dado, fazendo-os chegar a conclusão de que cada face deve sair aproximadamente 166 vezes. Diante disso, espera-se que os alunos observem que os valores encontrados no item (a) sejam parecidos ou próximos do resultado esperado de $16,6\%$, e dessa forma concluam que o dado é honesto.

4.4 Atividade 4 - Medidas de centralidade e de dispersão

Objetivo: Entender o conceito e a aplicação das medidas centrais e de dispersão.

Desenvolvimento: Os alunos calculam cada uma das medidas centrais e de dispersão e em seguida interpretam seus resultados.

A atividade é a seguinte: Em um treinamento de salto em altura, três atletas realizaram quatro saltos cada um, de acordo com os resultados da tabela abaixo. Trabalhando em duplas, os alunos devem calcular a média, a moda, a mediana, a variância e o desvio padrão de cada um dos atletas.

Tabela 4.4: Atividade 4

Atleta	1º salto	2º salto	3º salto	4º salto
Atleta A	158cm	170cm	165cm	157cm
Atleta B	173cm	167cm	150cm	160cm
Atleta C	168cm	162cm	165cm	155cm

A média aritmética é uma das medidas de centralidade. Ela resulta da divisão entre a soma dos números de uma lista e a quantidade de números somados. A discussão sobre a moda deve ser colocada pelo professor para que os alunos percebam que não existe um valor que se repete para os saltos de cada atleta. É importante que fique claro que nem sempre ela existirá, e que nesses casos a amostra é amodal. No exemplo em questão nenhum dos atletas têm moda. Para o cálculo da mediana, os alunos devem entender a necessidade de ordenar os dados, uma vez que a mediana deve ser vista como o valor que separa a metade maior e a metade menor de uma amostra, uma população ou uma distribuição de probabilidade. Quanto a variância e o desvio padrão, inicialmente, é interessante que o professor não trabalhe com fórmulas, elas podem parecer muito complicadas em um primeiro instante. Os alunos devem entender que a variância e o desvio padrão são medidas de dispersão que mostram o quão distante cada valor desse conjunto está do valor central (média). É importante nesse momento introduzir o conceito de desvio médio.

Atleta A:

$$\text{Média aritmética} = \frac{158 + 170 + 165 + 157}{4} = \frac{650}{4} = 162,5cm$$

Para o cálculo da mediana, colocando os dados em ordem crescente temos

Tabela 4.5: Saltos do Atleta A

Atleta A	157cm	158cm	165cm	170cm

$$\text{Mediana} = \frac{158 + 165}{2} = \frac{323}{2} = 161,5cm$$

Para o cálculo da variância, o professor deve pedir aos alunos para calcular o quanto cada um dos saltos do Atleta A está desviando-se da média, fazendo-os entender que a variância é uma medida que expressa a homogeniedade dos valores observados, ou seja, que quanto maior for a variação desses valores em torno da média, maior será a variância e, consequentemente, menos homogêneo será o conjunto de dados observados. Dessa forma, espera-se que os alunos encontrem os seguintes desvios:

- $157 - 162,5 = -5,5cm$
- $158 - 162,5 = -4,5cm$
- $165 - 162,5 = 2,5cm$
- $170 - 162,5 = 7,5cm$

O professor deve pedir aos alunos que somem os resultados econtrados nos desvios acima. Ou seja, $-5,5 - 4,5 + 2,5 + 7,5 = 0$. Eles devem questionar se o resultado sempre será igual a zero. Mostre que essa é uma importante propriedade da média aritmética. Nesse momento, o professor deverá definir então a variância como a média dos desvios médios elevados ao quadrado. Para o Atleta A, temos

$$\text{Variância} = \frac{(-5,5)^2 + (-4,5)^2 + (2,5)^2 + (7,5)^2}{4} = 28,25$$

Porém, como não é possível expressar a variância na mesma unidade de medida dos valores da variável, uma vez que os desvios são elevados ao quadrado, define-se a medida de dispersão chamada *Desvio Padrão* como a raiz quadrada da variância. Portanto, para o Atleta A temos

$$\text{Desvio Padrão} = \sqrt{28,25} \simeq 5,31$$

Mais uma vez, é relevante ressaltar que, quanto menor for o Desvio Padrão, mais homogêneo é o conjunto de dados. Dando continuidade ao processo, o professor deverá pedir aos alunos que façam o mesmo processo para o Atleta B e o Atleta C. As respostas esperadas são:

