

Felipe Ricardo Santos de Gusmão

**Um estudo sobre a teoria da identificabilidade paramétrica
e suas aplicações em novas distribuições e classes de
distribuições de probabilidade**

Recife

6 de maio de 2024



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**Um estudo sobre a teoria da identificabilidade paramétrica e suas aplicações em
novas distribuições e classes de distribuições de probabilidade**

Tese julgada adequada para obtenção
do título de Doutor em Biometria e Es-
tatística Aplicada, defendida e aprovada
por unanimidade em 20/02/2024 pela
comissão examinadora

**Área de concentração: Biometria e
Estatística Aplicada**

**Orientador: Dr. Frank Sinatra Gomes
da Silva**
**Coorientador: Dr. Cícero Carlos Ramos
de Brito**

Recife

6 de maio de 2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- G982e Gusmão, Felipe Ricardo Santos de
Um estudo sobre a teoria da identificabilidade paramétrica e suas aplicações em novas distribuições e classes de distribuições de probabilidade/ Felipe Ricardo Santos de Gusmão. – 2024.
69 p. : f.
- Orientador: Dr. Frank Sinatra Gomes da Silva
Coorientador: Dr. Cícero Carlos Ramos de Brito
inclui referências.
- Tese (doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, 2024.
1. Probabilidade. 2. Distribuição. 3. Misturas. 4. Identificabilidade. 5. Classes. I. Silva, Frank Sinatra Gomes da, orient. II. Brito, Cícero Carlos Ramos de, coorient. III. Título

CDD 519.5

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**Um estudo sobre a teoria da identificabilidade paramétrica e suas aplicações em
novas distribuições e classes de distribuições de probabilidade**

Felipe Ricardo Santos de Gusmão

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 20/02/2024 pela comissão examinadora

Orientador:

Dr. Frank Sinatra Gomes da Silva
Orientador

Banca examinadora:

**Dr. Abraão David Costa do
Nascimento**
Universidade Federal de Pernambuco

**Dr. Thiago Alexandre Nascimento de
Andrade**
Universidade Federal de Santa Maria

Dr. Cícero Carlos Ramos de Brito
Instituto Federal de Educação

PhD Leandro Chaves Rêgo
Universidade Federal do Ceará

*Este trabalho é dedicado à todos
que sonham com um dia melhor.*

Agradecimentos

Agradeço a Deus por esta oportunidade de ascensão intelectual, agradeço pelo amparo nos momentos de dificuldades e por toda ajuda que acredito ter recebido dele.

Agradeço a minha mãe Sônia Santos por todo esforço que ela fez durante toda minha vida para que este momento fosse possível.

Agradeço a minha esposa Pâmella Teixeira pelo amor, carinho, companheirismo, afinidade, apoio, incentivo e pelo desejo de ver a realização deste sonho que é de nós dois.

Agradeço a minha irmã Rafaela e ao meu cunhado Rosivaldo pela torcida e pelo apoio dado a mim e a minha família nas situações adversas que vivemos.

Agradeço a Marco, meu amigo de longa data, por ter lembrado de mim no momento em que eu mesmo já não lembrava de quem eu era.

Agradeço ao meu orientador Dr. Frank Sinatra Gomes da Silva e ao meu coorientador Dr. Cícero Carlos Ramos de Brito pela amizade, pela paciência, pelo incentivo, pela preocupação e pelo desejo de ver a mim com o título de doutor.

Possuo eterna gratidão a todos os meus professores por serem responsáveis pela pessoa, pelo cidadão e pelo acadêmico que sou hoje.

Não posso esquecer de agradecer a todos membros do Biometria e Estatística Aplicada e aos membros da pré-banca e da banca final de avaliação pelas contribuições e sugestões dadas para melhoria da qualidade deste trabalho.

*“O primeiro pecado da humanidade foi a fé;
a primeira virtude foi a dúvida.”
(Carl Sagan)*

Resumo

A proposição de novas classes de distribuições ou de funções de distribuição constituem generalizações de distribuições ou composições entre distribuições já descritas na literatura estatística. É sabido que a adição de parâmetros a funções de distribuição pode levar a problemas de identificabilidade e consequentemente trazer complicações para a estimação de parâmetros. No trabalho de [Cole \(2020\)](#) é citado um exemplo clássico de problema por identificabilidade, tal problema consiste de dois parâmetros que estão dispostos em uma fd como um produto e por isto não podem ser estimados individualmente, apenas o produto dos parâmetros pode ser estimado. Neste trabalho, foi apresentado definições, teoremas, lemas e proposições que são contribuições desta tese para a teoria da identificabilidade. Apresentadas e estudadas em termos da identificabilidade misturas de funções de distribuição aparecem como resultados das operações de adição e multiplicação. A classe de distribuições T-**G**-X [*Method for generating distributions and classes of probability distributions: the univariate case*, [Brito et al. \(2019\)](#)] foi exibida e discutida a identificabilidade de alguns subcasos da mesma. A T-**G**-X constitui uma extensão multi-baseline, por isto o **G** em negrito, da bem conhecida classe T-X. Cada G que compõem o vetor de baselines **G** é um função de distribuição univariada.

Palavras-chaves: Identificabilidade. Probabilidade. Distribuição. Misturas. Classes.

Abstract

The proposition of new classes of distributions or distribution functions constitute generalizations of distributions or compositions between distributions already described in the statistical literature. It is known that adding parameters to distribution functions can lead to problems of identifiability and consequently bring complications to parameter estimation. In the work of [Cole \(2020\)](#) a classic example of problem by identifiability is cited, this problem consists of two parameters that are arranged in a *fd* as a product and therefore cannot be estimated individually, only the product of the parameters can be estimated. In this work, definitions, theorems, lemmas and propositions which are contributions of this thesis for the theory of identifiability. Presented and studied in terms of identifiability mixtures of distribution functions appear as results of addition and multiplication operations. The class of distributions T- $\mathbf{G}$ -X [*Method for generating distributions and classes of probability distributions: the univariate case*, [Brito et al. \(2019\)](#)] was displayed and the identifiability of some subcases of it was discussed. T- $\mathbf{G}$ -X constitutes a multi-baseline extension, hence the $\mathbf{G}$ in black, of the well-known T-X class. Each $\mathbf{G}$ that makes up the baseline vector $\mathbf{G}$ is a distribution function univariate.

Key-words: Identifiability. Probability. Distribution. Mixture. Class.

Lista de tabelas

Tabela 1 – Prova que $\Theta_{\mathcal{N}}$ é identificável	37
Tabela 2 – Prova de que $\Theta_{\mathcal{F}}$ é identificável.	46

Sumário

1	Introdução	1
1.1	Contextualização do problema da Tese	1
1.2	Delineamento dos objetivos da Tese	3
1.3	Contribuições da Tese	4
2	Revisão da Literatura	10
2.1	Motivação para o estudo da identificabilidade	10
2.2	Teoria geral da identificabilidade	12
2.2.1	Relações binárias	12
2.2.2	Identificabilidade em conjuntos e funções	12
2.2.3	Identificabilidade em modelos estatísticos	13
2.2.3.1	Modelos paramétricos	14
2.3	Misturas de distribuições definidas por Henry Teicher	15
2.3.1	Misturas de famílias aditivamente fechadas	16
2.3.2	Identificabilidade em misturas finitas	18
2.4	A família T-G-X de distribuições de probabilidades	20
2.4.1	Método gerador distribuições e classes de distribuições de probabilidade no caso univariado T-G-X	20
2.4.1.1	Família T-X de distribuições	21
3	Teoria da Identificabilidade Paramétrica	23
3.1	Introdução	23
3.2	Definições, proposições, lemas e teoremas para teoria da identificabilidade	23
3.3	Estudo da identificabilidade nas distribuições <i>EGW</i> e <i>EW</i>	44
3.3.1	A distribuição <i>EGW</i> e um estudo sobre identificabilidade	44
3.3.2	A distribuição <i>EW</i> e um estudo sobre identificabilidade	45
3.3.3	Conclusão do Capítulo 3	46
4	Identificabilidade em Métodos Construtores de <i>fd</i>'s Pertencente a Classes não Finitas	47
4.1	Método construtor de <i>fd</i> a partir da combinação de <i>fd</i> 's generalizadas	47
4.1.1	Método construtor de <i>fd</i> 's pela soma de funções de distribuição generalizadas	48
4.1.2	Método construtor de <i>fd</i> 's pelo produto de potências de <i>fd</i> 's generalizadas	50

4.1.3	Método construtor de fd 's pela soma ponderada de produtórios de fd 's generalizadas	51
4.1.4	Método construtor de fd 's pelo produto de potências de somatórios de fd 's generalizadas ponderadas	52
4.1.5	Método construtor de fd 's pelo somatório ponderado de potências de somatórios ponderados de fd 's generalizadas	53
4.1.6	Método construtor de fd 's pelo somatório ponderado de potências de fd 's generalizadas somado ao produtório de potências de fd 's generalizadas	54
4.1.7	Método construtor de fd 's por sucessivas composições de funções f 's da fd generalizada	54
4.1.8	Conclusão do Capítulo 4	55
5	Um Estudo sobre Identificabilidade da Família T-G-X de Distribuições . . .	56
5.1	Um estudo da identificabilidade em alguns subcasos da família T-G-X de distribuições	56
5.1.1	A classe de distribuições de probabilidade $F_{\mathcal{H}}-(G_1, G_2)$ e um estudo de identificabilidade	56
5.1.2	A classe de distribuições de probabilidade $F_{\mathcal{H}}-G$ e um estudo de identificabilidade	58
5.1.3	A classe $F_{\mathcal{H}}(x)$ -Tangente-G de distribuições de probabilidade e um estudo da identificabilidade	59
5.1.4	Estudo de identificabilidade em casos particulares da Família T-X de distribuições	60
5.1.5	Conclusão do Capítulo 5	61
6	Conclusão	63
7	Trabalhos Futuros	64
	Referências Bibliográficas	65

1 Introdução

1.1 Contextualização do problema da Tese

A Inferência Estatística é fundamentalmente orientada para extrair conclusões a respeito de parâmetros populacionais com base em dados observados. Para esse fim, recorre-se à utilização de distribuições de probabilidade para modelar características específicas da população. Essas distribuições, agrupadas em classes ou famílias ($\mathcal{F}$), em geral não enumeráveis, estabelecem uma relação um-a-um com um espaço vetorial paramétrico, fenômeno conhecido como identificabilidade paramétrica. Desde a década de 1930, a literatura especializada tem sido enriquecida por estudos sobre a Identificabilidade Paramétrica. Destaca-se, por exemplo, o trabalho do economista [Frisch \(1934\)](#), que evidenciou essa questão em sua pesquisa. Anos mais tarde, [Koopmans \(1949\)](#) sublinhou a necessidade de uma distinção clara entre os problemas de inferência estatística, relacionados à variabilidade de amostras finitas, e os problemas de identificabilidade, que surgem quando os estatísticos exploram os limites da inferência.

[Hurwicz \(1950\)](#) e [Koopmans e Reiersol \(1950\)](#) desenvolveram uma teoria geral da identificabilidade paramétrica para modelos estatísticos, formalizando a ideia intuitiva de identificação. Essa concepção intuitiva é abordada na Seção 1.2 de [Koopmans e Reiersol \(1950\)](#). [Silvey \(1975\)](#) ressalta que a não identificabilidade paramétrica ocorre quando parâmetros distintos (no mesmo espaço paramétrico), pertencentes a um mesmo espaço amostral, resultam na mesma distribuição. Definições mais elaboradas de identificabilidade para modelos paramétricos foram apresentadas em trabalhos como [Rothenberg \(1971\)](#), [Hsiao \(1983\)](#) e [Paulino e Pereira \(1994\)](#).

[Jones et al. \(2010\)](#) menciona que uma função de distribuição (F) é não identificável se existirem vetores paramétricos γ_1 e γ_2 , distintos entre si, tais que $F(\gamma_1) = F(\gamma_2)$. Essa ideia é corroborada por [Cole \(2020\)](#). Introduzido como uma condição mais branda, o conceito de identificabilidade local é discutido por [Catchpole e Morgan \(1997\)](#), que definem a existência de uma distância $\delta > 0$, de modo que quaisquer dois conjuntos de parâmetros (produzindo a mesma distribuição) devem ser separados por pelo menos δ para garantir a identificabilidade.

As funções de distribuição podem ser classificadas como globalmente identificáveis, localmente identificáveis ou não identificáveis. Tais classificações são descritas por [Cole](#)

(2020). Uma função de distribuição $F_{\mathcal{F}}$, pertencente à família $\mathcal{F}$, é globalmente identificável se $F_{\mathcal{F}}(x; \boldsymbol{\theta}) = F_{\mathcal{F}}(x; \boldsymbol{\beta})$ para todo $x \in \mathbb{R}$ implicar em $\boldsymbol{\theta} = \boldsymbol{\beta}$. A distribuição é localmente identificável se houver uma vizinhança aberta no espaço vetorial paramétrico $\Theta_{\mathcal{F}}$, na qual a definição de identificabilidade global possa ser aplicada. Se a função de distribuição não for localmente identificável, então ela não será identificável nem global nem localmente. Cole (2020) também define que uma função de distribuição é redundante em parâmetros se for possível expressar $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ apenas como uma função do vetor paramétrico $\boldsymbol{\beta}$, em que $\boldsymbol{\beta} = g(\boldsymbol{\theta})$, com g sendo uma função de $\boldsymbol{\theta}$, e a dimensão de $\boldsymbol{\beta}$ for menor do que a dimensão de $\boldsymbol{\theta}$.

A identificabilidade de misturas de distribuições tem sido objeto de estudo por diversos pesquisadores, porém o problema da não identificabilidade é solúvel apenas para misturas finitas. Tais misturas são aquelas em que as funções de distribuição pertencem a uma classe com cardinalidade finita, ou seja, essas funções de distribuição têm uma relação um-a-um com elementos de um espaço vetorial paramétrico finito. Teicher (1961) foi um dos pioneiros na mistura de funções de distribuição com o objetivo de produzir uma classe identificável. No artigo de Teicher (1960), é apresentado como misturar distribuições através de uma classe de medidas de Lebesgue-Stieltjes, e é definida identificabilidade paramétrica para essa classe. Previamente, Teicher (1954) apresenta condições suficientes e necessárias para que uma família seja aditivamente fechada. Teicher (1961) apresenta um teorema que garante a identificabilidade de misturas de distribuições uniparamétricas e aditivamente fechadas. Finalmente, Teicher (1963) apresenta dois teoremas que garantem identificabilidade para a classe de misturas finitas.

No trabalho de Barndorff-Nielsen (1965), é demonstrada a identificabilidade para misturas finitas de membros da família exponencial. A identificabilidade do produto de medidas é estudada em Teicher (1967), e são apresentados dois teoremas para validar a afirmação de que o produto de medidas finitas é identificável. No artigo de Yakowitz e Spragins (1968), é provado que uma família $\mathcal{F}$ de funções de distribuição induz misturas finitas identificáveis se, e somente se, $\mathcal{F}$ for linearmente independente na extensão sobre o corpo dos números reais.

A identificabilidade em funções de distribuição que contemplam dados circulares foi estudada em artigos como Fraser, Hsu e Walker (1981), que demonstraram a identificabilidade da classe de distribuições de von Mises, e Kent (1983), que apresentou uma ampla classe de misturas finitas identificáveis compostas por distribuições direcionais.

A identificabilidade de misturas finitas é alcançada por uma nova abordagem, que utiliza a transformada - a função geradora de momentos do $\log(X)$ - como discutido por Ahmad (1988). Nesse mesmo estudo, é demonstrado que as misturas finitas das

famílias Weibull, log normal, χ^2 , Pareto e função potência são identificáveis. Além disso, [Chandra \(1977\)](#) investiga o problema das misturas de distribuições de probabilidade em uma estrutura geral, em que tanto o vetor paramétrico quanto o espaço amostral são mensuráveis. O mesmo artigo apresenta uma teoria abrangente para a identificabilidade de uma classe de distribuições mistas. Observações adicionais sobre a não identificabilidade de misturas finitas das distribuições beta e Pearson Tipo VI são feitas em uma nota de [Ahmad e Al-Hussaini \(1982\)](#).

Em [Al-Hussaini e Ahmad \(1981\)](#), foi mostrado que a classe de todas as misturas finitas das famílias de distribuições univariadas, tais como série logarítmica, retangular discreta, retangular, primeira lei de Laplace, χ^2 não central, logística, logística generalizada, secante hiperbólica generalizada, gaussiana inversa e passeio aleatório, são identificáveis. Além disso, no mesmo artigo, foi apresentada uma versão generalizada de um teorema anteriormente proposto por [Teicher \(1963\)](#), mostrando que a classe de todas as misturas finitas das seguintes distribuições multivariadas são identificáveis: binomial negativa, série logarítmica, Poisson, normal, gaussiana inversa e passeio aleatório.

Por meio das condições de cauda das funções características, são obtidos resultados sobre a identificabilidade de misturas finitas com aplicações a famílias de locação e escala com domínio na reta real e para distribuições circulares. [Holzmann, Munk e Stratmann \(2004\)](#) abordam esse tema. Em [Holzmann, Munk e Gneiting \(2006\)](#), é apresentado um lema e três teoremas sobre a identificabilidade de misturas finitas de distribuições elípticas através de funções características.

A identificabilidade em modelos de captura-recaptura para populações fechadas com misturas finitas foi estudada por [Holzmann, Munk e Gneiting \(2006\)](#). Esses modelos são usados por levar em conta a heterogeneidade individual nas probabilidades de captura, mas a validade dessas distribuições foi questionada por [Link \(2003\)](#) com base na não identificabilidade dos referidos modelos.

1.2 Delineamento dos objetivos da Tese

A partir do delineado na Subseção anterior e considerando os estudos fundamentais mencionados, esta subseção destaca os objetivos da presente Tese. Esses objetivos fundamentam as contribuições subsequentemente indicadas e são enumerados da seguinte maneira:

1. Resumir conceitos essenciais sobre identificabilidade paramétrica baseada no trabalho de [Basse e Bojinov \(2020\)](#), identificabilidade em misturas finitas de funções de

distribuição conforme explorado por Henry Teicher ([TEICHER, 1954](#); [TEICHER, 1960](#); [TEICHER, 1961](#); [TEICHER, 1963](#)), e detalhar a família T- $\mathbf{G}$ -X de distribuições com ênfase na subfamília T-X ([ALZAATREH; LEE; FAMOYE, 2013](#); [BRITO et al., 2019](#));

2. Introduzir definições fundamentais sobre a teoria da identificabilidade paramétrica que servirão como alicerces para o desenvolvimento de teoremas, lemas, corolários e proposições ao longo da pesquisa. Esses elementos teóricos serão essenciais para estabelecer critérios de identificabilidade para funções de distribuição;
3. Investigar a identificabilidade de misturas não finitas de funções de distribuição que não sejam uniparamétricas. Este estudo será conduzido através da análise de subcasos derivados de uma forma generalizada de misturas de distribuições, proposta neste trabalho;
4. Propor teoremas que asseguram a identificabilidade para subcasos específicos da família T- $\mathbf{G}$ -X de distribuições. Adicionalmente, examinar cenários nos quais a família T-X de distribuições não é identificável.

À luz dos objetivos supracitados, obtiveram-se as contribuições da presente tese, elencados na Subseção a seguir.

1.3 Contribuições da Tese

Esta tese representa um avanço na teoria da identificabilidade paramétrica, abordando aspectos pouco explorados na literatura estatística. Identificabilidade, sendo crucial para a validação de modelos estatísticos, requer uma compreensão profunda dos parâmetros que podem ser unicamente determinados a partir das distribuições observadas. Neste contexto, esta pesquisa contribui em três áreas específicas, cada uma delas expandindo os horizontes do que é atualmente conhecido e aplicado. Estas áreas incluem:

1. Funções de distribuição definidas sobre a reta real;
2. Misturas não finitas de funções de distribuição não uniparamétricas, também com domínio na reta real;
3. Composições de funções de distribuição utilizando o método T- $\mathbf{G}$ -X, pertencentes à família T- $\mathbf{G}$ -X de distribuições, estendendo-se ao longo da reta real.

Para facilitar uma compreensão abrangente da contribuição desta pesquisa em cada um dos subcampos mencionados, segue-se uma análise detalhada de cada um deles.

Funções de distribuição definidas sobre a reta real

No Capítulo 3 apresenta-se uma contribuição à teoria da identificabilidade paramétrica, fundamentando-se no trabalho seminal de [Basse e Bojinov \(2020\)](#). Uma definição meticulosa de modelo estatístico é apresentada para embasar uma nova formulação, aqui proposta, sobre identificabilidade paramétrica.

A identificabilidade paramétrica, conforme delineada por [Basse e Bojinov \(2020\)](#), aborda a capacidade de estimar os parâmetros de um modelo estatístico de maneira singular e precisa, com base nos dados observados. Esta teoria é de suma importância para assegurar a fidedignidade das inferências estatísticas, adentrando questões essenciais relativas à existência e unicidade das soluções dos parâmetros em contextos estatísticos, enriquecendo, assim, a compreensão acerca da inferência estatística e suas aplicações.

A Definição de Identificabilidade Paramétrica (Definição 9) emprega o conceito de injetividade dentro do espaço vetorial de parâmetros. Tal definição se destaca por sua precisão, comparativamente àquela delineada por [Basse e Bojinov \(2020\)](#), uma vez que utiliza a função de distribuição ao invés de uma medida de probabilidade genérica, estando em consonância com os propósitos deste estudo.

A partir da Definição de Identificabilidade Paramétrica, introduz-se a noção de Função Parte Paramétrica (Definição 11). Essa concepção aborda a decomposição de uma função de distribuição em n outras funções reais. Decorrente desta definição, emerge um teorema (Teorema 4) que assegura a identificabilidade da função de distribuição, contanto que haja injetividade em todas as funções componentes, bem como em suas subcomposições.

Do Teorema 4, deduz-se o Corolário 1, que discorre sobre a mínima redução necessária para alcançar a identificabilidade paramétrica em uma função de distribuição anteriormente não identificável. Além disso, decorrentes diretamente do Teorema 4, são apresentados cinco corolários adicionais ao longo da tese.

Duas definições são propostas: uma que aborda a relação de encaixe entre parâmetros ou vetores de parâmetros (Definição 12), e outra que relaciona parâmetros ou vetores paramétricos por meio da ampliação dos mesmos (Definição 13). A Definição de subfamília paramétrica (Definição 14) descreve como uma família de funções de distribuição pode ser um caso particular de outra família ou dela própria. Como consequência desta definição,

emergem duas proposições. A Definição 19 utiliza a contrapositiva de uma consequência direta da Definição 9 na demonstração de alguns lemas e teoremas ao longo da tese. Algumas outras definições são introduzidas com o intuito de padronizar a terminologia neste estudo.

Dois lemas (Lemas 2 e 3) são apresentados para fundamentar a demonstração do teorema que garante a identificabilidade de subfamílias provenientes de famílias identificáveis (Teorema 5). Em contrapartida ao Teorema 5, propõe-se o Teorema 6, que estabelece que se uma família de funções de distribuição for não identificável, então outra família que a generaliza também será não identificável.

Os últimos quatro teoremas do Capítulo 3 são introduzidos para evidenciar a possibilidade de estudar a identificabilidade de uma família de funções de distribuição por meio da derivação da função de distribuição (Teorema 7) ou pela integração da função densidade de probabilidade (Teorema 8) no caso contínuo. Para o caso discreto, esse estudo pode ser conduzido por meio da função de probabilidade (Teorema 10) ou pela função de distribuição discreta (Teorema 9).

Em um artigo recente, [Gusmão et al. \(2021\)](#) investigaram a identificabilidade das distribuições Weibull exponenciada generalizada ([OGUNTUNDE; ODETUNMIBI; ADEJUMO, 2015](#)) e Weibull exponenciada ([MUDHOLKAR; SRIVASTAVA; FREIMER, 1995](#)). Neste estudo, estabeleceu-se uma relação entre ambas as distribuições, demonstrando que a segunda é um caso particular da primeira. Esta relação foi detalhadamente explorada e, por meio de manipulações algébricas na primeira distribuição (que é não identificável), obteve-se a segunda distribuição (que é identificável). A Seção 3.3 deste trabalho incorpora parte deste estudo que é pertencente ao escopo desta tese.

Misturas não finitas de funções de distribuição não uniparamétricas com domínio na reta real

Nos estudos que abordaram a identificabilidade em misturas de distribuições, como os conduzidos por [Teicher \(1960\)](#), [Teicher \(1961\)](#), [Teicher \(1963\)](#), [Barndoff-Nielsen \(1965\)](#), [Fraser, Hsu e Walker \(1981\)](#), [Kent \(1983\)](#), [Ahmad \(1988\)](#), entre outros mencionados neste contexto, a atenção foi predominantemente direcionada às misturas finitas compostas por funções de distribuição ou por misturas compostas por funções de distribuição não uniparamétricas.

Henry Teicher figura como um dos pioneiros no estudo da identificabilidade em misturas finitas de funções de distribuição. Seus trabalhos, incluindo as contribuições de [Teicher \(1954\)](#), [Teicher \(1960\)](#), [Teicher \(1961\)](#), [Teicher \(1963\)](#), versam sobre questões

cruciais associadas à identificabilidade de misturas finitas de distribuições, especialmente quando se busca assegurar a estimativa única e confiável de parâmetros desconhecidos a partir dos dados observados. Tal linha de pesquisa detém implicações relevantes em diversas esferas da estatística, fornecendo fundamentos teóricos imprescindíveis para a modelagem e análise de dados.

Na presente tese, propom-se uma investigação sobre a identificabilidade de misturas não finitas ou / e não uniparamétricas. Em outras palavras, consideramos funções de distribuição que mantêm uma relação um-a-um com elementos de um espaço vetorial paramétrico não enumerável e contínuo, sob as condições de existência da família de funções de distribuição em questão.

Foi elaborado um método para a geração de misturas não finitas de distribuições de probabilidade, intitulado Método Construtor de Funções de Distribuição, conforme delineado na Seção 4.1. A partir deste método, foram estabelecidos seis submétodos ou subcasos. Para cada um desses, foi formulado um teorema que atesta a não identificabilidade de misturas não finitas compostas por funções de distribuição não uniparamétricas, todas pertencentes à mesma família.

Finalmente, apresenta-se um método gerador denominado Método Construtor de Funções de Distribuição por sucessivas composições de funções de uma função de distribuição generalizada. A identificabilidade deste método pode ser investigada mediante o Teorema 4.

Composições de funções de distribuição na família T-G-X

A família T-G-X de distribuições, introduzida por Brito et al. (2019), é uma classe flexível de distribuições estatísticas que engloba uma ampla variedade de modelos. Essa família tem como subcaso a família T-X de distribuições, proposta por Alzaatreh, Lee e Famoye (2013). Ambas as famílias têm sido amplamente estudadas devido à sua capacidade de modelar uma diversidade de fenômenos em diversas áreas, incluindo ciências naturais, engenharia, economia e muitas outras. O estudo da identificabilidade nessas famílias de distribuições é crucial para garantir a adequação e precisão dos modelos estatísticos utilizados para análise de dados empíricos.

Foi conduzido um estudo sobre a identificabilidade em alguns subcasos da família T-G-X de distribuições. Na Subseção 13, foi introduzida a classe de distribuições de probabilidade $F_{\mathcal{H}}(G_1, G_2)$. Para esta classe, foi proposto o Teorema 17, que afirma a identificabilidade da classe das funções de distribuição $F_{\mathcal{H}}(G_1, G_2)$. Em seguida, é apresentado um exemplo específico dessa classe, abordando "A classe das distribuições de

probabilidade Normal- (G_1, G_2) ". Nesse exemplo, é exposta uma proposição (Proposição 5), que é um caso particular do Teorema 17.

Na subsequente Subseção (5.1.2), é introduzida a classe de distribuições de probabilidade $F_{\mathcal{H}}$ -G. O Teorema 18 assegura a identificabilidade para a classe que é originada a partir do estudo mencionado. Um caso específico dessa classe, a "Classe de distribuições de probabilidade Normal-G", é apresentado. Como resultado direto do Teorema 18, uma proposição (Proposição 6) é exposta, garantindo a identificabilidade para classe Normal-G.

Na Subseção 5.1.3, é introduzida a classe $F_{\mathcal{H}}(x)$ -Tangente-G de distribuições de probabilidade. Seguem-se passos semelhantes aos descritos nos parágrafos anteriores, resultando na apresentação de mais um teorema que pode ser utilizado como critério para classificar uma função de distribuição como identificável.

Em conclusão, foram investigadas situações em que a família T-X de distribuições enfrenta problemas de identificabilidade paramétrica. Esses casos serão delineados a seguir:

- A distribuição de probabilidade Weibull-Weibull;
- A nova família Weibull-G de distribuições (TAHIR et al., 2016).

Embora apenas dois casos tenham sido demonstrados como exemplos de não identificabilidade na família T-X de distribuições, fica evidente que a composição sugerida por Alzaatreh, Lee e Famoye (2013) requer cautela na seleção das funções utilizadas no processo.

Artigos publicados e em fase de elaboração

Finalmente, apresentamos a seguir uma relação de artigos que foram publicados ou estão em fase de elaboração, desenvolvidos no contexto desta pesquisa de doutorado:

- GUSMÃO, F. R. S. et al. Analyzing and solving the identifiability problem in the exponentiated generalized weibull distribution. *Journal of the Egyptian Mathematical Society*, Springer, v. 29, n. 21, p. 2090–9128, 2021;
- GUSMÃO, F. R. S. et al. A study on the theory of parametric identifiability and its applications in new distributions and classes of probability distributions (em fase final de produção);
- GUSMÃO, F. R. S. et al. Identifiability in constructor methods of non-finite mixtures of distribution functions (em fase final de produção);

- GUSMÃO, F. R. S. et al. A study on the identifiability of the T-***G***-X family of distributions (em fase final de produção).

2 Revisão da Literatura

No presente segmento da tese, delineiam-se revisões literárias que abarcam tanto a erudição dos artigos científicos quanto o rigor das obras referenciais nas disciplinas de matemática, probabilidade e estatística. O desiderato subjacente a esta revisão é conferir ao leitor uma compreensão abalizada dos desafios inerentes à falta de identificabilidade paramétrica em distribuições e classes geradoras de distribuições de probabilidade. Ao sondar o contexto teórico por intermédio dessas fontes de conhecimento, os leitores são instados a familiarizarem-se com os dilemas prementes e as estratégias confeccionadas para contornar a problemática da identificabilidade em diversos cenários estatísticos.

Ademais, são delineadas definições, proposições, lemas, teoremas, aplicações e exemplos que funcionarão como esteios para a apreensão da teoria exposta nos capítulos subsequentes. Tal estruturação é concebida com o escopo de munir os leitores das ferramentas essenciais para assimilarem e aplicarem os princípios teóricos debatidos nos desdobramentos subsequentes.

2.1 Motivação para o estudo da identificabilidade

Recentemente, uma miríade de autores tem conjecturado novas classes de distribuições que se erigem como variantes das *fd's*, provendo funções de taxa de falha (*hf-hazard function*) com multifacetadas configurações.

A título ilustrativo, mencionamos a distribuição Weibull exponenciada, cunhada por [Mudholkar, Srivastava e Freimer \(1995\)](#), a qual ostenta uma função de risco com uma inusitada forma unimodal, conforme exegese de [Xie e Lai \(1996\)](#).

Por sua vez, [Carrasco, Ortega e Cordeiro \(2008\)](#) conceberam uma distribuição de quatro parâmetros, denominada distribuição de Weibull modificada e regenerada, cuja *hf* ostenta figuras não monótonas, como as configurações de banheira e banheira invertida; ao passo que [Gusmão, Ortega e Cordeiro \(2011\)](#) introduziram e investigaram a distribuição Weibull inversa generalizada com três parâmetros, a qual manifesta uma *hf* exibindo um perfil unimodal, crescente e decrescente. Por fim, [Pescim et al. \(2010\)](#) propuseram uma distribuição de quatro parâmetros designada distribuição beta semi-normal generalizada.

Um rol diversificado de famílias aventadas na literatura constitui uma fonte fecunda de distribuições de probabilidade para a modelagem de dados de vida, uma vez que, de

forma geral, o suporte da distribuição resultante é delineado a partir da *baseline*. [Cordeiro, Ortega e Cunha \(2013\)](#) propugnaram uma nova família, a saber, a classe de distribuições generalizada exponenciada (*EG*), com o fito de generalizar outras distribuições. Tendo em vista que uma variável aleatória T segue a distribuição G , propõe-se a aplicação da nova classe de distribuições a $G(t)$. Importa salientar que $0 < G(t) < 1$ para qualquer distribuição G , daí deduz-se a *fd EG* conforme expressa pela Eq. (3.39), em que $a > 0$ e $b > 0$ denotam dois parâmetros de forma adicionais. Não obstante, os autores sublinham que a nova classe de distribuições é menos intrincada que a família beta generalizada proposta por [Eugene, Lee e Famoye \(2002\)](#), sendo que mesmo as simulações são conduzidas com simplicidade, dado que a função quantílica (qf) possui uma formulação fechada.

As seguintes distribuições foram utilizadas e discutidas como baseline pelos autores [Cordeiro, Ortega e Cunha \(2013\)](#) e reconhecidas na literatura recente (esta lista não é exaustiva) são elas: A distribuição Birnbaum-Saunders ([CORDEIRO; LEMONTE, 2014](#)), a distribuição gama generalizada ([SILVA et al., 2015](#)), a distribuição Gumbel ([ANDRADE et al., 2015](#)), a distribuição Dagum ([GOMES-SILVA et al., 2017](#)), a distribuição Weibull generalizada exponenciada ([OGUNTUNDE; ODETUNMIBI; ADEJUMO, 2015](#)), a distribuição exponencial estendida ([ANDRADE et al., 2016](#)), a distribuição semi-logística padronizada ([CORDEIRO; LEMONTE; CAMPELO, 2017](#)), a distribuição de Pareto estendida ([ANDRADE; ZEA, 2018](#)), a distribuição Gumbel padronizada ([ANDRADE; GOMES-SILVA; ZEA, 2018](#)), e o Gompertz estendido ([ANDRADE et al., 2019](#)).

É amplamente sabido que a adição de parâmetros às classes de distribuição pode precipitar problemas de identificabilidade, e, por conseguinte, afetar a qualidade da estimação no modelo proposto.

A subsequência aduz uma relação de algumas distribuições de probabilidade. Estas distribuições, dentre outras, são reconhecidas por abrigarem problemáticas de identificabilidade paramétrica. Ei-las: Função de potência Kumaraswamy exponenciada ([KUMAR et al., 2021](#)), Weibull ponderada modificada ([MUHAMMED, 2021](#)), Weibull inversa generalizada transmutada exponenciada ([YASSMEN, 2018](#)), Kumaraswamy generalizada exponenciada ([ELGARHY; HAQ; AIN, 2018](#)), Kumaraswamy Rayleigh inversa exponenciada ([HAQ, 2016](#)), McDonald Weibull inversa generalizada ([SHAHBAZ; ISMAIL; BUTT, 2016](#)), Weibull generalizada exponenciada ([OGUNTUNDE; ODETUNMIBI; ADEJUMO, 2015](#)), Kumaraswamy inversa Weibull ([SHAHBAZ; SHAHBAZ; BUTT, 2012](#)), Weibull inversa generalizada ([GUSMÃO; ORTEGA; CORDEIRO, 2011](#)), e Beta Normal ([EUGENE; LEE; FAMOYE, 2002](#)).

Ainda, são discutidos ao longo do Capítulo 3 exemplos adicionais de distribuições de probabilidade não identificáveis, bem como suas correspondentes funções parte

paramétricas. As funções parte paramétricas serão definidas no Capítulo 3.

As Seções 2.2, 2.3 e 2.4 direcionam-se, respectivamente, aos Capítulos 3, 4 e 5.

2.2 Teoria geral da identificabilidade

Esta seção é revisada do artigo Basse e Bojinov (2020) além de outros trabalhos que serão devidamente citados quando necessário. Os estatísticos parecem concordar com uma definição no contexto de modelos estatísticos paramétricos e esta é apresentada a seguir. Seja um vetor paramétrico θ em um modelo estatístico ou estrutura estatística $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$, θ é identificável se o mapeamento $\theta \rightarrow P_\theta$ for injetivo.

2.2.1 Relações binárias

Sejam Θ e Λ dois conjuntos. Uma relação binária R de Θ para Λ é um subconjunto do produto cartesiano $\Theta \times \Lambda$. Para $\theta \in \Theta$ e $\lambda \in \Lambda$ diz-se que θ é R -relacionado a λ se $(\theta, \lambda) \in R$. Por convenção da obra de Halmos (2017) tem-se que $\theta R \lambda$ abrevia $(\theta, \lambda) \in R$. Serão definidas a seguir quatro importantes propriedades para as relações binárias, ver Freedman (2015).

Definição 1. A relação binária R de θ para Λ é dita ser:

- *injetiva* se $\forall \theta_1, \theta_2 \in \Theta, \forall \lambda \in \Lambda, \theta_1 R \lambda \text{ e } \theta_2 R \lambda \Rightarrow \theta_1 = \theta_2$;
- *sobrejetiva* se $\forall \lambda \in \Lambda, \exists \theta \in \Theta : \theta R \lambda$;
- *funcional* se $\forall \theta \in \Theta, \forall \lambda_1, \lambda_2 \in \Lambda, \theta R \lambda_1 \text{ e } \theta R \lambda_2 \Rightarrow \lambda_1 = \lambda_2$;
- *total à esquerda* se $\forall \theta \in \Theta, \exists \lambda \in \Lambda : \theta R \lambda$.

Uma relação binária que é funcional e total à esquerda é denominada de função.

2.2.2 Identificabilidade em conjuntos e funções

A definição seguinte forma a base para a unificação de como todas as outras definições de identificabilidade são obtidas pelas especificações adequadas para Θ , Λ e R .

Definição 2 (identificabilidade). *Sejam Θ e Λ dois conjuntos, e R uma relação binária sobrejetiva e total à esquerda de Θ para Λ . Então,*

- Θ é R -identificável em $\lambda_0 \in \Lambda$ se existe um único $\theta_0 \in \Theta$ tal que $\theta_0 R \lambda_0$;

- Θ é R -identificável em toda parte em Λ se é R -identificável em λ_0 para todo $\lambda_0 \in \Lambda$. Neste caso é dito que Θ é R -identificável.

Formalmente, identificável em toda parte é equivalente a relação binária ser injetiva.

Proposição 1. *Sejam Θ e Λ dois conjuntos, e R uma relação binária sobrejetiva e total à esquerda de Θ para Λ . Θ é R -identificável se e somente se R é injetiva.*

Definição 3 (Relação binária induzida). *Seja $\mathcal{S}$ um conjunto e $G(\mathcal{S})$ o conjunto de todas as funções com domínio em $\mathcal{S}$. Sejam $f_1, f_2 \in G(\mathcal{S})$, e $\Theta = \text{Img}(f_1)$ e $\Lambda = \text{Img}(f_2)$ as respectivas imagens. A relação binária de Θ para Λ definida como*

$$R_{f_1, f_2} = \{(f_1(S), f_2(S)) : S \in \mathcal{S}\}$$

é denominada de relação binária induzida associada com (f_1, f_2) .

A próxima proposição segue imediata da Definição 3.

Proposição 2. *Sejam $\mathcal{S}$ um conjunto e $f_1, f_2 \in G(\mathcal{S})$. A relação binária R_{f_1, f_2} é sobrejetiva e total à esquerda.*

Pela Definição 3 estende-se a noção de identificabilidade de conjuntos para funções.

Definição 4 (Identificabilidade de uma função). *Considere $\mathcal{S}, f_1, f_2 \in G(\mathcal{S})$, $\Theta = \text{Img}(f_1)$ e $\Lambda = \text{Img}(f_2)$.*

- A função f_1 é dita ser identificável em $\lambda_0 \in \Lambda$ se Θ é R -identificável em λ_0 . Isto é, para $\lambda_0 \in \Lambda$ com

$$\mathcal{S}_0 = \{S \in \mathcal{S} : f_2(S) = \lambda_0\},$$

então R_{f_1, f_2} é identificável em λ_0 se, e somente se, existir $\theta_0 \in \Theta$ tal que para todo $S \in \mathcal{S}$, tem-se que $f_1(S) = \theta_0$;

- A função f_1 é dita ser identificável em toda parte de Λ se Θ é R -identificável em toda parte de Λ . Neste caso é mais simples afirmar que f_1 é identificável.