Atleta B:

$$\text{Média Aritmética} = \frac{173 + 167 + 150 + 160}{4} = \frac{650}{4} = 162,5cm$$

Para o cálculo da mediana, colocando os dados em ordem crescente temos

Tabela 4.6: Saltos do Atleta B

Atleta B	150cm	160cm	167cm	173cm

$$\text{Mediana} = \frac{160 + 167}{2} = \frac{327}{2} = 163,5cm$$

Os desvios médios serão:

- $150 - 162,5 = -12,5cm$
- $160 - 162,5 = -2,5cm$
- $167 - 162,5 = 4,5cm$
- $173 - 162,5 = 10,5cm$

Logo,

$$\text{Variância} = \frac{(-12,5)^2 + (-2,5)^2 + (4,5)^2 + (10,5)^2}{4} = 73,25$$

Portanto, o desvio padrão será

$$\text{Desvio Padrão} = \sqrt{73,25} \simeq 8,56$$

Atleta C:

$$\text{Média Aritmética} = \frac{168 + 162 + 165 + 155}{4} = \frac{650}{4} = 162,5cm$$

Para o cálculo da mediana, colocando os dados em ordem crescente temos

Tabela 4.7: Saltos do Atleta C

Atleta B	155cm	162cm	165cm	168cm

$$\text{Mediana} = \frac{162 + 165}{2} = \frac{327}{2} = 163,5cm$$

Os desvios médios serão:

- $155 - 162,5 = -7,5cm$
- $162 - 162,5 = -0,5cm$
- $165 - 162,5 = 2,5cm$
- $168 - 162,5 = 5,5cm$

Logo,

$$\text{Variância} = \frac{(-7,5)^2 + (-0,5)^2 + (2,5)^2 + (5,5)^2}{4} = 23,25$$

Portanto, o desvio padrão será

$$\text{Desvio Padrão} = \sqrt{23,25} \simeq 4,82$$

Ao final de todo esse processo, o professor poderá pedir para que os alunos façam uma tabela contendo todos os resultados encontrados. A Tabela 4.8 mostra esses resultados.

Tabela 4.8: Resultados

Atleta	Média	Moda	Mediana	Variância	Desvio Padrão
Atleta A	162,5cm	Amodal	161,5cm	28,25	5,31
Atleta B	162,5cm	Amodal	163,5cm	73,25	8,56
Atleta C	162,5cm	Amodal	163,5cm	23,25	4,82

De posse da tabela uma análise dos resultados se faz necessária. Os três atletas possuem mesma média, o que poderia levar os alunos a um raciocínio errôneo de que os três atletas têm, no geral, mesmo desempenho. O professor deve chamar a atenção, mais uma vez, sobre a importância da variância e do desvio padrão. Os alunos devem entender que uma menor variação dos dados significa maior homogeneidade e, assim, concluirem que o atleta mais regular é o Atleta C.

Por fim, vale mais uma vez ressaltar a importância da Estatística, não só pelas recomendações curriculares do MEC, mas também para a sociedade como um todo, que lida quase que diariamente com as informações disponibiizadas pelos meios de comunicação em forma de gráficos e tabelas. É inegável a crescente importância de se trabalhar os conceitos de Estatística no ensino básico, e os Parâmetros Curriculares Nacionais são os grandes propulsores na motivação do professor para que em suas aulas de matemática sejam trabalhados conteúdos de Estatística com o objetivo de desenvolver nos alunos sua capacidade de lidar com as situações e as informações que estão disponíveis no nosso dia a dia. Assim, é necessário que os professores se conscientizem que o ensino da Estatística é de suma importância para o desenvolvimento de cidadãos críticos.

4.5 Atividade 5 - Introduzindo os conceitos de Função de Probabilidade e Função de Distribuição Acumulada

Como vimos, as orientações dos Parâmetros Curriculares Nacionais para o ensino de estatística no Ensino Médio são bastante limitadas, embora alguns conceitos tais como, esperânça, sobrevivência, taxa de falha, momentos, função densidade de probabilidade, função de distribuição acumulada, entre outras, possam ser introduzidos nesse nível de ensino sem muitos problemas. Vejamos um exemplo de como introduzir os conceitos de Função de Probabilidade e Função de Distribuição Acumulada.