2.2.3 Identificabilidade em modelos estatísticos

Existem dois benefícios significativos em usar a estrutura exibida nesta seção para lidar com a identificabilidade em modelos estatísticos, são eles:

- A flexibilidade da formalização matemática geral permite trabalhar diretamente com modelos paramétricos e não paramétricos, sem ter de introduzir definições separadas;

- Confiar em relações binárias em vez de funções enriquece a classe de questões que podem ser abordadas pelo olhar da identificabilidade.

Nota: Neste trabalho será abordados apenas os modelos paramétricos.

2.2.3.1 Modelos paramétricos

Considere o modelo paramétrico $\Lambda = \{P_{\theta} : \theta \in \Theta\}$ em que $\Theta \in \mathbb{R}^n$, P_{θ} é uma medida de probabilidade indexada por θ e $\mathbb{R}^n$ é o espaço Euclidiano n -dimensional.

A definição padrão de identificabilidade paramétrica é centrada na injetividade da parametrização. Duas definições com estas características são exibidas nos dois próximos parágrafos.

No livro *Statistical Inference* Casella e Berger (2001) na página 523 na Definição 11.2.2. é afirmado que um vetor paramétrico θ que indexa uma família de distribuições $\{f(x, \theta) : \theta \in \Theta\}$ é identificável se distintos valores de θ correspondem a distintas funções densidade de probabilidade *fdp's* ou funções de probabilidade (*fp's*). Isto é, se $\theta \neq \theta'$, então $f(x, \theta)$ será diferente de $f(x, \theta')$ para algum $x \in \mathbb{R}$.

Em Lehmann e Casella (2006) define-se que se $X \sim P_{\theta}$, então θ é dita ser não identificável sobre a base de X se existe $\theta_1 \neq \theta_2$ para os quais $P_{\theta_1} = P_{\theta_2}$.

Definição 5 (Identificabilidade paramétrica por Basse e Bojinov (2020)). *O vetor paramétrico θ do modelo estatístico $\Lambda = \{P_{\theta} : \theta \in \Theta\}$ é dito identificável se a função f , tal que $f : \theta \rightarrow P_{\theta}$, é injetiva.*

2.3 Misturas de distribuições definidas por Henry Teicher

Esta seção é revisada com base nos artigos [Teicher \(1954\)](#), [Teicher \(1960\)](#), [Teicher \(1961\)](#), [Teicher \(1963\)](#) sobre a identificabilidade de misturas e de classes de misturas de fd 's com base em teoria da medida.

Seja $\mathcal{F} = \{F_{\mathcal{F}}(x) : x \in \mathbb{R}\}$ uma família de fd 's. Seja $\mathcal{G} = \{\{y\} : y = F(x); \forall x \in \mathbb{R}\}$ e considere $\mathcal{A} = \sigma(\mathcal{G})$ a menor σ -álgebra gerada por $\mathcal{G}$. $\mathcal{A}$ coincide com a σ -álgebra de Borel de subconjuntos do intervalo $[0, 1]$. Seja $\mathcal{M} = \{\mu(A) : A \in \mathcal{A}\}$ com $\mu([0, 1]) = 1$ para todo $\mu \in \mathcal{M}$. Então, define-se a mistura H ou μ -mistura de $\mathcal{F}$ providenciando a mistura de medidas pelo mapeamento mensurável de g para $\mathcal{F}$ com domínio na linha real pela expressão

$$H(x) = \int_{\mathcal{F}} g[F_{\mathcal{F}}(x)] d\mu[F_{\mathcal{F}}(x)] = \int_{\mathcal{F}} F_{\mathcal{F}}(x) d\mu[F_{\mathcal{F}}(x)] \quad (2.1)$$

em que $g = g_x(F_{\mathcal{F}}) = F_{\mathcal{F}}$. Desde que μ não atribua valor 1 (um) para um específico membro de $\mathcal{F}$. Desta forma o termo mistura significa uma média ponderada de fd 's.

Para a estipulada classe $\mathcal{F}$, a família $\mathcal{H} = \{H(x) : x \in \mathbb{R}\}$ com $H(x)$ expressa como em (2.1) será denotada pela classe de $\mathcal{H}$ -misturas em $\mathcal{F}$ ou classe de misturas em $\mathcal{F}$.

Definição 6 (Identificabilidade da família $\mathcal{H}$). *Uma μ -mistura de $\mathcal{F}$, $H(x)$, será chamada identificável se, para qualquer medida de probabilidade μ^* a igualdade*

$$H(x) = \int_{\mathcal{F}} F_{\mathcal{F}}(x) d\mu[F_{\mathcal{F}}(x)] = \int_{\mathcal{F}} F_{\mathcal{F}}(x) d\mu^*[F_{\mathcal{F}}(x)] \quad (2.2)$$

implicar em $\mu[F_{\mathcal{F}}(x)] = \mu^[F_{\mathcal{F}}(x)]$.*

Se cada membro da classe $\mathcal{H}$ de μ -misturas de $\mathcal{F}$ é identificável, $\mathcal{H}$ será dita identificável.

Considere agora que a família de fd 's $\mathcal{F} = \{F_{\mathcal{F}}(x) : \alpha \in \Theta_{\mathcal{F}} \subset \mathbb{R}^n, x \in \mathbb{R}\}$ em que $\Theta_{\mathcal{F}}$ é o *evp* (*evp*) e $\mathbb{R}^n$ é o espaço Euclidiano n -dimensional. Seja $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ um vetor de parâmetros ou vetor paramétrico tal que $\alpha_i \in \mathbb{R}$ para todo $1 \leq i \leq n$. Sejam $\Theta_{\mathcal{F}}$ um *evp* mensurável ($\Theta_{\mathcal{F}}$ -mensurável) e a família $\mathcal{F}$ (como aqui descrita) $\mathbb{R} \times \Theta_{\mathcal{F}}$ -mensurável com dimensão $n + 1$. Seja $\mathcal{G}$ uma classe de fd 's n -dimensionais expressa por $\mathcal{G} = \{G(\alpha) : \alpha \in \Theta_{\mathcal{F}} \subset \mathbb{R}_1^n\}$ em que G é $\Theta_{\mathcal{F}}$ -mensurável.

A classe de medidas de Lebesgue-Stieltjes μ_G , que não concentra toda a massa em um simples ponto $\alpha \in \Theta_{\mathcal{F}}$ (classe não degenerada), induzida por $G \in \mathcal{G}$ atribui medida 1 (um) para $\Theta_{\mathcal{F}}$, ou seja, $\mu_G(\Theta_{\mathcal{F}}) = G(\Theta_{\mathcal{F}}) = 1$. Então, pode-se escrever a fd unidimensional ($x \in \mathbb{R}$) H ou mistura H ou ainda a G -mistura de $\mathcal{F}$ como

$$H(x) = \int_{\Theta_{\mathcal{F}}} F_{\mathcal{F}}(x; \alpha) dG(\alpha) \quad (2.3)$$

mais especificamente a *fd* H é uma G -mistura de elementos de $\mathcal{F}$ que providencia a mistura de medidas. Assim, o termo mistura (neste contexto) significa uma média ponderada por G de *fd*'s em $\mathcal{F}$.

Seja agora $\mathcal{H} = \{H(x) : \alpha \in \Theta_{\mathcal{F}} \subset \mathbb{R}^n, x \in \mathbb{R}\}$ a classe das misturas. Seja $\tilde{\mathcal{F}}$ um mapeamento tal que

$$\begin{aligned}\tilde{\mathcal{F}} : \mathcal{G} &\longrightarrow \mathcal{H} \\ G &\mapsto H(x).\end{aligned}$$

Se $\tilde{\mathcal{F}}$ é um mapeamento um-a-um sobre $\mathcal{H}$, então $\mathcal{H}$ é dita uma classe identificável.

Considere $\mathcal{H}_0 \subset \mathcal{H}$ uma subclasse qualquer e $\mathcal{G}_0 \subset \mathcal{G}$ outra subclasse qualquer tais que

$$\begin{aligned}\tilde{\mathcal{F}} : \mathcal{G}_0 &\longrightarrow \mathcal{H}_0 \\ G &\mapsto H(x)\end{aligned}$$

em que $\mathcal{G}_0 = \tilde{\mathcal{F}}^{-1}(\mathcal{H}_0)$ (imagem inversa) com $\tilde{\mathcal{F}}$ sendo um mapeamento um-a-um. Assim, $\mathcal{H}_0$ é dita uma classe identificável.

Considere $\mathcal{J}$ a classe das distribuições degeneradas $\Theta_{\mathcal{F}}$ -mensuráveis. Então $\mathcal{H}$ será dita identificável em $\mathcal{G}$, com respeito a $\mathcal{F}$, se houver correspondência um-a-um entre $\mathcal{H} \cup \mathcal{F}$ e $\mathcal{G} \cup \mathcal{J}$; equivalente, se a identidade

$$\int_{\Theta_{\mathcal{F}}} F(x; \alpha) dG(\alpha) = \int_{\Theta_{\mathcal{F}}} F(x; \alpha) dG^*(\alpha) \quad (2.4)$$

implicar em $G = G^*$ para todo G^* pertencente a $\mathcal{G} \cup \mathcal{J}$, $\mathcal{H}$ é identificável.

2.3.1 Misturas de famílias aditivamente fechadas

Nesta subseção é revisado um teorema que garante que a classe de todas as misturas de *fd*'s uniparamétricas aditivamente fechadas são identificáveis [Teicher \(1961\)](#).

Uma família de *fd*'s $\mathcal{F} = \{F_{\mathcal{F}}(x; \alpha) : \alpha \in D, x \in \mathbb{R}\}$ é chamada de aditivamente fechada para cada $\alpha, \beta \in D$ se

$$F(x; \alpha) * F(x; \beta) = F(x; \alpha + \beta) \quad (2.5)$$

em que $*$ é uma convolução. Seja D a notação genérica de um semigrupo aditivo abeliano; $D(\mathbb{Z})$, $D(\mathbb{Q})$ e $D(\mathbb{R})$ denotam os semigrupos dos números inteiros, racionais e reais; respectivamente. De forma análoga $D(\mathbb{Z}+)$, $D(\mathbb{Q}+)$ e $D(\mathbb{R}+)$ são semigrupos restritos a valores positivos. Considere $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ e seja $D = (D_1, D_2, \dots, D_n)$ com $D_i \subset \mathbb{R}$ um semigrupo aditivo abeliano em que $\alpha_i \in D_i$ para todo $1 \leq i \leq n$.

Definição 7. A fd expressa em (2.3) é dita uma mistura da família aditivamente fechada $\mathcal{F}$ se atender a:

1. $\mathcal{F}$ é a aditivamente fechada;
2. $F_{\mathcal{F}}(x; \alpha)$ é mensurável;
3. D é um subconjunto mensurável de $\mathbb{R}^n$ com $\mu_G \{D\} = 1$.

Foi demonstrado em Teicher (1954) que para $n = 1$, $D(\mathbb{Z}+)$ e $D(\mathbb{Q}+)$ uma família aditivamente fechada $\mathcal{F} = \{F_{\mathcal{F}}(x; \alpha) : \alpha \in D, x \in \mathbb{R}\}$ possui funções característica na forma

$$\phi(t; \alpha) = [\phi(t)]^\alpha \quad (2.6)$$

em que $\phi(t) = \phi(t; 1)$ é uma função característica independente de α . Examinando as demonstrações dos Teoremas 1 e 2 de Teicher (1954) encontra-se que (2.6) também pode ser obtida para $D(\mathbb{R}+)$ apenas para $F_{\mathcal{F}}(x; \alpha)$ mensurável. Nos casos $D(\mathbb{Z})$, $D(\mathbb{Q})$ e $D(\mathbb{R})$ e também nos casos $D(\mathbb{Z}_+)$, $D(\mathbb{Q}_+)$ e $D(\mathbb{R}_+)$ a função característica da mistura H , dita $\psi_H(t)$, com família $\mathcal{F}$ aditivamente fechada é expressa na forma

$$\psi_H(t) = \int_D [\phi(t)]^\alpha dG(\alpha). \quad (2.7)$$

Teorema 1. Se $n = 1$ e D é $D(\mathbb{Z}+)$ ou $D(\mathbb{Q}+)$ ou $D(\mathbb{R}+)$, a classe de misturas $\mathcal{H} = \{H(x) : \alpha \in D, x \in \mathbb{R}\}$ tal que

$$H(x) = \int_D F(x; \alpha) dG(\alpha)$$

de uma família $\mathcal{F} = \{F_{\mathcal{F}}(x; \alpha) : \alpha \in D, x \in \mathbb{R}\}$ aditivamente fechada é identificável.

Demonstração. Se, em (2.7), $|\phi(t)| \equiv 1$ então $\phi(t) = \exp(it\theta)$ com $\theta \in \mathbb{R}$ e $\theta \neq 0$ já que $\mathcal{F}$ contém pelo menos dois elementos. Portanto, G é unicamente determinada por $\psi_H(t)$ a qual é unicamente gerada por H .

Se D contém apenas valores não negativos, a transformada

$$\psi(z, t) = \int_D z^\alpha dG(\alpha) \quad (2.8)$$

é analítica em $0 < |z| < 1$; se duas distintas fd's G_1 e G_2 gerassem a mesma mistura H , então $\psi(z, G_1)$ e $\psi(z, G_2)$ coincidiriam para $z = \phi(t)$ e para $0 < |z| < 1$. Isto implica em $\psi(\rho \exp(it\theta), G_1) = \psi(\rho \exp(it\theta), G_2)$ para todo $\rho \in (0, 1)$ e por isto, pelo Teorema da Convergência Dominada, para $\rho = 1$. Mas $\psi(\rho \exp(it\theta), G_1) = \psi(\rho \exp(it\theta), G_2)$ implica $G_1 = G_2$ pelo Teorema da identificabilidade por transformadas de Fourier.

Portanto, fica provado o Teorema 1. □

Quando $n > 1$ a função característica de uma família aditivamente fechada é da forma

$$\prod_{i=1}^n [u_i(t)]^{\alpha_i}$$

pelo menos para um adequado D . Alguns dos parâmetros α_i podem assumir valores negativos e positivos. Como um breve exemplo, exibido em [Teicher \(1961\)](#), a família das distribuições Normais bi-paramétrica não gera uma família identificável.

2.3.2 Identificabilidade em misturas finitas

No artigo de [Teicher \(1963\)](#), é afirmado que as classes de misturas pertencentes à família de distribuições normais, distribuições gama ou distribuições binomiais em geral são consideradas não identificáveis. No entanto, o mesmo artigo apresenta dois teoremas que asseguram que todas as misturas finitas de fd 's de probabilidade são, de fato, identificáveis.

Seja G uma distribuição ou a correspondente μ_G uma medida discreta e que atribui massa positiva a um número finito de pontos $\alpha_i \in \Theta_{\mathcal{F}}$ tal que $1 \leq i \leq k$, desta forma H' é dita uma mistura finita.

Seja $\mathcal{F} = \{F_{\mathcal{F}}(x; \alpha_i) : \alpha_i \in \Theta_{\mathcal{F}} \subset \mathbb{R}^n, x \in \mathbb{R}, 1 \leq i \leq k\}$ uma família finita de fd 's com $k \in \mathbb{N}$.

Teorema 2. *Uma condição necessária e suficiente para a classe $\mathcal{H}' = \left\{ \sum_{i=1}^k c_i F_{\mathcal{F}}(x; \alpha_i) \right\}$, com $c_i \geq 0$ e $\sum_{i=1}^k c_i = 1$, de todas as misturas finitas da família finita $\mathcal{F} = \{F_{\mathcal{F}}(x; \alpha_i)\}$ seja identificável é que exista k valores reais $x_1, x_2, \dots, x_k$ para os quais o determinante de $F_{\mathcal{F}}(x_j; \alpha_i)$, com $1 \leq i, j \leq k$ não se anule.*

Teorema 3. *Seja $\mathcal{F}$ uma família de fd 's com transformada $\phi(t)$ definida em $t \in D_{\phi}$ (domínio da definição de ϕ) tal que o mapeamento $M : F \rightarrow \phi$ seja linear e um-a-um. Suponha que existe um total ordenamento ($\prec$) de elementos de $\mathcal{F}$ tal que $F_{\mathcal{F}}(x; \alpha_1) \prec F_{\mathcal{F}}(x; \alpha_2)$ implique em*

1. $D_{\phi_1} \subset D_{\phi_2}$;
2. A existência de algum $t_1 \in [D_{\phi_1}]^c$ (t_1 sendo independente de ϕ_2) tal que

$$\lim_{t \rightarrow t_1} \phi_2(t)/\phi_1(t) = 0.$$

Então a classe $\mathcal{H}$ de todas as misturas finitas de $\mathcal{F}$ é identificável.

Exemplo 1. *Considere a distribuição de probabilidade a seguir*

$$F(x; \lambda, \gamma) = \gamma^{\lambda} [\Gamma(\lambda)]^{-1} \int_0^x u^{\lambda-1} \exp(-\gamma u) du$$

em que $x > 0$, $\lambda > 0$ e $\gamma > 0$; e função característica da distribuição 1 dada por

$$\phi(t; \lambda, \gamma) = [1 - (it/\gamma)]^{-\lambda}.$$

Considere $\gamma = 1$ e por isto segue que a família $\{F(x; \lambda, 1)\}$ é aditivamente fechada. Pelo Teorema 1 a classe de misturas da referida família é identificável. Seja $\mathcal{H}$ a classe de todas as misturas finitas da família $\{F(x; \lambda, \gamma)\}$, através de 1 e pelo Teorema 3 a classe $\mathcal{H}$ é identificável.

2.4 A família T-**G**-X de distribuições de probabilidades

Brito et al. (2019) introduzem um método que se vale de múltiplas baselines para a geração tanto de distribuições de probabilidade quanto de classes correspondentes. Este método engendra classes de distribuição a partir de funções monotônicas preestabelecidas e de distribuições pré-existent. Como consequência, é possível introduzir uma variada gama de classes de distribuições de probabilidade, tanto aquelas já reconhecidas na literatura estatística quanto aquelas até então desconhecidas. Ademais, os resultados obtidos não apenas oferecem insights sobre a amplitude de modelagem, mas também sobre a natureza intrínseca das distribuições produzidas.

2.4.1 Método gerador distribuições e classes de distribuições de probabilidade no caso univariado T-**G**-X

Seja $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_{\mathcal{F}})$ um espaço de probabilidade em que $\mathcal{B}(\mathbb{R})$ é a σ -álgebra de Borel formada por subconjuntos de $\mathbb{R}$ e $F_{\mathcal{F}}$ é uma *fd*. Seja $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ uma família de *fd*'s de probabilidade indexadas por $\boldsymbol{\theta}$, em que $\boldsymbol{\theta}$ é um *vp* e $\Theta_{\mathcal{F}}$ é o *evp* da família $\mathcal{F}$.

Seja $g(x; \boldsymbol{\alpha}) = g[F_{\mathcal{G}_1}(x; \boldsymbol{\theta}_1), F_{\mathcal{G}_2}(x; \boldsymbol{\theta}_2), \dots, F_{\mathcal{G}_n}(x; \boldsymbol{\theta}_n)]$ uma função real monótona, contínua à direita e não decrescente tal que $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ em que $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$, $\boldsymbol{\alpha} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n)$. Seja $h(x; \boldsymbol{\beta}) = h[F_{\mathcal{G}_{n+1}}(x; \boldsymbol{\theta}_{n+1}), F_{\mathcal{G}_{n+2}}(x; \boldsymbol{\theta}_{n+2}), \dots, F_{\mathcal{G}_m}(x; \boldsymbol{\theta}_m)]$ uma função real monótona, contínua à direita e não crescente tal que $h : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ em que $\boldsymbol{\beta} = (\boldsymbol{\theta}_{n+1}, \boldsymbol{\theta}_{n+2}, \dots, \boldsymbol{\theta}_m)$. Seja $\mathcal{G}_i = \{F_{\mathcal{G}_i}(x; \boldsymbol{\theta}_i) : \boldsymbol{\theta}_i \in \Theta_{\mathcal{G}_i}, x \in \mathbb{R}\}$ uma família de *fd*'s de probabilidade para cada $1 \leq i \leq m$ com $n, m \in \mathbb{N}$. Seja $\mathcal{H} = \{F_{\mathcal{H}}(x; \boldsymbol{\rho}) : \boldsymbol{\rho} \in \Theta_{\mathcal{H}}, x \in \mathbb{R}\}$ uma família de *fd*'s. As funções $g(x; \boldsymbol{\alpha})$ e $h(x; \boldsymbol{\beta})$ são tais que:

1. $\lim_{x \rightarrow -\infty} g(x; \boldsymbol{\alpha}) \leq \lim_{x \rightarrow -\infty} h(x; \boldsymbol{\beta})$;
2. $\lim_{x \rightarrow \infty} g(x; \boldsymbol{\alpha}) \geq \sup \{x \in \mathbb{R} : F_{\mathcal{H}}(x; \boldsymbol{\rho}) < 1\}$;
3. $\lim_{x \rightarrow \infty} h(x; \boldsymbol{\beta}) \leq \inf \{x \in \mathbb{R} : F_{\mathcal{H}}(x; \boldsymbol{\rho}) > 0\}$;
4. $F_{\mathcal{H}}(x; \boldsymbol{\rho})$ não possui pontos de descontinuidade ou $g(x; \boldsymbol{\alpha})$ e $h(x; \boldsymbol{\beta})$ são constantes à direita na vizinhança dos pontos cujas as imagens são pontos de descontinuidade de $F_{\mathcal{H}}$.

Pode-se escrever o método gerador distribuições e classes de distribuições de probabilidade T-**G**-X para o caso univariado como segue

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = \int_{h(x; \boldsymbol{\beta})}^{g(x; \boldsymbol{\alpha})} dF_{\mathcal{H}}(x; \boldsymbol{\rho}) = F_{\mathcal{H}}[g(x; \boldsymbol{\alpha}); \boldsymbol{\rho}] - F_{\mathcal{H}}[h(x; \boldsymbol{\beta}); \boldsymbol{\rho}]$$

$$= F_{\mathcal{N}}(x; \boldsymbol{\delta}) - M(x; \boldsymbol{\psi}) \quad (2.9)$$

em que $\mathcal{N} = \{F_{\mathcal{N}}(x; \boldsymbol{\delta}) : \boldsymbol{\delta} \in \Theta_{\mathcal{N}}, x \in \mathbb{R}\}$ é uma família de *fd*'s de probabilidade. Os *vp*'s $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\rho}, \boldsymbol{\beta}, \boldsymbol{\rho})$, $\boldsymbol{\delta} = (\boldsymbol{\alpha}, \boldsymbol{\rho})$ e $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\rho})$ estão assim dispostos. A função real M não é uma *fd* de probabilidade, pois ela é uma função não-crescente e tem-se que

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) : \overline{\mathbb{R}} \rightarrow [0, 1]$$

em que $\overline{\mathbb{R}} = [-\infty, \infty]$.

O método gerador T- $\boldsymbol{G}$ -X é um caso geral aplicável à família T-X de distribuições. Em outras palavras, a família T-X de distribuições pode ser considerada um caso específico deste método mais abrangente. Quando o vetor de baselines $\boldsymbol{G}$ do método T- $\boldsymbol{G}$ -X é configurado para resultar em uma única baseline, a família T-X de distribuições é obtida. O método gerador T- $\boldsymbol{G}$ -X apresenta limites superiores e inferiores, os quais podem incorporar parâmetros adicionais, influenciando assim a regra de correspondência entre eles e, conseqüentemente, afetando a modelagem da *fd* $F_{\mathcal{F}}(x; \boldsymbol{\theta})$.

2.4.1.1 Família T-X de distribuições

Uma nova classe de generalizações foi apresentada em [Alzaatreh, Lee e Famoye \(2013\)](#), a família T-X de distribuições, que generaliza algumas outras classes e distribuições de probabilidade já conhecidas na literatura estatística.

Seja $r(x)$ a fdp de uma variável aleatória contínua $T \in [a, b]$ ($a, b \in \mathbb{R}$), para $-\infty \leq a < b \leq \infty$. Seja $W[G(x)]$ uma função de $G(x)$ para qualquer variável aleatória X tal que $W[G(x)]$ satisfaça as seguintes condições:

1. $W[G(x)] \in [a, b]$;
2. $W[G(x)]$ é diferenciável e monótona não decrescente;
3. $W[G(x)] \rightarrow a$ quando $x \rightarrow -\infty$ e $W[G(x)] \rightarrow b$ quando $x \rightarrow \infty$.

Considere X uma variável aleatória com fdp $g(x)$ e *fd* $G(x)$. Seja a fdp $r(t) \in [a, b]$. A *fd* da família de distribuições é definida como segue:

$$F_{T-X}(x) = \int_a^{W[G(x)]} r(t) dt = R_{(T-X)}\{W[G(x)]\}, \quad (2.10)$$

em que a fdp da família T-X de distribuições é dada por

$$f_{T-X}(x) = r\{W[G(x)]\} \left\{ \frac{d[W[G(x)]]}{dx} \right\}. \quad (2.11)$$

Dependendo do suporte de T , tem-se a definição para $W[G(x)]$. Quando T possui suporte no intervalo $(0, 1)$ $W[G(x)]$ pode ser definida como segue:

- $G(x)$;
- $[G(x)]^\alpha$.

Quando T possui suporte no intervalo $(0, \infty)$ $W[G(x)]$ pode ser definida como segue:

- $-\log[1 - G(x)]$;
- $-\log\{1 - [G(x)]^\alpha\}$;
- $G(x)/1 - G(x)$;
- $G(x)^\alpha/1 - [G(x)]^\alpha$.

Ainda, quando T possui suporte no intervalo $(-\infty, \infty)$; $W[G(x)]$ pode ser definida como segue:

- $\log[-\log[1 - G(x)]]$;
- $\log\{-\log\{1 - [G(x)]^\alpha\}\}$;
- $\log\{G(x)/[1 - G(x)]\}$;
- $\log\{G(x)^\alpha/[1 - G(x)^\alpha]\}$.

Em que $\alpha > 0$.

3 Teoria da Identificabilidade Paramétrica

3.1 Introdução

Este capítulo é organizado em três seções distintas: uma introdução, uma seção dedicada à apresentação de definições, teoremas, lemas, corolários e proposições, acompanhadas de exemplos com ênfase didática, e uma terceira seção que ilustra a aplicação da teoria desenvolvida na segunda seção. Em geral, os exemplos são fornecidos com o propósito de facilitar a compreensão das teorias apresentadas. Eles são apresentados imediatamente após a introdução de alguma proposição, lema ou teorema relevante na respectiva seção. As contribuições contidas na seção 3.2 constituem uma parte crucial da teoria de identificabilidade paramétrica abordada e exposta neste trabalho. Tais contribuições são aplicáveis às *fd*'s e às classes geradoras de distribuições de probabilidade cujo domínio está na reta real. As contribuições discutidas na seção 3.2 podem ser empregadas como um critério para determinar se uma *fd* é identificável.

3.2 Definições, proposições, lemas e teoremas para teoria da identificabilidade

Nesta seção, são delineadas diversas definições, teoremas, lemas, corolários e proposições que constituem contribuições originais para a teoria da identificabilidade paramétrica. Tais contribuições não só enriquecem este estudo, mas também estabelecem os fundamentos necessários para os desenvolvimentos subsequentes nos Capítulos 4 e 5.

Definição 8 (Modelos estatísticos). *O modelo estatístico (ou estrutura estatística) é a trinca $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{F})$ em que $\mathcal{B}(\mathbb{R})$ é a σ -álgebra de Borel formada por subconjuntos de $\mathbb{R}$, $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ é uma família de *fd*'s $F_{\mathcal{F}}$ sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\boldsymbol{\theta}$ é um vetor paramétrico (vp) e $\Theta_{\mathcal{F}}$ é o evp.*

A definição para a identificabilidade paramétrica utiliza a ideia de injetividade de funções que possuem como domínio o evp Θ ou espaço paramétrico Θ , no caso da *fd* de probabilidade possuir um único parâmetro.

Definição 9 (Identificabilidade paramétrica). *Seja a trinca $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{F})$ uma estrutura estatística. Seja $\mathcal{F}$ uma família de *fd*'s ou de acordo com Basse e Bojinov (2020) um*

modelo paramétrico indexado por θ . Seja R uma relação binária $\Theta_{\mathcal{F}}$ -identificável de $\Theta_{\mathcal{F}}$ espaço vetorial paramétrico (evp) para $\mathcal{F}$ de acordo com a Definição 2. Considere f uma função tal que $f \in R$ e que

$$f : \Theta_{\mathcal{F}} \rightarrow \mathcal{F}$$

$$\theta \mapsto F_{\mathcal{F}}(x; \theta).$$

Define-se que $\Theta_{\mathcal{F}}$ é identificável (ver Definição 3). Também pode-se afirmar que a família $\mathcal{F}$, o espaço de probabilidade, a estrutura estatística e o modelo paramétrico são identificáveis.

Caso R seja uma relação não injetiva, tal como na Definição 2, diz-se que $\Theta_{\mathcal{F}}$ é não identificável ou o espaço de probabilidade ou a estrutura estatística é não identificável ou $\mathcal{F}$ é não identificável ou ainda a fd é não identificável.

O estudo de Fraser, Hsu e Walker (1981) aborda a identificabilidade de fd's aplicadas a dados circulares, uma temática que está além do escopo desta tese. O foco do presente trabalho reside na análise de fd's e de classes geradoras de distribuições de probabilidade com domínio na reta real.

O exemplo a seguir exhibe uma fd que de acordo com a Definição 9 é não identificável.

Exemplo 2. No artigo Elgarhy, Haq e Ain (2018) foi realizada a proposta de uma nova fd chamada de Kumaraswamy generalizada exponenciada (EG-Kum) cuja expressão é dada por

$$F_{\mathcal{F}}(x; a, b, \alpha, \beta) = \left[1 - (1 - x^a)^{\alpha b}\right]^{\beta} \quad (3.1)$$

em que a , b , α e β são parâmetros de forma, reais positivos e $x \in (0, 1)$. Seja $\mathcal{F} = \{F_{\mathcal{F}}(x; \theta) : \theta \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ a família de fd's EG-Kum e $\theta = (a, b, \alpha, \beta)$.

Pelo produto entre os parâmetros α e b e devido ao fato de que nenhum dos dois parâmetros está explícito em qualquer outro lugar na fd (3.1), também pela Definição 9, o evp $\Theta_{\mathcal{F}}$ é não identificável, ou seja, a família $\mathcal{F}$ da distribuição EG-Kum é não identificável.

Em fd's paramétricas a quantidade elevada de parâmetros pode ocasionar problemas no momento de estimá-los e podem gerar o problema de sobreajuste (overfitting) dos dados.

A consideração de uma distribuição que apresente uma parcimônia em relação à quantidade de parâmetros se revela como um aspecto de notável interesse, uma vez que tal característica propicia uma via menos intrincada no processo de estimação destes parâmetros. Ademais, viabiliza-se a aplicação destas distribuições tanto em conjuntos de dados reais quanto em simulações com um nível de adequação satisfatório. Nesse contexto, é pertinente apresentar uma definição que estabeleça uma relação entre os parâmetros de

distribuições paramétricas, considerando a quantidade destes parâmetros em duas ou mais distribuições.

A definição de parcimônia paramétrica, aqui delineada, ostenta um caráter específico que se coaduna com os propósitos intrínsecos deste estudo. Ela se erige como uma ferramenta fundamental para a realização de comparações entre a quantidade de parâmetros presentes em duas ou mais distribuições, independentemente de pertencerem à mesma família de *fd*'s ou não.

Definição 10 (Parcimônia paramétrica entre *fd*'s). *A parcimônia paramétrica entre fd 's paramétricas ocorre quando compara-se duas ou mais distribuições, com relação ao número de parâmetros de cada uma delas, e localiza-se a fd com um menor número de parâmetros. É dita ser a fd paramétrica mais parcimoniosa aquela que possuir uma menor quantidade de parâmetros com relação as outras distribuições em questão.*

Considerando as distribuições Normal e Exponencial, conforme discutido em [Johnson, Kotz e Balakrishnan \(1994\)](#), percebe-se que a segunda distribuição demonstra uma característica de parcimônia superior à primeira, o que facilita a estimação dos parâmetros da segunda. É importante ressaltar que, em geral, a preferência recai sobre distribuições mais parcimoniosas, embora seja prudente reconhecer que a distribuição mais parcimoniosa nem sempre é a mais adequada para ajustar determinados conjuntos de dados.

É relevante observar que a distribuição Exponencial não se configura como um caso particular (ou distribuição filha) da distribuição Normal. Dessa forma, é possível realizar comparações entre distribuições filhas e não filhas em termos da parcimônia paramétrica entre *fd*'s.

Ao analisar a comparação entre a quantidade de parâmetros das distribuições Weibull biparamétrica e Weibull triparamétrica, conforme abordado por [Johnson, Kotz e Balakrishnan \(1994\)](#), constata-se que a Weibull biparamétrica apresenta uma maior parcimônia em relação à triparamétrica, sendo esta última um caso particular daquela (distribuição filha). Ambas as distribuições têm a capacidade de modelar o mesmo conjunto de dados, e, nesse contexto, a superioridade em termos de parcimônia implica em uma maior viabilidade matemática.

A próxima definição a ser apresentada refere-se à expressão de uma *fd* como uma composição de n funções reais. Tal definição assume relevância ao longo desta tese, pois será empregada na formulação de definições, lemas, proposições e teoremas, além de ser um recurso recorrente em diversas demonstrações.

Definição 11 (Função parte paramétrica da *fd*). *Considere uma família de funções reais $\mathcal{H} = \{h_n(x; \theta_n) : \theta_n \in \Theta_{h_n}, x \in \mathbb{R}\}$ com domínio em Θ_{h_n} (evp associado h_n com $n \in \mathbb{N}$).*

Seja f uma função real tal que

$$f : \Theta_{h_n} \rightarrow \mathcal{H}. \quad (3.2)$$

Seja $\mathcal{G}_i = \{g_i[x; (\boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_n)] : (\boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_n) \in \Theta_{h_i, \dots, h_n}, x \in \mathbb{R}\}$ uma classe de funções reais compostas por funções h 's, em que $g_i[x; (\boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_n)] = h_i(x; \boldsymbol{\theta}_i) \circ \dots \circ h_n(x; \boldsymbol{\theta}_n)$, com $1 \leq i \leq (n-1)$.

Sejam as funções h 's tais que

$$h_{n-1} : \mathcal{H} \rightarrow \mathcal{G}_{n-1} \quad e \quad h_j : \mathcal{G}_{j+1} \rightarrow \mathcal{G}_j \quad (3.3)$$

em que $1 \leq j \leq (n-2)$.

Seja $\mathcal{G}_1 = \{g_1(x; \boldsymbol{\theta}) ; \boldsymbol{\theta} \in \Theta, x \in \mathbb{R}\}$ em que $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n)$ é o vp transposto de $\Theta = \Theta_{h_i, \dots, h_n}$.

Se uma fd $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ puder ser escrita como uma composição das funções reais $h_1, h_2, \dots, h_n$ tal como

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = h_1(x; \boldsymbol{\theta}_1) \circ h_2(x; \boldsymbol{\theta}_2) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ h_n(x; \boldsymbol{\theta}_n) \quad (3.4)$$

em que $\boldsymbol{\theta} \in \Theta$ e $\Theta = \Theta_{\mathcal{F}}$, então define-se que a função real h_n na expressão (3.4) é a função parte paramétrica da fd $F_{\mathcal{F}}$.

As funções h 's tal como exibido na expressão (3.4) possuem parâmetros, porém é em h_n que serão alocados de forma arbitrária os possíveis parâmetros redundantes.

O seguinte teorema está embasado nas Definições 11 e 9, e poderá ser utilizado como um método para testar se uma fd paramétrica é identificável.

Teorema 4 (Identificabilidade pela injetividade das funções que compõem a fd). *Considere a fd $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ da família $\mathcal{F}$ tal como descrita na Definição 11. Seja*

$$\mathcal{G}_i = \{g_i[x; (\boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_n)] : (\boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_n) \in \Theta_{h_i, \dots, h_n}, x \in \mathbb{R}\}$$

uma classe de funções injetivas em $\Theta_{h_i, \dots, h_n}$, para todo $1 \leq i \leq (n-1)$. Sejam injetivas as funções f e h_i 's para todo $1 \leq i \leq n$, então o evp $\Theta_{\mathcal{F}}$ é identificável.

Demonstração. Pela Definição 11 e pela Definição 9 é direta a demonstração de que $\Theta_{\mathcal{F}}$ é identificável. $\square$

De acordo com Catchpole e Morgan (1997), uma fd que possui uma quantidade m de parâmetros redundantes, ou seja, parâmetros que não contribuem para a identificação

da distribuição (identificabilidade), pode ser submetida a uma redução na sua dimensionalidade paramétrica, eliminando-se os parâmetros redundantes, e, desse modo, tornar-se identificável. A formalização deste processo, fundamentada na Definição 11 e a assertiva deste parágrafo, é apresentada pelo Corolário 1.

Corolário 1 (Redução mínima até identificabilidade paramétrica). *Seja $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ tal como na Definição 11. Considere que haja apenas em h_n uma quantidade m parâmetro(s) redundante(s) e que na mesma ocorra uma operação matemática que resulte na eliminação do(s) parâmetro(s) redundante(s) produzindo uma outra função $v(x; \boldsymbol{\beta})$ injetiva em Θ_v , originária de h_n , que não possua mais parâmetro(s) redundante(s). Considere $\mathcal{G}_i = \{g_i[x; (\boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_{n-1}, \boldsymbol{\beta})] : (\boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_{n-1}, \boldsymbol{\beta}) \in \Theta_{h_i, \dots, h_{n-1}, v}, x \in \mathbb{R}\}$ uma classe de funções injetivas em $\Theta_{h_i, \dots, h_{n-1}, v}$ em que $g_i[x; (\boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_{n-1}, \boldsymbol{\beta})] = h_i(x; \boldsymbol{\theta}_i) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ v(x; \boldsymbol{\beta})$, para todo $1 \leq i \leq (n-1)$. Sejam injetivas as funções f e h_i 's para todo $1 \leq i \leq (n-1)$.*

Seja uma outra fd pertencente a uma família $\mathcal{H}$ tal que

$$F_{\mathcal{H}}(x; \boldsymbol{\beta}) = h_1(x; \boldsymbol{\theta}_1) \circ h_2(x; \boldsymbol{\theta}_2) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ v(x; \boldsymbol{\beta}) \quad (3.5)$$

a função real v é a função parte paramétrica na expressão (3.5) e a dimensão de $\boldsymbol{\beta}$ é menor do que a dimensão de $\boldsymbol{\theta}_n$.

Então, a família $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ não identificável sofreu redução mínima até produzir uma família $\mathcal{H}$ identificável.

Demonstração. A demonstração é direta pelo Teorema 4 e também pela hipótese de não haver mais parâmetros redundantes na família $\mathcal{H}$. $\square$

Definição 12 (Vetores encaixados). *Seja a fd*

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = h_1(x; \boldsymbol{\theta}_1) \circ h_2(x; \boldsymbol{\theta}_2) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ h_n(x; \boldsymbol{\theta}_n) \quad (3.6)$$

em que $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n)$. O vetor $\boldsymbol{\theta}$ terá dimensão igual ao somatório das dimensões dos vetores $\boldsymbol{\theta}_i$'s.

Considere que da expressão (3.6) surja uma outra fd de uma família $\mathcal{H}$, com uma menor quantidade de parâmetros e as vezes com algum ou alguns parâmetros que não estavam em $F_{\mathcal{F}}$.