Considere o lançamento de duas moedas, em sequência, de forma independente. A pergunta é: qual a probabilidade de sair cara em pelo menos uma moeda? O professor deve começar descrevendo o Espaço Amostral

$$\Omega = \{CC, CK, KC, KK\},$$

onde C:"cara"e K:"coroa". É óbvio que, em três das quatro possibilidades que temos no Espaço Amostral temos pelo menos uma cara, o que nos daria uma probabiliade de 75%. Mas esse não é o objetivo desse exemplo.

Vamos chamar de X :"**número de caras que saem nos dois lançamentos**". X é uma varável aleatória que pode assumir os seguintes valores:

$$\begin{cases} 0, & \text{caso saia "KK";} \\ 1, & \text{caso saiam "CK"ou "KC";} \\ 2, & \text{caso saia "CC".} \end{cases}$$

Já conhecendo a definição de probabilidade, o professor deverá mostrar que

- $P(X = 0) = \dfrac{1}{4} = 0,25$
- $P(X = 1) = \dfrac{2}{4} = 0,5$
- $P(X = 2) = \dfrac{1}{4} = 0,25$

e assim mostrar aos alunos que a probabilidade de sair pelo menos uma vez cara é

$$P(X \geq 1) = P(X = 1) + P(X = 2) = 0,5 + 0,25 = 0,75 = 75\%.$$

Esse raciocínio é comum. Instigar os alunos na busca de outras formas de resolução deve sempre ser feito. Uma outra maneira para responder a pergunta acima seria

$$P(X \geq 1) = 1 - P(X = 0),$$

ou seja

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0,25 = 0,75 = 75\%.$$

Nessa hora os alunos devem estar preparados para entender o que são a função de probabilidade e a função de distribuição acumulada, entender que essas funções caracterizam uma variável aleatória. Nesse exemplo temos que nossa variável aleatória é discreta, conceito já trabalhado no Ensino Médio. Defina função de probabilidade $f(x)$ como a função que associa a cada valor de X sua probabilidade de ocorrer, ou seja

$$f(x) : x_i \mapsto P(X = x_i),$$

onde $i = 0, 1, 2$.

Do estudo das probabilidades no Ensino Médio, já sabemos que a soma das probabilidades de um evento sempre dará 1. Daí, duas propriedades podem ser introduzidas

(i) $0 \leq P(X = x_i) \leq 1$

(ii) $\displaystyle\sum_{i=0}^{n} P(X = x_i) = 1$

No nosso exemplo $n = 3$ e, portanto, da propriedade (ii), temos

$$\sum_{i=0}^{3} P(X = x_i) = P(X = x_0) + P(X = x_1) + P(X = x_2) = 0,25 + 0,5 + 0,25 = 1.$$

A definição da função de distribuição acumulada $F(x)$ é imediata

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

FDP e FDA para variáveis aleatórias contínuas:

O seguinte roteiro servirá de base para a introdução das definições da função densidade de probabilidade e da função de distribuição acumulada no Ensino Médio. Nesse exemplo, nossa variável aleatória X será o tempo gasto por um estudante para respoder uma prova do ENEM. Suponha que o primeiro estudante a entregar a prova saia depois de 2 horas e 30 minutos e que o último estudante saia com 4 horas de prova. Assim, nós já estabelecemos o tempo mínimo e o tempo máximo que os estudantes gastaram na prova. Nesse caso, os valores que podem ser assumidos por nossa variável aleatória X estão entre o valor máximo e mínimo, ou seja, a variável é contínua. É importante que os alunos tenham consciência que quando estivermos trabalhando com uma variável aleatória contínua torna-se impossível calcularmos probabilidades de valores fixos, considerando então, intervalos.