Define-se que todo vp ou parâmetro pertencente a fd da família $\mathcal{F}$ está encaixado em $\boldsymbol{\theta}$, inclusive o próprio $\boldsymbol{\theta}$ e que todo vp ou parâmetro pertencente a fd da família $\mathcal{H}$ está encaixado em $\boldsymbol{\theta}$.

A notação utilizada para expressar que $\boldsymbol{\delta}$ está encaixado em $\boldsymbol{\theta}$ é:

$$\boldsymbol{\delta} \in \boldsymbol{\theta}.$$

Lê-se: O vp δ está encaixado no vp θ .

Na área da estatística, os modelos de regressão encaixados são conceituados como aqueles nos quais um modelo é obtido mediante a imposição de restrições nos parâmetros do modelo inicial (DOBSON, 2018). Contudo, é importante salientar que, na Definição 12, o conceito de encaixe transcende significativamente a abordagem convencional dos modelos de regressão encaixados.

Para exemplificar as Definições 11 e 12 (Função parte paramétrica e vetores encaixados), Corolário 1 e o Teorema 4, tem-se o próximo exemplo.

Exemplo 3. *Considere a fd Weibull inversa generalizada (GiW), apresentada por Gusmão, Ortega e Cordeiro (2011), escrita como*

$$F_{\mathcal{F}}(x; \lambda, \alpha, \beta) = \exp \left[-\lambda \left(\frac{\alpha}{x} \right)^{\beta} \right] \quad (3.7)$$

e pode ser expressa, pela Definição 11, como composição de funções reais tal como

$$F_{\mathcal{F}}(x; \lambda, \alpha, \beta) = h_1(x) \circ h_2(x; \lambda, \alpha, \beta)$$

em que $h_1(x) = \exp[-x]$, $h_2(x; \lambda, \alpha, \beta) = \lambda \left(\frac{\alpha}{x} \right)^{\beta}$ e h_2 é a função parte paramétrica em $F_{\mathcal{F}}(x; \lambda, \alpha, \beta)$.

A função h_2 é não injetiva em Θ_{h_n} por apresentar os parâmetros redundantes λ e α , e verificar esta afirmação dar-se de forma direta. A afirmação deste parágrafo classifica $F_{\mathcal{F}}$ como não identificável pela Definição 9 e pelo Teorema 4.

Tomando, em h_2 , a relação $c = \lambda\alpha^{\beta}$ tem-se a função real $g(x; c, \beta) = \frac{c}{x^{\beta}}$ (que é uma função injetiva em Θ_g). Considere a composição

$$h_1(x) \circ g(x; c, \beta)$$

que produz uma função injetiva em Θ_g . Seja uma fd $F_{\mathcal{G}}(x; c, \beta)$ definida como

$$F_{\mathcal{G}}(x; c, \beta) = h_1(x) \circ g(x; c, \beta).$$

A fd $F_{\mathcal{G}}(x; c, \beta)$ pertence a família $\mathcal{G}$ e não a família $\mathcal{F}$. A família $\mathcal{F}$ é não identificável, mas a família $\mathcal{G}$ é identificável com base nas Definições 9, 11, Corolário 1 e pelo Teorema 4. Pela Definição 12 $(c, \beta) \in (\lambda, \alpha, \beta)$, $c \in (c, \beta)$ e $c \in (\lambda, \alpha, \beta)$.

A $F_{\mathcal{F}}(x; \lambda, \alpha, \beta)$ não identificável produziu, através do processo descrito, uma fd $F_{\mathcal{G}}(x; c, \beta)$ identificável. Assim, pela Corolário 1 houve uma redução mínima até identificabilidade paramétrica.

Por meio das Definições 9 e 11, juntamente com o Corolário 1 e o Teorema 4, é viável classificar a identificabilidade de uma fd e, desse modo, estabelecer um critério para a finalidade mencionada.

Corolário 2 (Não identificabilidade pela falta de injetividade em qualquer função que compõe a fd). *Considere a fd $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ da família $\mathcal{F}$ tal como descrita na Definição 11 e no Teorema 4. Considere pelo menos uma das h_i 's com $1 \leq i \leq n$ não injetiva, então o evp $\Theta_{\mathcal{F}}$ não é identificável.*

Demonstração. Pelas Definições 11 e 9 e pelo Teorema 4 é direto ver que $\mathcal{F}$ é não identificável. $\square$

O próximo corolário constitui um caso específico do Corolário 2 e tem por objetivo ressaltar a importância da injetividade da função baseline empregada nas classes geradoras de distribuições de probabilidade, especialmente no domínio de evp. A questão levantada será devidamente elucidada em detalhes posteriormente nesta tese.

Corolário 3 (Não identificabilidade pela falta de injetividade na função parte paramétrica da fd). *Considere a fd $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ da família $\mathcal{F}$ tal como descrita na Definição 11 e no Teorema 4. Considere a função h_n não injetiva, então o evp $\Theta_{\mathcal{F}}$ não é identificável.*

Demonstração. A demonstração é consequência direta do Teorema 4. $\square$

Em geral uma distribuição uniparamétrica é mais tratável matematicamente, porém este fato não garante que é identificável. O Corolário 3 a seguir estuda esta questão.

Corolário 4 (fd's uniparamétricas não identificáveis pela paridade). *Seja uma família de fd's uniparamétricas $\mathcal{F} = \{F_{\mathcal{F}}(x; \theta) : \theta \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$. Seja $F_{\mathcal{F}}(x; \cdot)$ uma função par com domínio no espaço paramétrico $\Theta_{\mathcal{F}} \subset \mathbb{R}$. Considere $-\theta, \theta \in \Theta_{\mathcal{F}}$, então $\mathcal{F}$ é uma família não identificável.*

Demonstração. A demonstração é consequência direta do Corolário 2. $\square$

Corolário 5 (Identificabilidade em fd's com funções partes paramétricas uniparamétricas). *Considere a fd $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ da família $\mathcal{F}$ tal como descrita na Definição 11 e no Teorema 4, porém com a função h_n possuindo um único parâmetro. Então, o evp $\Theta_{\mathcal{F}}$ é identificável.*

Demonstração. A demonstração é consequência direta do Teorema 4. $\square$

Corolário 6 (Identificabilidade em fd's uniparamétricas). *Considere a fd $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ da família $\mathcal{F}$ tal como descrita na Definição 11 e no Teorema 4, porém com a função h_n sendo a única que possui parâmetro e um único parâmetro. Então, o ep $\Theta_{\mathcal{F}}$ é identificável.*

Demonstração. A demonstração é consequência direta do Teorema 4. $\square$

Ter o menor número possível de parâmetros em uma *fd* paramétrica, melhor ainda se for um único parâmetro, em termos da identificabilidade é interessante. Todavia nem sempre o conjunto de dados pode ser ajustado por uma *fd* uniparamétrica.

Definição 13 (Vetores paramétricos ampliados). *Seja a fd $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ e considere a introdução de parâmetro(s) em $F_{\mathcal{F}}$ tal que surja uma outra fd $F_{\mathcal{G}}(x; \boldsymbol{\delta})$ pertencente a família $\mathcal{G}$. Define-se que $\boldsymbol{\theta}$ sofreu uma ampliação paramétrica produzindo $\boldsymbol{\delta}$ e será utilizada a notação:*

$$\boldsymbol{\delta} \ni \boldsymbol{\theta}.$$

Lê-se: o vp $\boldsymbol{\delta}$ é uma ampliação do vp $\boldsymbol{\theta}$.

Definição 14 (Subfamília paramétrica). *Sejam $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ e $\mathcal{G} = \{F_{\mathcal{G}}(x; \boldsymbol{\beta}) : \boldsymbol{\beta} \in \Theta_{\mathcal{G}}, x \in \mathbb{R}\}$ duas famílias de *fd*'s paramétricas. Se $\boldsymbol{\alpha} \subseteq \boldsymbol{\beta}$ define-se que $\mathcal{F}$ é subfamília de $\mathcal{G}$ ou se $\boldsymbol{\alpha} \ni \boldsymbol{\beta}$ define-se que $\mathcal{G}$ é subfamília de $\mathcal{F}$.*

Como consequência direta da Definição 14 tem-se a Proposição 3.

Proposição 3 (Toda família é subfamília dela própria). *Toda e qualquer família de *fd*'s é subfamília dela própria.*

Demonstração. Pelas Definições 12 e 14 fica demonstrado que toda e qualquer família de *fd*'s é subfamília dela própria. $\square$

Proposição 4 (Encaixe pela ampliação). *Sejam $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ e $F_{\mathcal{G}}(x; \boldsymbol{\delta})$ tais que $\boldsymbol{\delta} \ni \boldsymbol{\theta}$ como na Definição 13, então $\boldsymbol{\theta} \subseteq \boldsymbol{\delta}$.*

Demonstração. Pela hipótese $\boldsymbol{\delta} \ni \boldsymbol{\theta}$ e pela Definição 14 $\mathcal{F}$ é subfamília de $\mathcal{G}$, então $\boldsymbol{\theta} \subseteq \boldsymbol{\delta}$ pela Definição 14.

Portanto, fica demonstrado a Proposição 4. $\square$

Definição 15 (Redução paramétrica). *Sejam $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ e $F_{\mathcal{H}}(x; \boldsymbol{\delta})$ *fd*'s tais que $\boldsymbol{\delta} \subseteq \boldsymbol{\theta}$. Define-se que $\boldsymbol{\theta}$ sofreu uma redução paramétrica produzindo $\boldsymbol{\delta}$.*

O Exemplo 4 utiliza as Definições 11, 12, 14, 15 e Teorema 4.

Exemplo 4. *Seja $\mathcal{G} = \{F_{\mathcal{G}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{G}}, x \in \mathbb{R}\}$ a família de *fd*'s Weibull generalizada exponenciada (EGW) estudada por Oguntunde, Odetunmbi e Adejumo (2015) com $\boldsymbol{\theta} = (a, b, \alpha, \beta)$ sendo o vp da EGW e $F_{\mathcal{G}}(x; \boldsymbol{\theta})$ expressa por*

$$F_{\mathcal{G}}(x; \boldsymbol{\theta}) = \left\{ 1 - \exp \left[-a (\alpha x)^{\beta} \right] \right\}^b \quad (3.8)$$

em que a , b e β são os parâmetros de forma e α é o parâmetro de escala, todos pertencentes ao conjunto $\mathbb{R}$ positivos. A fd $F_G(x; \boldsymbol{\theta})$ possui suporte em $x \in (0, \infty)$.

A expressão (3.8) pode ser escrita, pela Definição 11, como uma composição de funções reais tal como

$$F_G(x; \boldsymbol{\theta}) = h_1(x; b) \circ h_2(x; a, \alpha, \beta) \quad (3.9)$$

em que $h_1(x; b) = \{1 - \exp[-x]\}^b$ e $h_2(x; a, \alpha, \beta) = a(\alpha x)^\beta$. Claramente a função real h_2 é não injetiva em Θ_{h_2} e pelo Teorema 4, F_G é não identificável. Também pelos Corolários 2 e 3 fica evidente a não identificabilidade de F_G .

No artigo de [Gusmão et al. \(2021\)](#) foi realizada uma redução paramétrica (Definição 15) na distribuição EGW que conduziu a uma outra distribuição de probabilidade, a Weibull exponenciada (EW) estudada por [Nadarajah, Cordeiro e Ortega \(2013\)](#), cujo processo será exibido no próximo parágrafo.

Considere a função parte paramétrica h_2 da expressão 3.9 e tome $c = \alpha a^{\frac{1}{\beta}}$ com $c > 0$ e fixado. Dessa forma uma outra fd $F_{\mathcal{F}}(x; \boldsymbol{\delta})$ advinda de $F_G(x; \boldsymbol{\theta})$ surgirá. Esta outra distribuição pode ser expressa, pela Definição 11, como segue:

$$F_{\mathcal{F}}(x; \boldsymbol{\delta}) = h_1(x; b) \circ g(x; c, \beta) = \{1 - \exp[-(cx)^\beta]\}^b \quad (3.10)$$

em que $\boldsymbol{\delta} = (b, c, \beta)$, $h_1(x; b) = \{1 - \exp[-x]\}^b$ e $g(x; c, \beta) = (cx)^\beta$.

A função g é injetiva em Θ_g e a composição $h_1(x; b) \circ g(x; c, \beta)$ é injetiva em Θ_F , pelo Corolário 1 e pelo Teorema 4, $F_{\mathcal{F}}$ é identificável.

Pela Definição 12 $\boldsymbol{\delta} \in \boldsymbol{\theta}$ e pela Definição 14 a família $\mathcal{F}$ é subfamília da família $\mathcal{G}$. Ainda, pela Definição 15 o vp $\boldsymbol{\theta}$ sofreu uma redução paramétrica produzindo $\boldsymbol{\delta}$.

O Exemplo 5 utiliza as Definições 12, 14 e 15.

Exemplo 5. Considere $\mathcal{G} = \{F_G(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_G, x \in \mathbb{R}\}$ a família de fd's Weibull triparamétrica e seja $F_G(x; \boldsymbol{\theta})$ dada por

$$F_G(x; \alpha, \lambda, \beta) = 1 - \exp\left\{-[\alpha(x - \lambda)]^\beta\right\} \quad (3.11)$$

em que $\boldsymbol{\theta} = (\alpha, \lambda, \beta)$ com α , β e λ sendo parâmetros reais positivos de escala, locação e forma, respectivamente e $x \in (0, \infty)$. Tomando $\lambda = 0$ encontra-se a família $\mathcal{H} = \{F_{\mathcal{H}}(x; \boldsymbol{\delta}) : \boldsymbol{\delta} \in \Theta_{\mathcal{H}}, x \in \mathbb{R}\}$ de fd's Weibull biparamétrica que é subfamília da família $\mathcal{G}$ (pela Definição 14) e pode ser expressa por

$$F_{\mathcal{H}}(x; \alpha, \beta) = 1 - \exp\left\{-[\alpha x]^\beta\right\} \quad (3.12)$$

em que $\boldsymbol{\delta} = (\alpha, \beta)$.

Consequentemente, $\delta \in \theta$ pela Definição 12 e pela Definição 15 o vp θ sofreu uma redução paramétrica produzindo δ .

Definição 16 (Função distribuição uniparamétrica). *Seja a fd*

$$F_{\mathcal{F}}(x; \theta) = h_1(x) \circ h_2(x) \circ \dots \circ h_{n-1}(x) \circ h_n(x; \theta)$$

expressa no molde da Definição 11. Define-se que $F_{\mathcal{F}}(x; \theta)$ é uma fd uniparamétrica.

Considere agora uma fd paramétrica que passa por um processo de redução, resultando em uma distribuição caracterizada por um único parâmetro. Este é o objeto de estudo abordado na Definição 17.

Definição 17 (Redução paramétrica até tornar-se uniparamétrica). *Seja a fd*

$$F_{\mathcal{F}}(x; \theta) = h_1(x; \theta_1) \circ h_2(x; \theta_2) \circ \dots \circ h_{n-1}(x; \theta_{n-1}) \circ h_n(x; \theta_n)$$

em que $\theta = (\theta_1, \theta_2, \dots, \theta_{n-1}, \theta_n)$. Considere que em $F_{\mathcal{F}}(x; \theta)$ foi realizada uma operação matemática e que como resultado deste procedimento surja uma outra fd $F_{\mathcal{G}}(x; \delta)$ pertencente a família $\mathcal{G}$. Define-se que θ sofreu uma redução paramétrica até tornar-se uniparamétrica.

A seguir será exibida uma redução que produz uma fd não-paramétrica.

Definição 18 (Redução máxima até não ser mais paramétrica). *Seja a fd*

$$F_{\mathcal{F}}(x; \theta) = h_1(x; \theta_1) \circ h_2(x; \theta_2) \circ \dots \circ h_{n-1}(x; \theta_{n-1}) \circ h_n(x; \theta_n)$$

em que $\theta = (\theta_1, \theta_2, \dots, \theta_{n-1}, \theta_n)$. Considere que em $F_{\mathcal{F}}(x; \theta)$ foram atribuídos valores para todos os parâmetros de θ e como resultado deste procedimento surja uma outra fd $F_{\mathcal{G}}(x)$ pertencente a família $\mathcal{G}$. Define-se que $F_{\mathcal{F}}(x; \theta)$ sofreu uma redução máxima e produziu uma outra distribuição $F_{\mathcal{G}}(x)$ que não é mais paramétrica.

O caso descrito na Definição 18 pode ser ilustrado quando um pesquisador, ao atribuir valores específicos para a média e a variância, opta por utilizar tal conhecimento na modelagem de um experimento por meio da distribuição normal. Nesse cenário, a distribuição deixa de ser paramétrica, transformando-se em uma distribuição não paramétrica, isto é, desprovida de parâmetros.

A próxima definição consiste na contrapositiva de uma implicação direta decorrente da Definição 9.

Definição 19 (Igualdade identificável de funções distribuição). *Considere $\mathcal{F}$ uma família de fd's. Se $F_{\mathcal{F}}(x; \theta) = F_{\mathcal{F}}(x; \beta)$, para todo $x \in \mathbb{R}$, implicar em $\theta = \beta$, diz-se que $\mathcal{F}$ é uma família identificável.*

Vetores paramétricos iguais na mesma família de distribuições, seja ela identificável ou não, sempre produzem fd 's iguais. É do que trata o lema a seguir:

Lema 1 (Igualdade de vetores implicam em igualdade de distribuições). *Considere $\mathcal{F}$ uma família de fd 's e sejam θ e β pertencentes ao evp de $\mathcal{F}$ tais que $\theta = \beta$, então*

$$F_{\mathcal{F}}(x; \theta) = F_{\mathcal{F}}(x; \beta) \quad \text{para todo } x \in \mathbb{R}.$$

Demonstração. Pelas definições de função real e de igualdade de funções (descritas em Lima (2008)) e considerando $\theta = \beta \in \Theta_{\mathcal{F}}$ com

$$f : \Theta_{\mathcal{F}} \rightarrow \mathcal{F} \quad \text{para cada } x \in \mathbb{R}$$

em que f é uma função real, então que $F_{\mathcal{F}}(x; \theta) = F_{\mathcal{F}}(x; \beta)$ para cada $x \in \mathbb{R}$.

Portanto, fica assim provado o Lema 1. □

Este resultado reveste-se de significativa importância, uma vez que é inviável investigar a identificabilidade de uma família de distribuições quando se dispõe apenas de vetores paramétricos idênticos.

Nota: Lima (2008) se refere à regra que associa cada elemento de um conjunto a um único elemento de outro conjunto como "regra de correspondência da função". Na literatura matemática, é comum utilizar o termo "mapeamento" para denotar essa regra de relacionamento entre conjuntos.

Quando depara-se com fd 's idênticas, ou seja, pertencentes à mesma família e resultando em vp 's iguais, estas funções são consideradas identificáveis. Embora essa afirmação possa parecer similar ao Lema 1, uma sutil diferença é discernível, resultando em uma consequência importante.

Ao analisar a Definição 19 e o Lema 1, a discrepância entre ambos torna-se evidente. Enquanto o Lema 1 não é suficiente para afirmar a identificabilidade, a Definição 19 possibilita o estudo da identificabilidade de uma família de distribuições.

Exemplo 6. Na introdução do artigo *Silveira et al. (2023)* é afirmado que se

$$F_{\mathcal{F}}(x; \theta) = F_{\mathcal{F}}(x; \beta), \quad \text{para cada } x \in \mathbb{R} \Leftrightarrow \theta = \beta,$$

então a família $\mathcal{F}$ é identificável.

Porém, para afirmar que $\mathcal{F}$ é identificável basta que atenda a afirmação

$$F_{\mathcal{F}}(x; \theta) = F_{\mathcal{F}}(x; \beta), \quad \text{para cada } x \in \mathbb{R} \Rightarrow \theta = \beta \tag{3.13}$$

de acordo com a Definição 19.

Pelo Lema 1, a afirmação

$$\boldsymbol{\theta} = \boldsymbol{\beta} \Rightarrow F_{\mathcal{F}}(x; \boldsymbol{\theta}) = F_{\mathcal{F}}(x; \boldsymbol{\beta}), \forall x \in \mathbb{R} \quad (3.14)$$

é verdade sempre e por isto não consegue-se afirmar que a família $\mathcal{F}$ é identificável ou não pela afirmação 3.14.

Se existem dois vp 's ($\boldsymbol{\theta}$ e $\boldsymbol{\beta}$) que diferem entre si em pelo menos duas entradas (coordenadas), ambos pertencentes à mesma família de fd 's $\mathcal{F}$, não há garantia de que $F_{\mathcal{F}}(x; \boldsymbol{\theta}) \neq F_{\mathcal{F}}(x; \boldsymbol{\beta})$ para algum valor de x pertencente ao domínio de $F_{\mathcal{F}}$. O Exemplo 7 ilustra essa situação para o caso de duas coordenadas distintas, corroborando o que foi afirmado neste parágrafo.

Exemplo 7. Seja $\mathcal{F}$ a família de fd 's McDonald generalizada (Mc-G), apresentada por Alexander et al. (2012). A fd Mc-G é dada por:

$$F_{\mathcal{F}}(x; a, b, c) = \frac{1}{B(a, b)} \int_0^{[G(x)]^c} t^{a-1} (1-t)^{b-1} dt \quad (3.15)$$

em que G é a fd (baseline) que pertence a uma família $\mathcal{G}$ qualquer; a , b e c são parâmetros reais positivos da expressão 3.15, x pertence ao domínio de G , $t \in (0, 1)$ e $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$.

Seja G a fd da família GiW (GUSMÃO; ORTEGA; CORDEIRO, 2011). A fd McDonald Weibull inversa generalizada (Mc-GiW), apresentada em Shahbaz, Ismail e Butt (2016), é resultado da composição da fd da família Mc-G com a fd da família GiW em que o vp da fd Mc-GiW é dado por $\boldsymbol{\beta} = (a, b, c, \alpha, \beta, \lambda)$. Seja $\mathcal{H} = \{F_{\mathcal{H}}(x; \boldsymbol{\beta}) : \boldsymbol{\beta} \in \Theta_{\mathcal{H}}, x \in \mathbb{R}\}$ a família de fd 's Mc-GiW com $F_{\mathcal{H}}(x; \boldsymbol{\beta})$ dada por:

$$F_{\mathcal{H}}(x; \boldsymbol{\beta}) = \frac{1}{B(a, b)} \int_0^{\{\exp[-\lambda(\alpha/x)^{\beta}]\}^c} t^{a-1} (1-t)^{b-1} dt. \quad (3.16)$$

Tomando $b = 1$ na expressão (3.16) (redução paramétrica) consequentemente a referida expressão poderá ser escrita como segue:

$$F_{\mathcal{N}}(x; \boldsymbol{\psi}) = a \int_0^{\{\exp[-\lambda(\alpha/x)^{\beta}]\}^c} t^{a-1} dt = \left\{ \left\{ \exp[-\lambda(\alpha/x)^{\beta}] \right\}^c \right\}^a \quad (3.17)$$

em que $\boldsymbol{\psi} \subseteq \boldsymbol{\beta}$ pela Definição 12, $\mathcal{N} = \{F_{\mathcal{N}}(x; \boldsymbol{\psi}) : \boldsymbol{\psi} \in \Theta_{\mathcal{N}}, x \in \mathcal{B}(\mathbb{R})\}$ é uma outra família de fd 's a qual é subfamília da família $\mathcal{H}$, pela Definição 14, e $\boldsymbol{\psi} = (a, c, \beta, \alpha, \lambda)$.

Sejam $\boldsymbol{\delta} = (a, c, \alpha_l, \beta, \lambda_l)$, $\boldsymbol{\psi} = (a, c, \alpha_j, \beta, \lambda_j) \in \Theta_{\mathcal{N}}$, em que $l, j \in I$ e I é um certo conjunto de índices, tais que $\alpha_l \neq \alpha_j$ e $\lambda_l \neq \lambda_j$ para todo $l \neq j$, $l, j \in I$ e I é um certo conjunto de índices; então tem-se que $\boldsymbol{\delta} \neq \boldsymbol{\psi}$.

Pela hipótese, pode-se afirmar que

$$\exists \alpha_l \neq \alpha_j \text{ e } \lambda_l \neq \lambda_j : -\lambda_l (\alpha_l/x)^\beta = -\lambda_j (\alpha_j/x)^\beta$$

e pode-se escrever

$$\begin{aligned} -c\alpha_l (\alpha_l/x)^\beta &= -c\alpha_j (\alpha_j/x)^\beta \Rightarrow \\ \left\{ \left[\exp \left[-\lambda_l (\alpha_l/x)^\beta \right] \right]^c \right\}^a &= \left\{ \left[\exp \left[-\lambda_j (\alpha_j/x)^\beta \right] \right]^c \right\}^a \Rightarrow \\ F_{\mathcal{N}}(x; a, c, \alpha_l, \lambda_l, \beta) &= F_{\mathcal{N}}(x; a, c, \alpha_j, \lambda_j, \beta) \Rightarrow \\ F_{\mathcal{H}}(x; a, b=1, c, \alpha_l, \lambda_l, \beta) &= F_{\mathcal{H}}(x; a, b=1, c, \alpha_j, \lambda_j, \beta) \Rightarrow \\ F_{\mathcal{H}}(x; a, b, c, \alpha_l, \lambda_l, \beta) &= F_{\mathcal{H}}(x; a, b, c, \alpha_j, \lambda_j, \beta). \end{aligned} \quad (3.18)$$

Foi verificado que se dois vp 's δ e ψ que diferem entre si por duas coordenadas, ambos pertencentes a mesma família $\mathcal{H}$ de fd 's, os mesmos podem produzir $F_{\mathcal{H}}(x; \delta) = F_{\mathcal{H}}(x; \psi)$ para todo x pertencente ao domínio de $F_{\mathcal{H}}$.

Se ocorrer uma redução paramétrica e for realizada uma comparação entre a fd inicial e a mais recente (reduzida), é esperado que a fd reduzida apresente uma maior parcimônia e uma menor flexibilidade em relação à fd inicial.

A parcimônia é assegurada pela definição de parcimônia paramétrica, conforme descrita neste capítulo, e, se a fd inicial for não identificável, a redução paramétrica pode resultar em uma nova fd identificável ou, dependendo da regra de correspondência da distribuição, a não identificabilidade pode persistir.

Exemplo 8. Considere uma fd que inicialmente era não identificável e sofreu uma redução paramétrica, porém essa redução não resultou em uma outra fd identificável. Eugene, Lee e Famoye (2002) propuseram a fd beta normal, a qual pertence à subfamília da família de distribuições Mc-Normal. Posteriormente, foi proposta a família beta generalizada (Beta-G). A fd pertencente à família Beta-G é expressa por:

$$F_{\mathcal{W}}(x; a, b) = \frac{1}{B(a, b)} \int_0^{G(x)} t^{a-1} (1-t)^{b-1} dt \quad (3.19)$$

em que $\mathcal{W} = \{F_{\mathcal{W}}(x; \beta) : \beta \in \Theta_{\mathcal{W}}, x \in \mathbb{R}\}$ (Beta-G) é uma subfamília da família Mc-G.

A fd da família Mc-GiW (expressão 3.15) sofrendo uma redução paramétrica (tomando $c = 1$) produz uma fd mais parcimoniosa conhecida como Beta generalizada inversa Weibull (Beta-GiW) representada por $F_{\mathcal{F}}(x; a, b, \alpha, \beta, \lambda)$ e pode ser expressa por

$$F_{\mathcal{F}}(x; \alpha) = \frac{1}{B(a, b)} \int_0^{\left[\exp \left[-\lambda (\alpha/x)^\beta \right] \right]} t^{a-1} (1-t)^{b-1} dt \quad (3.20)$$

em que $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ com $\text{vp } \boldsymbol{\alpha} = (a, b, \alpha, \beta, \lambda)$ e $x \in (0, \infty)$.

Pela discussão do Exemplo 7 a família $\mathcal{F}$ é não identificável, mesmo sendo mais parcimoniosa do que a família Mc-GiW. A redução paramétrica aplicada na distribuição Mc-GiW não conseguiu produzir uma distribuição identificável.

Uma fd não identificável em que a redução na quantidade de parâmetros produz uma outra fd identificável é a situação exibida no exemplo seguinte.

Exemplo 9. Considere a fd da iW que tem recebido alguma atenção na literatura estatística. Keller, Kamath e Perera (1982) estudaram as formas da fdp e da função de risco desta distribuição inversa. Voltando ao exemplo, inserindo a fd iW na classe Kumaraswamy-G (Kum-G) apresentada por Cordeiro e Castro (2011), tem-se que

$$F_{\mathcal{M}}(x; a, b, c, \beta) = 1 - \left\{ 1 - \exp \left[-a \left(\frac{c}{x} \right)^{\beta} \right] \right\}^b = h_1(x; b) \circ h_2(x; a, c, \beta) \quad (3.21)$$

pela Definição 11 pode-se escrever

$$h_1(x; b) = 1 - \{1 - x\}^b \quad e \quad h_2(x; a, c, \beta) = \exp \left[-a \left(\frac{c}{x} \right)^{\beta} \right]$$

em que $\boldsymbol{\theta} = (a, b, c, \beta)$ é o vetor de parâmetros da expressão (3.21) com a, b, c e β sendo parâmetros reais positivos e $x \in (0, \infty)$. É direto ver que a função h_2 é não injetiva no vp .

Seja $\mathcal{M} = \{F_{\mathcal{M}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{M}}, x \in \mathbb{R}\}$ a família de fd's da Kumaraswamy inversa Weibull (Kum- iW). A fd 3.21 foi introduzida por Shahbaz, Shahbaz e Butt (2012). É direto ver a distribuição Kumaraswamy inversa Weibull é não identificável.

Procedendo uma redução paramétrica (tomando $d = ca^{\frac{1}{\beta}}$ com $d > 0$) surge a distribuição Kumaraswamy reparametrizada inversa Weibull (Kum-RiW), a qual recebe este nome apenas para fins de distinção da distribuição Kum- iW . Seja a fd Kum-RiW expressa por

$$F_{\mathcal{N}}(x; \boldsymbol{\psi}) = 1 - \left\{ 1 - \exp \left[- \left(\frac{d}{x} \right)^{\beta} \right] \right\}^b = h_1(x; b) \circ g(x; d, \beta) \quad (3.22)$$

em que $h_1(x; b) = 1 - \{1 - x\}^b$, $g(x; d, \beta) = \exp \left[- \left(\frac{d}{x} \right)^{\beta} \right]$, $\boldsymbol{\psi} = (b, d, \beta)$ é o vp da expressão (3.22); b, d e β são parâmetros reais positivos e $\mathcal{N} = \{F_{\mathcal{N}}(x; \boldsymbol{\psi}) : \boldsymbol{\psi} \in \Theta_{\mathcal{N}}, x \in \mathbb{R}\}$ e é importante salientar que $\mathcal{N}$ é subfamília da família $\mathcal{M}$, pela Definição 14.

Sejam $\boldsymbol{\psi} = (b_l, d_l, \beta_l)$ e $\boldsymbol{\rho} = (b_j, d_j, \beta_j) \in \Theta_{\mathcal{N}}$ tais que $\boldsymbol{\psi} \neq \boldsymbol{\rho}$ para todo $l \neq j$; $l, j \in I$ e I é um certo conjunto de índices. Os dois vp 's, citados neste parágrafo, diferem

Tabela 1 – Prova que $\Theta_{\mathcal{N}}$ é identificável

Casos	Hipóteses $\psi \neq \rho$	Cadeia de implicações
1	$b_l \neq b_j, d_l = d_j$ e $\beta_l = \beta_j$	$F_{\mathcal{N}}(x, \psi) \neq F_{\mathcal{N}}(x, \rho)$ para algum $x > 0$
2	$b_l = b_j, d_l = d_j$ e $\beta_l \neq \beta_j$	$F_{\mathcal{N}}(x, \psi) \neq F_{\mathcal{N}}(x, \rho)$ para algum $x > 0$
3	$b_l \neq b_j, d_l \neq d_j$ e $\beta_l = \beta_j$	$F_{\mathcal{N}}(x, \psi) \neq F_{\mathcal{N}}(x, \rho)$ para algum $x > 0$
4	$b_l = b_j, d_l \neq d_j$ e $\beta_l \neq \beta_j$	$F_{\mathcal{N}}(x, \psi) \neq F_{\mathcal{N}}(x, \rho)$ para algum $x > 0$
5	$b_l \neq b_j, d_l \neq d_j$ e $\beta_l = \beta_j$	$F_{\mathcal{N}}(x, \psi) \neq F_{\mathcal{N}}(x, \rho)$ para algum $x > 0$
6	$b_l \neq b_j, d_l = d_j$ e $\beta_l \neq \beta_j$	$F_{\mathcal{N}}(x, \psi) \neq F_{\mathcal{N}}(x, \rho)$ para algum $x > 0$
7	$b_l \neq b_j, d_l \neq d_j$ e $\beta_l \neq \beta_j$	$F_{\mathcal{N}}(x, \psi) \neq F_{\mathcal{N}}(x, \rho)$ para algum $x > 0$

de sete formas distintas. Estas formas e as respectivas implicações disto são expostas na Tabela 1.

Logo, pela Tabela 1 tem-se que $\Theta_{\mathcal{N}}$ é identificável.

Lema 2. Considere a função parte paramétrica h_n tal como descrita no Teorema 4, se h_n puder ser expressa tal como segue

$$h_n(x; \theta_n) = u(x; \alpha) \circ v(x; \beta).$$

Seja $\mathcal{V} = \{v(x; \beta) : \beta \in \Theta_{\mathcal{V}}, x \in \mathbb{R}\}$ uma família de funções reais e seja f uma função real tal que

$$f : \Theta_{\mathcal{V}} \rightarrow \mathcal{V}.$$

Seja u uma função real tal que

$$u : \mathcal{V} \rightarrow \mathcal{H}$$

em que $\mathcal{H}$ é descrita no Teorema 4. Seja o vp $\theta_n = (\alpha, \beta)$ e $\beta \in \theta_n$, então v é uma função injetiva com domínio no evp $\Theta_{\mathcal{V}}$.

Demonstração. Sejam $\theta_{n_1}, \theta_{n_2} \in \Theta_{h_n}$ tais que $\theta_{n_1} = (\alpha, \beta_1)$, $\theta_{n_2} = (\alpha, \beta_2)$, β_1 e $\beta_2 \in \Theta_{\mathcal{V}}$ e $\beta_1 \neq \beta_2$. Por hipótese $h_n(x; \theta_n)$ é injetiva em Θ_{h_n} , então

$$\theta_{n_1} \neq \theta_{n_2} \Rightarrow h_n(x; \theta_{n_1}) = u(x; \alpha) \circ v(x; \beta_1) \neq h_n(x; \theta_{n_2}) = u(x; \alpha) \circ v(x; \beta_2), \quad (3.23)$$

as funções h 's diferem em um conjunto de medida de Lebesgue não nulo.

Segue da sentença 3.23 que

$$v(x; \beta_1) \neq v(x; \beta_2) \Rightarrow u[v(x; \beta_1); \alpha] \neq u[v(x; \beta_2); \alpha],$$

as funções u 's diferem em um conjunto de medida de Lebesgue não nulo e u é uma função injetiva em $\mathcal{V}$.

Segue da sentença 3.2 que

$$\beta_1 \neq \beta_2 \Rightarrow v(x; \beta_1) \neq v(x; \beta_2),$$

as funções v 's diferem em um conjunto de medida de Lebesgue não nulo e v é uma função injetiva em $\Theta_{\mathcal{V}}$.

Portanto, as funções u , v e f são injetivas em seus respectivos domínios. $\square$

Lema 3. *Seja $\mathcal{G}_i = \{g_i[x; (\theta_i, \dots, \theta_n)] : (\theta_i, \dots, \theta_n) \in \Theta_{h_i, \dots, h_n}, x \in \mathbb{R}\}$ uma classe de funções injetivas em $\Theta_{h_i, \dots, h_n}$, com $1 \leq i \leq (n-1)$. Sejam injetivas as funções f e h_i 's com $1 \leq i \leq n$, tais como no Teorema 4. A função real v é injetiva pelo Lema 2.*

Considere $\mathcal{F}_i = \{f_i[x; (\theta_i, \dots, \theta_{n-1}, \beta)] : (\theta_i, \dots, \theta_{n-1}, \beta) \in \Theta_{\mathcal{F}_i}, x \in \mathbb{R}\}$ uma classe de funções reais compostas com domínio em $\Theta_{\mathcal{F}_i} = \Theta_{h_i, \dots, h_n, \mathcal{V}}$ com $1 \leq i \leq (n-1)$ e

$$f_i[x; (\theta_i, \dots, \theta_{n-1}, \beta)] = h_i(x; \theta_i) \circ \dots \circ h_{n-1}(x; \theta_{n-1}) \circ v(x; \beta),$$

então a classe $\mathcal{F}_i$ (para todo $1 \leq i \leq (n-1)$) é uma família de funções compostas injetivas.

Demonstração. Por hipótese a composição

$$h_i(x; \theta_i) \circ \dots \circ h_n(x; \theta_n) \tag{3.24}$$

é injetiva.

Pela sentença 3.24 e pelo Lema 2 a função composta

$$h_i(x; \theta_i) \circ \dots \circ h_{n-1}(x; \theta_{n-1}) \circ u(x; \alpha) \circ v(x; \beta) \tag{3.25}$$

é injetiva.

Sejam $\theta_{n_1} = (\alpha, \beta_1)$, $\theta_{n_2} = (\alpha, \beta_2) \in \Theta_{h_n}$ tais que $\theta_{n_1} \neq \theta_{n_2}$ (ver 2). Então, (pela hipótese) tem-se que

$$\theta_{n_1} \neq \theta_{n_2} \Rightarrow h_i(x; \theta_i) \circ \dots \circ h_{n_1}(x; \theta_{n_1}) \neq h_i(x; \theta_i) \circ \dots \circ h_{n_2}(x; \theta_{n_2}). \tag{3.26}$$

Pela sentença 3.26 tem-se

$$\theta_{n_1} \neq \theta_{n_2} \Rightarrow h_i(x; \theta_i) \circ \dots \circ u(x; \alpha) \circ v(x; \beta_1) \neq h_i(x; \theta_i) \circ \dots \circ u(x; \alpha) \circ v(x; \beta_2). \tag{3.27}$$

Sejam $\delta_{n_1} = (\theta_1, \dots, \theta_{n-1}, \beta_1)$ e $\delta_{n_2} = (\theta_1, \dots, \theta_{n-1}, \beta_2)$ em que $\delta_1, \delta_2 \in \Theta_{\mathcal{F}_i}$. Pela sentença 3.27 tem-se

$$\delta_1 \neq \delta_2 \Rightarrow h_i(x; \theta_i) \circ \dots \circ h_{n-1}(x; \theta_{n-1}) \circ v(x; \beta_1) \neq h_i(x; \theta_i) \circ \dots \circ h_{n-1}(x; \theta_{n-1}) \circ v(x; \beta_2)$$

em que $\delta_1, \delta_2 \in \Theta_{\mathcal{F}_i}$.

Portanto, a classe $\mathcal{F}_i$ é injetiva para todo $1 \leq i \leq (n-1)$. $\square$

Teorema 5 (Subfamília identificável). *Sejam $\mathcal{F}$ e $\mathcal{G}$ duas famílias de fd's expressas por*

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = h_1(x; \boldsymbol{\theta}_1) \circ h_2(x; \boldsymbol{\theta}_2) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ h_n(x; \boldsymbol{\theta}_n) \quad (3.28)$$

e

$$F_{\mathcal{G}}(x; \boldsymbol{\theta}) = h_1(x; \boldsymbol{\theta}_1) \circ h_2(x; \boldsymbol{\theta}_2) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ v(x; \boldsymbol{\beta}). \quad (3.29)$$

Seja $F_{\mathcal{F}}$ identificável e seja $F_{\mathcal{G}}$ tal como nos Lemas 2 e 3, então $\mathcal{G}$ é identificável e é subfamília de $\mathcal{F}$.