O que queremos deixar claro aos alunos é que seria praticamente impossível calcular a probabilidade de um estudante terminar a prova do ENEM depois de 3 horas e vinte minutos, ou seja, calcular $P(X = 3:20)$. Agora, torna-se necessário o cálculo de probabilidades de intervalos, como por exemplo, calcular a probabilidade de um estudante terminar a prova entre 3 horas e 10 minutos e 3 horas e 15 minutos, ou seja

$$P(3:10 < X < 3:15).$$

Nesses casos a função de probabilidade que caracteriza nossa variável aleatória contínua X chama-se função densidade de probabilidade. O gráfico dessa função é uma curva.

O gráfico 4.2 é da função densidade de probabilidade da distribuição Weibull dada por (2.3). Fizemos $\alpha = 2$ e $\beta = 1$, donde temos que

$$f(x) = 2x\, e^{-x^2}$$

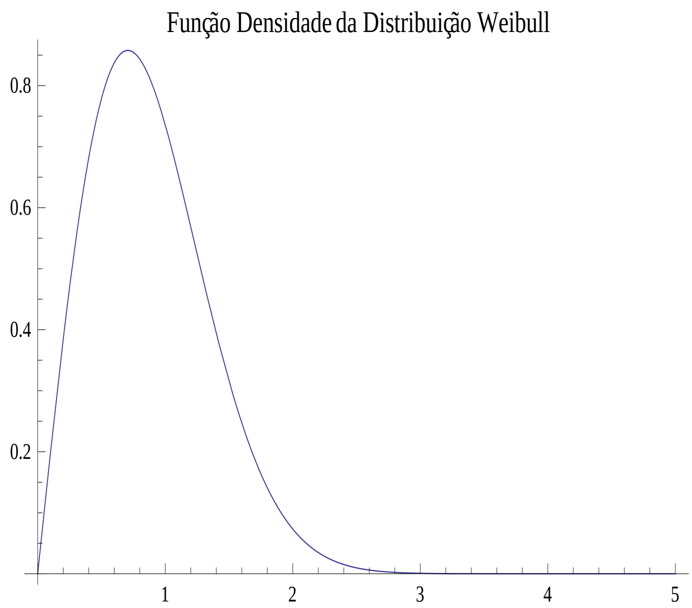

Figura 4.2: fdp Weibull $\alpha = 2$ e $\beta = 1$

Podemos destacar no gráfico que:

- $f(x) \geq 0$ para todo valor real de x.
- A área total abaixo da curva do gráfico é igual a 1, ou seja, a probabilidade máxima possível é 1.
- Para quaisquer valores de a e b, $P(a < X < b)$ é igual a àrea abaixo do gráfico de $f(x)$ delimitada por esse intervalo.

A definição para a função de distribuição acumulada continua a mesma para as variáveis aleatórias contínuas, ou seja, a FDA $F(x)$ determina $P(X \leq x)$. Para esses cálculos precisamos usar integração, o que não vem ao caso no Ensino Médio.

Capítulo 5

Considerações finais

Neste trabalho, fizemos um tratamento matemático para a nova distribuição Zeta-Weibull que inclui a função densidade de probabilidade, a função de distribuição acumulada, a funçao de sobrevivência, a função taxa de falha, a função quantílica, a função geratriz de momentos, a entropia de Rényi e a confiabilidade. Além disso, escrevemos a função densidade de probabilidade da distribuição Zeta-Weibull como combinação linear de densidades Weibull, o que se faz muito importante uma vez que propriedades matemáticas já conhecidas da distribuição Weibull podem ser aplicadas aos cálculos da nova distribuição. Também escrevemos a fdp e a fda da distribuição Zeta-Weibull como combinação de densidades exponencializadas. A estimação dos parâmetros foi feita atravéz do método da máxima verossimilhança, utilizando para isso o pacote bbmle, de autoria de Benjamin Bolker, do software R(versão 3.3.3).

Variando alguns parâmetros das funções densidade, sobrevivência e taxa de falha, encontramos seus gráficos. As curvas da função taxa de falha apresentaram as formas crescente, decrescente e, considerando a Figura 3.9, uma curva que decresce e, a partir de determinado ponto, cresce. Analisamos o conjunto de dados reais encontrado em [17] com a distribuição Zeta-Weibull, comparando-a com as distribuições Zeta-Rayleigh, Zeta Gumbel, Zeta-Fréchet, Weibull Estendida e Beta Weibull Poisson. A Zeta-Weibull se ajustou melhor aos dados o que faz dessa nova distribuição uma ferramenta competitiva para análise de dados de sobrevivência, comparando-a com outras distribuições existentes na literatura.