Demonstração. É direto ver que $\mathcal{G}$ é subfamília de $\mathcal{F}$ pela Definição 14. Pelos Lemas 2 e 3, e pelo Teorema 4 é direto ver que $\mathcal{G}$ é identificável. $\square$

Considere uma outra situação em que $\mathcal{G}$ é não identificável e $\mathcal{G}$ é uma subfamília da família $\mathcal{F}$. Será que neste caso $\mathcal{F}$ é não identificável? Esta situação é analisada no Teorema 6.

Teorema 6 (Não identificabilidade pelo acréscimo de parâmetros na fd não identificável). *Seja $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ uma família de fd's e $\mathcal{G} = \{F_{\mathcal{G}}(x; \boldsymbol{\phi}) : \boldsymbol{\phi} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ uma subfamília não identificável da família $\mathcal{F}$, então a família $\mathcal{F}$ é não identificável.*

Demonstração. Sejam os vp's $\boldsymbol{\phi} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{n-1}, \boldsymbol{\delta})$ e $\boldsymbol{\xi} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{n-1}, \boldsymbol{\eta})$ com $\boldsymbol{\phi}, \boldsymbol{\xi} \in \Theta_{\mathcal{G}}$ e $\boldsymbol{\delta} \neq \boldsymbol{\eta}$, então $\boldsymbol{\phi} \neq \boldsymbol{\xi}$.

Pela hipótese a família $\mathcal{G}$ é não identificável. Considerando afirmação do parágrafo anterior e a não identificabilidade de $\mathcal{G}$ segue que

$$\begin{aligned} \boldsymbol{\delta} \neq \boldsymbol{\eta} &\Rightarrow \boldsymbol{\phi} \neq \boldsymbol{\xi} \Rightarrow F_{\mathcal{G}}(x, \boldsymbol{\phi}) = F_{\mathcal{G}}(x, \boldsymbol{\xi}) \Rightarrow \\ &h_1(x; \boldsymbol{\theta}_1) \circ h_2(x; \boldsymbol{\theta}_2) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ g(x; \boldsymbol{\delta}) = \\ &h_1(x; \boldsymbol{\theta}_1) \circ h_2(x; \boldsymbol{\theta}_2) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ g(x; \boldsymbol{\eta}) \Rightarrow \\ &g(x; \boldsymbol{\delta}) = g(x; \boldsymbol{\eta}) \end{aligned}$$

para todo x pertencente ao domínio de g .

Considere que em g foi realizada uma operação matemática de composição na qual como resultado deste procedimento surge a função real $h_n(x; \cdot)$ com domínio nos parâmetros pertencentes a h_n (tal como no Lema 3) e expressa como

$$h_n(x; \boldsymbol{\theta}_n) = u(x; \boldsymbol{\nu}) \circ g(x; \boldsymbol{\delta})$$

em que u e h_n são funções reais, $\boldsymbol{\theta}_n = (\boldsymbol{\nu}, \boldsymbol{\delta})$ e como consequencia de h_n surja uma outra fd dada por

$$F_{\mathcal{F}}(x, \boldsymbol{\theta}) = h_1(x; \boldsymbol{\theta}_1) \circ h_2(x; \boldsymbol{\theta}_2) \circ \dots \circ h_{n-1}(x; \boldsymbol{\theta}_{n-1}) \circ h_n(x; \boldsymbol{\theta}_n).$$

Sejam os vp 's $\theta = (\theta_1, \theta_2, \dots, \theta_{n-1}, \alpha)$ e $\tau = (\theta_1, \theta_2, \dots, \theta_{n-1}, \beta)$ com $\theta, \tau \in \Theta_{\mathcal{F}}$ e $\alpha \neq \beta$ tais que α e β sejam pertencentes a h_n , então $\theta \neq \tau$.

Considerando afirmação do parágrafo anterior segue que

$$\alpha \neq \beta \Rightarrow h_n(x; \alpha) = h_n(x; \beta) \text{ para cada } x \text{ pertencente ao domínio de } h_n \Rightarrow$$

$$h_1(x; \theta_1) \circ h_2(x; \theta_2) \circ \dots \circ h_{n-1}(x; \theta_{n-1}) \circ h_n(x; \alpha) =$$

$$h_1(x; \theta_1) \circ h_2(x; \theta_2) \circ \dots \circ h_{n-1}(x; \theta_{n-1}) \circ h_n(x; \beta) \Rightarrow F_{\mathcal{F}}(x, \theta) = F_{\mathcal{F}}(x, \tau),$$

então $\theta \neq \tau$ implica em $F_{\mathcal{F}}(x, \phi) = F_{\mathcal{F}}(x, \xi)$ para todo x pertencente ao domínio de $F_{\mathcal{F}}$ e a família $\mathcal{F}$ é não identificável.

Portanto, o Teorema 6 está demonstrado. $\square$

Quando aumenta-se o número de parâmetros em uma fd específica, não há garantia de que a fd resultante seja identificável em relação à fd original. A validade dessa afirmação será discutida nos Exemplos 10 e 11.

Exemplo 10. Considere a fd iW com vp $\alpha = (\alpha, \beta)$ e seja a fd GiW com vp $\beta = (\alpha, \beta, \lambda)$. Realizando uma operação matemática na distribuição iW na qual resulte em um acréscimo de um parâmetro (λ) e assim surgindo uma outra distribuição a GiW , segue que houve uma ampliação paramétrica na iW e que $\beta \ni \alpha$. As distribuições iW 's pertencem a uma subfamília da família de distribuições GiW 's e mesmo havendo esta relação entre as duas distribuições a iW é identificável e a distribuição GiW é não identificável.

Segue da argumentação do Exemplo 10 que ampliações paramétricas de famílias identificáveis podem produzir famílias não identificáveis.

A composição de funções é, de fato, um exemplo específico de ampliação paramétrica. Consequentemente, não se pode assegurar que a composição entre duas fd 's identificáveis resulte em uma distribuição identificável. O próximo exemplo discutirá essa situação em detalhes.

Exemplo 11. Considere a classe fd Kumaraswamy- G , que é identificável, e a fd baseline iW , que é identificável. A composição destas duas funções produz uma distribuição não identificável, como já visto no Exemplo 9. Considere a fd baseline GiW , que é não identificável. A composição da fd GiW com a classe de fd Kumaraswamy- G produz a fd Kum- GiW que é não identificável. Será demonstrado que a distribuição Kum- GiW é não identificável nos próximos parágrafos deste exemplo.

Pelo Exemplo 9 e pelo Exemplo 3 a expressão da distribuição Kum- GiW é dada por

$$F_{\mathcal{M}}(x; a, b, c, \beta, \lambda) = 1 - \left\{ 1 - \exp \left[-a\lambda \left(\frac{c}{x} \right)^{\beta} \right] \right\}^b.$$

Considere que a distribuição Kum-GiW seja fruto de uma ampliação paramétrica na fd Kum-iW tal como descrito pelos próximos passos.

Seja fd Kum-iW dada por

$$F_{\mathcal{M}}(x; a, b, c, \beta) = 1 - \left\{ 1 - \exp \left[-a \left(\frac{c}{x} \right)^{\beta} \right] \right\}^b = h_1(x; b) \circ h_2(x; a, c, \beta) \quad (3.30)$$

em que $h_1(x; b) = 1 - \{1 - x\}^b$, $h_2(x; a, c, \beta) = \exp \left[-a \left(\frac{c}{x} \right)^{\beta} \right]$ e a expressão 3.30 pertence a uma família não identificável (Exemplo 9).

Considere uma função real g tal como

$$g(x; a, b, c, \beta, \lambda) = \exp \left[-a\lambda \left(\frac{c}{x} \right)^{\beta} \right] = u(x; \lambda) \circ h_2(x; a, c, \beta) \quad (3.31)$$

em que $u(x; \lambda) = x^{\lambda}$ e $h_2(x; a, c, \beta) = \exp \left[-a \left(\frac{c}{x} \right)^{\beta} \right]$.

Seja uma fd de probabilidade descrita como segue

$$\begin{aligned} F_{\mathcal{M}}(x; a, b, c, \beta, \lambda) &= h_1(x; b) \circ u(x; \lambda) \circ h_2(x; a, c, \beta) = \\ h(x; b) \circ g(x; a, c, \beta, \lambda) &= 1 - \left\{ 1 - \exp \left[-a\lambda \left(\frac{c}{x} \right)^{\beta} \right] \right\}^b \end{aligned} \quad (3.32)$$

a expressão 3.32 é a fd Kum-GiW e pelo Teorema 6 a distribuição Kum-GiW pertence a uma família não identificável.

Portanto, fica provada a não identificabilidade da família Kum-GiW.

De fato, se for realizado uma diferenciação em uma distribuição identificável, isso não perturbará sua identificabilidade. Em outras palavras, se uma distribuição for identificável e aplicarmos diferenciação a ela, o resultado dessa operação será uma distribuição também identificável.

Teorema 7 (Identificabilidade pela diferenciação). *Seja $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{F})$ uma estrutura estatística em que $\mathcal{B}(\mathbb{R})$ é uma σ -álgebra de Borel formada por subconjuntos de $\mathbb{R}$ e seja $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ uma família de fd's absolutamente contínuas sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. A distribuição $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ pode ser escrita como segue:*

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = \int_{-\infty}^x dF_{\mathcal{F}}(y; \boldsymbol{\theta}) \quad \text{para cada } x \in \mathbb{R}.$$

em que $dF_{\mathcal{F}}(y; \boldsymbol{\theta})$ representa uma fdp (fdp).

Se a função $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ é identificável, então $dF_{\mathcal{F}}(x; \boldsymbol{\theta})$ é identificável.

Demonstração. Sejam $\theta, \xi \in \Theta_{\mathcal{F}}$. Por $F_{\mathcal{F}}(x; \cdot)$ ser identificável (hipótese) e pela Definição 19, tem-se que

$$F_{\mathcal{F}}(x; \theta) = F_{\mathcal{F}}(x; \xi), \forall x \in \mathbb{R} \Rightarrow \theta = \xi. \quad (3.33)$$

Então, derivando F em 3.33 tem-se

$$dF_{\mathcal{F}}(x; \theta) = dF_{\mathcal{F}}(x; \xi) \Rightarrow \theta = \xi, \quad (3.34)$$

pela contrapositiva da sentença 3.34, tem-se que

$$\theta \neq \xi \Rightarrow dF_{\mathcal{F}}(x; \theta) \neq dF_{\mathcal{F}}(x; \xi) \quad (3.35)$$

em que x pertence a um conjunto de medidas de Lebesgue não nulas e pela Definição 9 $dF_{\mathcal{F}}$ é identificável.

Portanto, pode-se estudar a identificabilidade pela *fdp* de uma *fd* $F_{\mathcal{F}}(x; \cdot)$ e pode-se afirmar que a identificabilidade é invariante para a diferenciação de *fd*'s absolutamente contínuas. Assim fica provado o Teorema 7. $\square$

Teorema 8 (Identificabilidade pela integração). *Seja $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{F})$ um estrutura estatística em que $\mathcal{B}(\mathbb{R})$ é uma σ -álgebra de Borel formada por subconjuntos de $\mathbb{R}$ e seja $\mathcal{F} = \{F_{\mathcal{F}}(x; \theta) : \theta \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ uma família de *fd*'s absolutamente contínuas sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. A distribuição $F_{\mathcal{F}}(x; \theta)$ pode ser escrita como segue que*

$$F_{\mathcal{F}}(x; \theta) = \int_{-\infty}^x dF_{\mathcal{F}}(y; \theta) \quad \text{para cada } x \in \mathbb{R}$$

em que $dF_{\mathcal{F}}(y; \theta)$ *fdp*.

A função $dF_{\mathcal{F}}(x; \theta)$ é identificável, então $F_{\mathcal{F}}(x; \theta)$ é identificável.

Demonstração. Sejam $\theta, \xi \in \Theta_{\mathcal{F}}$. Se $dF_{\mathcal{F}}(x; \cdot)$ é identificável (hipótese), pelo Teorema 7 e pela Definição 19, tem-se

$$dF_{\mathcal{F}}(x; \theta) = dF_{\mathcal{F}}(x; \xi), \forall x \in \mathbb{R} \Rightarrow \theta = \xi. \quad (3.36)$$

Então, integrando dF em 3.36 no intervalo $(-\infty, x]$, $\forall x \in \mathbb{R}$, tem-se

$$F_{\mathcal{F}}(x; \theta) = F_{\mathcal{F}}(x; \xi), \forall x \in \mathbb{R} \Rightarrow \theta = \xi, \quad (3.37)$$

pela contrapositiva da sentença 3.37, tem-se que

$$\theta \neq \xi \Rightarrow F_{\mathcal{F}}(x; \theta) \neq F_{\mathcal{F}}(x; \xi) \text{ em um conjunto de medidas de Lebesgue não nulas} \quad (3.38)$$

e pela Definição 9 $F_{\mathcal{F}}$ é identificável.

Portanto, se $dF_{\mathcal{F}}(x; \cdot)$ é identificável segue que $F_{\mathcal{F}}(x; \cdot)$ é identificável e pode-se concluir que a identificabilidade é invariante para a integração de funções absolutamente contínuas e assim fica provado o Teorema 8. $\square$

Teorema 9 (Identificabilidade em *fd*'s de probabilidade discretas). *Seja $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{F})$ uma estrutura estatística em que $\mathcal{B}(\mathbb{R})$ é uma σ -álgebra de Borel formada por subconjuntos de $\mathbb{R}$ e $F_{\mathcal{F}}$ é uma *fd* discreta sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Seja $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{F}}, x \in \mathbb{R}\}$ uma família de *fd*'s discretas tal que $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ possa ser escrita como segue:*

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = \sum_{i \in I} p_{\mathcal{F}}(x_i; \boldsymbol{\theta}) \text{ para todo } x_i \in \mathbb{R} \text{ e } i \in I$$

em que $I = \{i : x_i \leq x\}$ e $p_{\mathcal{F}}(x_i; \boldsymbol{\theta})$ representa uma função de probabilidade (*fp*).

Se a distribuição $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ é identificável, então $p_{\mathcal{F}}(x; \boldsymbol{\theta})$ é identificável.

Demonstração. Sejam $\boldsymbol{\theta}$ e $\boldsymbol{\xi}$, dois vetores de parâmetros pertencentes a $\Theta_{\mathcal{F}}$, tais que $\boldsymbol{\theta} \neq \boldsymbol{\xi}$ implique em $F_{\mathcal{F}}(x; \boldsymbol{\theta}) \neq F_{\mathcal{F}}(x; \boldsymbol{\xi})$ para algum $x \in \mathbb{R}$. Então, pela hipótese tem-se que

$$\boldsymbol{\theta} \neq \boldsymbol{\xi} \Rightarrow \sum_{i \in I} p_{\mathcal{F}}(x_i; \boldsymbol{\theta}) \neq \sum_{i \in I} p_{\mathcal{F}}(x_i; \boldsymbol{\xi}) \Rightarrow p_{\mathcal{F}}(x_i; \boldsymbol{\theta}) \neq p_{\mathcal{F}}(x_i; \boldsymbol{\xi})$$

para algum $x_i \in \mathbb{R}$ e $i \in I = \{i : x_i \leq x\}$.

Portanto, $\boldsymbol{\theta} \neq \boldsymbol{\xi}$ implica em $p_{\mathcal{F}}(x_i; \boldsymbol{\theta}) \neq p_{\mathcal{F}}(x_i; \boldsymbol{\xi})$ para algum $x_i \in \mathbb{R}$ e $i \in I$ e $\Theta_{\mathcal{F}}$ é identificável. Assim fica provado o Teorema 9. $\square$

Teorema 10 (Identificabilidade em funções de probabilidade). *Seja $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_{\mathcal{F}})$ uma ep em que $\mathcal{B}(\mathbb{R})$ é uma σ -álgebra de Borel formada por subconjuntos de $\mathbb{R}$ e $F_{\mathcal{F}}$ é uma *fd* discreta sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Seja $\mathcal{F} = \{F_{\mathcal{F}}(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{F}}, x \in \mathcal{B}(\mathbb{R})\}$ uma família de *fd*'s discretas tal que $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ possa ser escrita como segue:*

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = \sum_{i \in I} p_{\mathcal{F}}(x_i; \boldsymbol{\theta}) \text{ para cada } x \in \mathbb{R}.$$

em que $I = \{i : x_i \leq x\}$ e $p_{\mathcal{F}}(x_i; \boldsymbol{\theta})$ representa uma *fp*.

Se a *fp* $p_{\mathcal{F}}(x; \boldsymbol{\theta})$ é identificável, então $F_{\mathcal{F}}(x; \boldsymbol{\theta})$ é identificável.

Demonstração. Sejam $\boldsymbol{\theta}, \boldsymbol{\xi} \in \Theta_{\mathcal{F}}$ tais que $\boldsymbol{\theta} \neq \boldsymbol{\xi}$ implique em $p_{\mathcal{F}}(x_i; \boldsymbol{\theta}) \neq p_{\mathcal{F}}(x_i; \boldsymbol{\xi})$ para algum $x \in \mathbb{R}$, então pela hipótese tem-se que

$$\boldsymbol{\theta} \neq \boldsymbol{\xi} \Rightarrow p_{\mathcal{F}}(x_i; \boldsymbol{\theta}) \neq p_{\mathcal{F}}(x_i; \boldsymbol{\xi}) \Rightarrow$$

$$\sum_{i \in I} p_{\mathcal{F}}(x_i; \boldsymbol{\theta}) \neq \sum_{i \in I} p_{\mathcal{F}}(x_i; \boldsymbol{\xi}) \Rightarrow F_{\mathcal{F}}(x; \boldsymbol{\theta}) \neq F_{\mathcal{F}}(x; \boldsymbol{\xi})$$

para algum $x_i \in \mathbb{R}$ e $i \in I = \{i : x_i \leq x\}$.

Logo, $\boldsymbol{\theta} \neq \boldsymbol{\xi}$ implica em $F_{\mathcal{F}}(x; \boldsymbol{\theta}) \neq F_{\mathcal{F}}(x; \boldsymbol{\xi})$ e $\Theta_{\mathcal{F}}$ é identificável. Assim fica provado o Teorema 10. $\square$

3.3 Estudo da identificabilidade nas distribuições *EGW* e *EW*

Esta seção é fundamentada no trabalho de [Gusmão et al. \(2021\)](#), resultante desta tese de doutorado. Alguns resultados apresentados no artigo foram deliberadamente omitidos neste contexto, mas podem ser consultados na obra mencionada anteriormente.

Nesta seção, examinamos a identificabilidade de duas *fd*'s: a Weibull exponenciada generalizada proposta por [Oguntunde, Odetunmibi e Adejumo \(2015\)](#) e a Weibull exponenciada introduzida por [Mudholkar, Srivastava e Freimer \(1995\)](#), esta última obtida por meio de uma reparametrização da primeira.

Várias famílias de distribuições propostas na literatura representam uma fonte significativa para modelagem de dados de vida útil, visto que, em geral, o suporte da distribuição resultante é condicionado pela baseline G , em que $0 < G(x) < 1$. [Cordeiro, Ortega e Cunha \(2013\)](#) introduziram uma nova classe de distribuições exponenciadas generalizadas (*EG*) com o intuito de oferecer alternativas para modelar diversos tipos de dados. Supondo que uma variável aleatória X siga uma distribuição G , a nova classe de distribuições *EG* é expressa por:

$$F_{\mathcal{F}}(x; a, b) = \{1 - [1 - G(x)]^a\}^b, \quad (3.39)$$

em que $a > 0$ e $b > 0$ são dois parâmetros de forma adicionais. Os autores comentam que a nova classe de distribuições é uma função menos complicada do que a família beta generalizada introduzida por [Eugene, Lee e Famoye \(2002\)](#), pois as simulações são tratadas com facilidade, pois a função quantil (*qf*) tem forma fechada.

Com base nas ideias previamente mencionadas, ao considerarmos a composição entre a classe exponenciada generalizada de distribuições e a distribuição Weibull como candidata a baseline, obtem-se a distribuição Weibull exponenciada generalizada (*EGW*), conforme introduzida por [Oguntunde, Odetunmibi e Adejumo \(2015\)](#).