No capítulo 4 apresentamos sugestões de atividades para o ensino de Estatística no Ensino Médio, dentro do que propõe os PCNEN [1], onde procuramos abordar os conteúdos sugeridos quase que em sua totalidade para que, dessa forma, a leitura servisse de norte para profissionais que atendem nesse nível de ensino.

O Anexo contém um modelo de script para a construção dos gráficos utilizando o software Mathemática (versão 9.0.1.0). Para o software R(versão 3.3.3), um modelo de script para a otimização dos parâmetros e um modelo para encontrar os critérios de informação.

Anexos

Script para a construção do gráfico 3.3

shadowbox[legend] :=

Graphics[{{Black, Rectangle[0.05, −0.05, 1.05, 0.95]}, White, EdgeForm[Gray], Rectangle[ ],

Inset[legend,0.5, 0.5, Center]}, ImageSize− > 78]

Framed[Plot[Evaluate@Table[4 ∗ 0.5 ∗ $x^{(4-1)}$ ∗ PolyLog[s − 1, Exp[−0.5 ∗ x^4]]

/Zeta[s], s, 1.1, 1.2, 1.4, 1.8], {x, 0, 5}, Frame → True, Filling → None, FrameLabel →

{x, Densidade}, PlotStyle → {{Red, DotDashed}, {Green, Dashed}, {Orange, Thickness},

{Blue, Dotted}}, Background → Lighter[Blend[{Yellow, Orange}], 0.97],

PlotLabel → "Densidade da Distribuição Zeta-Weibull", PlotRange → {{0, 2}, {0, 2}},

PlotLegends →

Placed[LineLegend[{Directive[Red, DotDashed], Directive[Green, Dashing[Medium]],

Directive[Orange, Thickness], Directive[Blue, Dotted]},

{"s = 1.1", "s = 1.2", "s = 1.4", "s = 1.8"}, LegendFunction → shadowbox,

LegendLayout → "Column"], {{0.9, 1}, {0.7, 1.1}}]]]

Script para a otimização dos parâmetros da distribuição Zeta-Weibull aplicada aos dados de [17]

require(VGAM)

zetaw< −function(x,s,alfa,beta){

$fx <-(zeta(s) \wedge (-1)) * alfa * beta * x \wedge (alfa-1) * exp(-beta * x \wedge (alfa))$

$*lerch(exp(-beta * x \wedge (alfa)), s-1, 1)$

return(fx)

}

lfr < − function(parametro) {

x < −sort(c(2.160, 0.746, 0.402, 0.954, 0.491, 6.560, 4.992, 0.347,

0.150, 0.358, 0.101, 1.359, 3.465, 1.060, 0.614, 1.921,

4.082, 0.199, 0.605, 0.273, 0.070, 0.062, 5.320))

s < − parametro[1]

alfa < − parametro[2]
beta < − parametro[3]
sum(log(zetaw(x,s,alfa,beta)))
}
chute< −c(5.0045,1.0213,0.0069)
lfr(chute)
optim(chute, lfr,method="BFGS",control=list(fnscale=-1))

Script para encontrar os critérios de informação para a distribuição Zeta-Weibull aplicada aos dados de [17]

require(VGAM)
require(bbmle)
$x <$ −sort(c(2.160, 0.746, 0.402, 0.954, 0.491, 6.560, 4.992, 0.347,
0.150, 0.358, 0.101, 1.359, 3.465, 1.060, 0.614, 1.921,
4.082, 0.199, 0.605, 0.273, 0.070, 0.062, 5.320))
zetaw< −function(x,s,alfa,beta){
$fx < -(zeta(s) \wedge (-1)) * alfa * beta * x \wedge (alfa - 1) * exp(-beta * x \wedge (alfa))$
$*lerch(exp(-beta * x \wedge (alfa)), s - 1, 1)$
return(fx) }
lfr < − function(s,alfa,beta) { -sum(log(zetaw(x,s,alfa,beta))) }
max.zetaw < − mle2(lfr, start=list(s= 10.6115,alfa= 0.9490,beta= 0.0270),
method="BFGS",skip.hessian=FALSE,use.ginv=TRUE,gr=NULL)
logLik(max.zetaw)
AIC:
AIC(max.zetaw)
BIC:
AIC(max.zetaw,k=log(213))
HQIC:
AIC(max.zetaw,k= 2*log(log(213)))
AIC CORRIGIDO:
AIC(max.zetaw,k=426/209)

Referências Bibliográficas

[1] BRASIL. Ministério da Educação. Secretaria de Educação Média e Tecnológica. *Parâmetros Curriculares Nacionais: Ensino Médio.* Brasília, 1999.

[2] FARIAS,Ana Maria Lima de. *Variáveis aleatórias contínuas.* Dísponível em: http://www.est.uff.br/images/ArqGET/EnsMatDidat/get00118_I-0.pdf. Acesso em 13 de jul. 2017.

[3] BERLINGHOFF, W.P.; GOUVÊA, F.Q. *A matemática através dos tempos: um guia fácil e prático para professores e entusiastas*, tradução; Elza Gomide, Helena Castro, 2 Ed. São Paulo, Blucher, 2010.

[4] CÉSAR, Kelly Araújo.*Análise estatística de sobrevivência: um estudo com pacientes com câncer de mama.* 2005. 12f. Monografia(Graduação) - Universidade Católica de Brasília, Brasília, 2005.

[5] EVES, Howard. *Introdução à histótia da matemática*, tradução: Hygino H. Domingues. Campinas, SP: Editora da UNICAMP, 2004.

[6] FREITAS, C.C; RENNÓ, C.D; JUNIOR, M.A.S. *Estatística - Curso 1.* São José dos Campos, INPE, 2013

[7] JAMES, B.R. *Probabilidade: um curso em nível intermediário.* Projeto Euclides, 3.ed. Rio de Janeiro, IMPA, 2013.

[8] JUNIOR, P.L.; SILVEIRA, F.L; OSTERMANN, F. *Análise de sobrevivência aplicada ao estudo do fluxo escolar nos cursos de graduação em física: um exemplo de uma universidade brasileira.* Revista Brasil.eira de Ensino de Física, V.34, n.1, 2012.

[9] LOPES, C.A.E. *Experiências e Perspectivas do Ensino da Estatística - Desafios para o século XXI.* Florianópolis, p.167-174, setembro.1999.

[10] MAGALHÃES, M.N; LIMA, A.C.P. *Noções de Probabilidade e Estatística*, 7 Ed. São Paulo, EDUSP, 2013.

[11] MORGADO, A.C; CARVALHO, P.C.P. *Matemática Discreta*, 1 Ed. Rio de Janeiro, SBM, 2013.

[12] MUNIZ NETO, Antonio Caminha. *Fundamentos de Cálculo*, 1 Ed. Rio de Janeiro, SBM, 2015.

[13] NAVARRO,M.C.K.A.; AGUILERA-NAVARRO,V.C.; FERREIRA,R.C.; TERAMON,N. *A função zeta de Riemann.*Revista Ciências Exatas e Naturais, V.1,n.1,p.23-47,1999.

[14] SPIEGEL, Murray Ralph. *Probabilidade e Estatística*, tradução: Alfredo Alves de Farias, São Paulo, SP, Pearson Education do Brasil, 1978.

[15] PAIXÃO, A.C.P.da. *New extended lifetime distributions.*2014. 115f. Tese(Doutorado)-Universidade Federal de Pernanbuco. Recife, 2014.

[16] SANTOS, Rosilda Sousa. *Estudo sobre algumas Famílias de Distribuições de Probabilidades Generalizadas.* 2012. 102f. Dissertação(Mestrado) - Universidade Federal de Campina Grande, Campina Grande, 2012.

[17] SUPRAWHARDANA, M.S; PRAYOTO; SANGADJI. *Total Time On Test Plot Analysis For Mechanical Components Of The RSG-GAS Reactor.* Atom Indones, V.25, n.2, 1999.

[18] ROSS, Sheldon. *Probabilidade: um curso moderno com aplicações*, tradução; Alberto Resende de Conti, 8 Ed. São Paulo, Bookman, 2010.