3.3.1 A distribuição *EGW* e um estudo sobre identificabilidade

A distribuição Weibull tem recebido atenção considerável na literatura estatística. Muitos autores estudaram as formas da densidade e funções taxa de falha para o modelo clássico da distribuição Weibull. Seja X uma variável aleatória com distribuição Weibull biparamétrica, então a *fd* de probabilidade pode ser escrita como

$$G(x; \alpha, \beta) = 1 - \exp \left\{ - [\alpha x]^\beta \right\}$$

em que $\alpha > 0$, $\beta > 0$ e $x > 0$.

Seja $\mathcal{G}$ a família de *fd*'s Weibull generalizada exponenciada (*EGW*) estudada por [Oguntunde, Odetunmbi e Adejumo \(2015\)](#) com $\boldsymbol{\theta} = (a, b, \alpha, \beta)$ sendo o *vp* da *EGW* e $F_{\mathcal{G}}(x; \boldsymbol{\theta})$ expressa por

$$F_{\mathcal{G}}(x; \boldsymbol{\theta}) = \left\{ 1 - \exp \left[-a (\alpha x)^{\beta} \right] \right\}^b \quad (3.40)$$

em que a , b e β são os parâmetros de forma e α é o parâmetro de escala, todos pertencentes ao conjunto $\mathbb{R}_+$ positivos. A *fd* $F_{\mathcal{G}}(x; \boldsymbol{\theta})$ possui suporte em $(0, \infty)$ e a expressão 3.40 é a composição de $G(x; \alpha, \beta)$ com a $F_{\mathcal{F}}(x; \boldsymbol{\theta})$.

A expressão 3.40 pode ser escrita, pela Definição 11, como uma composição de funções reais tal como

$$F_{\mathcal{G}}(x; \boldsymbol{\theta}) = h_1(x; a, b) \circ h_2(x; \alpha, \beta)$$

em que $h_1(x; a, b) = \{1 - [1 - x]^a\}^b$ e $h_2(x; \alpha, \beta) = 1 - \exp \left\{ -[\alpha x]^{\beta} \right\}$.

Pelos argumentos já expostos ao longo desta tese em exemplos, é óbvio a não identificabilidade da expressão 3.40. Para ver uma demonstração formal sobre isto ver [Gusmão et al. \(2021\)](#).

3.3.2 A distribuição *EW* e um estudo sobre identificabilidade

A reparametrização $c = \alpha a^{\frac{1}{\beta}}$ realizada nos parâmetros da distribuição $F_{\mathcal{G}}(x; \boldsymbol{\theta})$ promove identificabilidade, veja o artigo [Gusmão, Ortega e Cordeiro \(2011\)](#). Sem esta reparametrização vários valores de a e α satisfazem a relação $c^{\beta} = \alpha a^{\beta}$ para um valor fixo c . Com a relação a citada reparametrização é possível reescrever a expressão (3.40), obtendo a distribuição *EW* como segue

$$F_{\mathcal{H}}(x; b, c, \beta) = \left\{ 1 - \exp \left[-(cx)^{\beta} \right] \right\}^b,$$

em que $b > 0$ é o parâmetro de forma e $c > 0$ o parâmetro de escala. A expressão $F_{\mathcal{H}}(x; b, c, \beta)$ pertence a uma família identificável de *fd* de probabilidade. A expressão $F_{\mathcal{H}}(x; b, c, \beta)$ foi introduzida por [Mudholkar, Srivastava e Freimer \(1995\)](#).

O *vp* $\boldsymbol{\theta}_i = (b_i, c_i, \beta_i)$ pode diferir do *vp* $\boldsymbol{\theta}_j = (b_j, c_j, \beta_j)$ por sete caminhos, em que $i \neq j$; $i, j \in I$ e I é um certo conjunto de índices. Considere $\boldsymbol{\theta}_i, \boldsymbol{\theta}_j \in \Theta_{\mathcal{H}}$, em que $i \neq j$; $i, j \in I$ e I é um certo conjunto de índices.

De acordo com o parágrafo anterior $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_j$ e considere como o Caso 1 $b_i \neq b_j$, $c_i = c_j = c$ e $\beta_i = \beta_j = \beta$, conseqüentemente pela hipótese tem-se a seguinte cadeia de implicações:

$$b_i \neq b_j \Rightarrow \left\{ 1 - \exp \left[-(cx)^{\beta} \right] \right\}^{b_i} \neq \left\{ 1 - \exp \left[-(cx)^{\beta} \right] \right\}^{b_j} \Rightarrow \\ F_{\mathcal{H}}(x; \boldsymbol{\theta}_i) \neq F_{\mathcal{H}}(x; \boldsymbol{\theta}_j)$$

para algum x pertencente ao domínio de $F_{\mathcal{H}}(x; \cdot)$. Os demais casos, quanto o Caso 1, estão expressos na Tabela 2. A referida tabela resume a prova de identificabilidade para a família EW de distribuições.

Tabela 2 – Prova de que $\Theta_{\mathcal{F}}$ é identificável.

Casos	Hipóteses: $\theta_i \neq \theta_j$	Implicações
1	$b_i \neq b_j, c_i = c_j$ e $\beta_i = \beta_j = \beta$	$F_{\mathcal{H}}(x; \theta_i) \neq F_{\mathcal{H}}(x; \theta_j)$
2	$b_i = b_j = b, c_i = c_j = c$ e $\beta_i \neq \beta_j$	$F_{\mathcal{H}}(x; \theta_i) \neq F_{\mathcal{H}}(x; \theta_j)$
3	$b_i = b_j = b, c_i \neq c_j$ e $\beta_i = \beta_j = \beta$	$F_{\mathcal{H}}(x; \theta_i) \neq F_{\mathcal{H}}(x; \theta_j)$
4	$b_i = b_j = b, c_i \neq c_j$ e $\beta_i \neq \beta_j$	$F_{\mathcal{H}}(x; \theta_i) \neq F_{\mathcal{H}}(x; \theta_j)$
5	$b_i \neq b_j, c_i \neq c_j$ e $\beta_i = \beta_j = \beta$	$F_{\mathcal{H}}(x; \theta_i) \neq F_{\mathcal{H}}(x; \theta_j)$
6	$b_i \neq b_j, c_i = c_j = c$ e $\beta_i \neq \beta_j$	$F_{\mathcal{H}}(x; \theta_i) \neq F_{\mathcal{H}}(x; \theta_j)$
7	$b_i \neq b_j, c_i \neq c_j$ e $\beta_i \neq \beta_j$	$F_{\mathcal{H}}(x; \theta_i) \neq F_{\mathcal{H}}(x; \theta_j)$

Portanto, pela Tabela 2 o *evp* $\Theta_{\mathcal{H}}$ é identificável.

3.3.3 Conclusão do Capítulo 3

Foram estabelecidas definições que formalizam os conceitos discutidos na teoria da identificabilidade paramétrica. A teoria delineada neste capítulo proporciona um arcabouço substancial para a classificação da identificabilidade de distribuições ou de classes de distribuições de probabilidades. Exemplos ilustrativos sobre como realizar tal classificação são encontrados ao longo desta seção da tese.

4 Identificabilidade em Métodos Construtores de fd 's Pertencente a Classes não Finitas

Nesta seção é apresentado um novo método construtor de fd a partir da combinação de fd 's generalizadas. Seja $h_{ijl} : [0, 1] \rightarrow [0, 1]$ uma função real, considere uma família de fd 's não enumeráveis e não aditivamente fechadas $\mathcal{F}_{ijl} = \{F_{\mathcal{F}_{ijl}}(x; \beta_{ijl}) : \beta_{ijl} \in \Theta_{\mathcal{F}_{ijl}}, x \in \mathbb{R}\}$ com o vp β_{ijl} tendo dimensão $n \geq 2$ e seja $\mathcal{G}_{ijl} = \{F_{\mathcal{G}_{ijl}}(x; \alpha_{ijl}) : \alpha_{ijl} \in \Theta_{\mathcal{G}_{ijl}}, x \in \mathbb{R}\}$ uma família de fd 's não enumeráveis e não aditivamente fechadas com o vp α_{ijl} tendo dimensão $n \geq 2$; para cada $i \in I, j \in J$ e $l \in L$. Considere os conjuntos $I, J, L \subset \mathbb{N}$ com cardinalidade finita e os elementos dos mesmos escolhidos de forma arbitrária. Seja a composição entre h_{ijl} e $F_{\mathcal{F}_{ijl}}$ tal como segue

$$F_{\mathcal{G}_{ijl}}(x; \alpha_{ijl}) = h_{ijl}[F_{\mathcal{F}_{ijl}}(x; \beta_{ijl}); \delta_{ijl}]. \quad (4.1)$$

Seja 4.1 uma fd pertencente a família $\mathcal{G}_{ijl}$. A expressão 4.1 é dita uma fd generalizada.

4.1 Método construtor de fd a partir da combinação de fd 's generalizadas

Seja $g : [0, 1]^n \rightarrow [0, 1]$ uma função. O método construtor de distribuição de probabilidade pertencente a uma classe de misturas não finitas a partir da combinação de fd 's generalizadas $F_{\mathcal{G}_{ijl}}$ é apresentado a seguir

$$\begin{aligned} F_{\mathcal{F}}(x; \alpha) &= g \left[F_{\mathcal{G}_{i_0 j_0 l_0}}(x; \alpha_{i_0 j_0 l_0}), \dots, F_{\mathcal{G}_{ijl}}(x; \alpha_{ijl}); \zeta \right] \\ &= \sum_{i \in I} \left\{ a_i \prod_{j \in J} \left[\sum_{l \in L} c_l F_{\mathcal{G}_{ijl}}(x; \alpha_{ijl}) \right]^{\lambda_{ij}} \right\} \end{aligned} \quad (4.2)$$

em que $\alpha = (\alpha_{i_0 j_0 l_0}, \dots, \alpha_{ijl}, \zeta)$, com $i_0 j_0 l_0$ e ijl pertencentes aos conjuntos I, J, L . Considere $0 \leq a_i \leq 1, \sum_{i \in I} a_i = 1, 0 \leq c_l \leq 1, \sum_{l \in L} c_l = 1$ e $\lambda_{ij} \geq 0$ parâmetros para todo $i \in I, j \in J$ e $l \in L$.

Vale ressaltar que a expressão 4.2 generaliza, com alguns pequenos ajustes sobre o suporte dos parâmetros, o “Método gerador de modelos de crescimento e decrescimento via combinações de modelos existentes” de Santos (2019).

No artigo [Teicher \(1961\)](#) é apresentado um teorema que garante identificabilidade de misturas de distribuições uniparamétricas e aditivamente fechadas e em [Teicher \(1963\)](#) a identificabilidade é alcançada para a classe de misturas finitas através de dois teoremas. Neste capítulo são consideradas misturas não finitas de distribuições, o que difere dos estudos realizados por todos os pesquisadores citados na Introdução e na Revisão de literatura desta tese. Nas subseções a seguir é mostrado alguns casos particulares do método construtor [4.2](#) tal como os respectivos estudos sobre identificabilidade para cada submétodo apresentado.

4.1.1 Método construtor de fd 's pela soma de funções de distribuição generalizadas

O método construtor de fd pela soma de fd 's generalizadas é um caso particular do método construtor [4.2](#) e é obtido considerando os itens a seguir:

- $c_l = 0$ para todo $l \in L - \{l_0\}$ tal que $l_0 \in L$ é fixo e $c_{l_0} = 1$;
- $\lambda_{ij} = 0$ para todo $i \in I$ e para todo $j \in J - \{j_0\}$ tal que $j_0 \in J$ é fixo e $\lambda_{ij_0} = 1$ para todo $i \in I$.

Sejam $I = \{1, 2, \dots, n\}$, $J = \{1, 2, \dots, m\}$ e $L = \{1, 2, \dots, k\}$ com $n, m, k \in \mathbb{N}$ e $n, m, k \geq 2$, então levando-se em consideração os itens anteriores o método construtor [4.2](#) pode ser reescrito como segue

$$F_{\mathcal{M}}(x; \boldsymbol{\rho}) = \sum_{i=1}^n a_i F_{\mathcal{G}_i}(x; \boldsymbol{\alpha}_i) \quad (4.3)$$

em que a expressão [4.3](#) é o método construtor de fd pela soma de fd 's generalizadas e $\mathcal{G}_i$ é uma subfamília de $\mathcal{F}$, para todo $i \in I$, com $\boldsymbol{\rho} \subseteq \boldsymbol{\alpha}$ e $\mathcal{M}$ é uma outra família de fd 's distinta da família $\mathcal{F}$.

Teorema 11. *Considere o método construtor [4.1.1](#) tal que $F_{\mathcal{G}_i}(x; \boldsymbol{\alpha}_i) = F_{\mathcal{Z}}(x; \boldsymbol{\alpha}_i)$ para todo $i \in I$ e seja $F_{\mathcal{Z}}$ identificável, então a soma ponderada das fd 's expressa por [\(4.3\)](#) e com as condições ora descritas é não identificável.*

Demonstração. Considere $\boldsymbol{\rho} = (a_1, \boldsymbol{\alpha}_1, a_2, \boldsymbol{\alpha}_2, \dots, a_n, \boldsymbol{\alpha}_n) = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_n)$. Seja $\boldsymbol{\theta} = (\boldsymbol{\psi}_n, \boldsymbol{\psi}_{n-1}, \dots, \boldsymbol{\psi}_1)$ um vp tal que $\boldsymbol{\theta} \in \Theta_{\mathcal{M}}$. Segue que $\boldsymbol{\theta} \neq \boldsymbol{\rho}$, então

$$F_{\mathcal{M}}(x; \boldsymbol{\rho}) = \sum_{i=1}^n F(x; \boldsymbol{\psi}_i) = \sum_{i=1}^n F(x; \boldsymbol{\psi}_{n-i+1}) = F_{\mathcal{M}}(x; \boldsymbol{\theta}).$$

Portanto, o método construtor [4.1.1](#) da forma como descrito no Teorema [11](#) é não identificável. $\square$

Lema 4. *Seja $\mathcal{W} = \{F_{\mathcal{W}}(x; \alpha) : \alpha \in \Theta_{\mathcal{W}}, x \in \mathbb{R}\}$ uma família não identificável de fd's. Considere $\beta > 0$ um parâmetro de potência e seja a fd expressa por*

$$F_{\mathcal{U}}(x; \theta) = [F_{\mathcal{W}}(x; \alpha)]^{\beta}. \quad (4.4)$$

A fd 4.4 é não identificável.

Demonstração. Pelo Teorema 6 fica provado o Lema 4. □

Exemplo 12. *O Capítulo 2 do livro [Everitt e Hand \(1981\)](#) é dedicado inteiramente ao estudo da mistura de distribuições normais, a mistura citada é apresentada como segue*

$$F_{\mathcal{N}}(x; \theta) = \sum_{i=1}^n p_i F_{\mathcal{F}}(x; \sigma_i, \eta_i) \quad (4.5)$$

em que $\mathcal{F}$ é a família de fd's de probabilidades normais, $\theta = (p_1, \sigma_1, \eta_1, \dots, p_n, \sigma_n, \eta_n)$ é o vp da fd $F_{\mathcal{N}}(x; \theta)$, $0 < p_i < 1$ e $\sum_{i=1}^n p_i = 1$ com $1 \leq i \leq n$ e $i \in \mathbb{N}$. Pelo Teorema 11 a fd de misturas de normais ponderadas 4.5 é não identificável.

No Capítulo 3 do livro [Everitt e Hand \(1981\)](#) é apresentado o estudo da mistura de distribuições exponenciais, a referida mistura é expressa como segue

$$F_{\mathcal{E}}(x; \varphi) = \sum_{i=1}^n p_i F_{\mathcal{G}}(x; \lambda_i) \quad (4.6)$$

em que $\mathcal{G}$ é a família de fds exponenciais, $\varphi = (p_1, \lambda_1, \dots, p_n, \lambda_n)$ é o vp da fd $F_{\mathcal{E}}(x; \varphi)$, $0 < p_i < 1$ e $\sum_{i=1}^n p_i = 1$ com $1 \leq i \leq n$ e $i \in \mathbb{N}$. A fd de misturas de exponenciais ponderadas 4.6 é identificável pelo Teorema 1.

Por último, no Capítulo 4 do livro [Everitt e Hand \(1981\)](#) é trabalhado o caso de misturas com modelos discretos e o mesmo capítulo apresenta uma expressão de misturas de fp's binomiais tal como segue

$$F_{\mathcal{B}}(x; \nu) = \sum_{i=1}^n p_i F_{\mathcal{H}}(x; \lambda_i) \quad (4.7)$$

em que $\mathcal{H}$ é a família de fd de probabilidades binomiais, $\nu = (p_1, \rho_1, \dots, p_n, \rho_n)$ é o vp da fd $F_{\mathcal{B}}(x; \nu)$, $0 < p_i < 1$ e $\sum_{i=1}^n p_i = 1$ com $1 \leq i \leq n$ e $i \in \mathbb{N}$. A fd de misturas de binomiais ponderadas 4.7 é identificável pelo Teorema 1.

Pelo Exemplo 12 tem-se que devido a não preocupação com a identificabilidade o trabalho de [Everitt e Hand \(1981\)](#) apresenta no capítulo 2 misturas de distribuições de probabilidade não identificáveis.

4.1.2 Método construtor de fd 's pelo produto de potências de fd 's generalizadas

Para se encontrar o método construtor de fd pelo produto de potências de fd 's de probabilidade generalizadas, que é um caso particular do método construtor 4.2, considere os itens a seguir:

- $c_l = 0$ para todo $l \in L - \{l_0\}$ tal que $l_0 \in L$ é fixo e $c_{l_0} = 1$;
- $a_i = 0$ para todo $i \in I - \{i_0\}$ tal que $i_0 \in I$ é fixo e $a_{i_0} = a$ com $0 \leq a \leq 1$.

Sejam $I = \{1, 2, \dots, n\}$, $J = \{1, 2, \dots, m\}$ e $L = \{1, 2, \dots, k\}$ com $n, m, k \in \mathbb{N}$ e $n, m, k \geq 2$, então levando-se em consideração os itens anteriores o método construtor 4.2 pode ser escrito como segue

$$F_{\mathcal{M}}(x; \boldsymbol{\rho}) = a \prod_{j=1}^m [F_{\mathcal{G}_j}(x; \boldsymbol{\alpha}_j)]^{\lambda_j} \quad (4.8)$$

em que a expressão 4.8 é o método construtor de fd pelo produto de potências de fd 's de probabilidade generalizadas.

Teorema 12. *Considere o método construtor 4.1.2 tal que $F_{\mathcal{G}_j}(x; \boldsymbol{\alpha}_j) = F_{\mathcal{Z}}(x; \boldsymbol{\alpha}_j)$ para todo $j \in J$ e seja $F_{\mathcal{Z}}$ identificável, então o produto das fd 's generalizadas expresso por (4.8) com as considerações aqui descritas é não identificável.*

Demonstração. Para provar que o método construtor 4.1.2 com as condições impostas pelo Teorema 12 é não identificável será seguida uma ideia semelhante a prova do teorema anterior, considere

$$F_{\mathcal{M}}(x; \boldsymbol{\rho}) = a \prod_{j=1}^m [F(x; \boldsymbol{\alpha}_j)]^{\lambda_j}$$

em que $\boldsymbol{\rho} = (a, \boldsymbol{\alpha}_1, \lambda_1, \boldsymbol{\alpha}_2, \lambda_2, \dots, \boldsymbol{\alpha}_m, \lambda_m)$.

Seja $\boldsymbol{\theta} = (a, \boldsymbol{\alpha}_m, \lambda_m, \boldsymbol{\alpha}_{m-1}, \lambda_{m-1}, \dots, \boldsymbol{\alpha}_1, \lambda_1) \in \Theta_{\mathcal{G}}$, note que $\boldsymbol{\theta} \neq \boldsymbol{\rho}$, então

$$F_{\mathcal{M}}(x; \boldsymbol{\rho}) = a \prod_{j=1}^m [F(x; \boldsymbol{\alpha}_j)]^{\lambda_j} = a \prod_{j=1}^m [F(x; \boldsymbol{\psi}_{m-j+1})]^{\lambda_{m-j+1}} = F_{\mathcal{M}}(x; \boldsymbol{\theta}).$$

Logo, o método construtor 4.1.2 tal como descrito no Teorema 12 é não identificável. $\square$

Corolário 7. *Considerando o método construtor 4.1.2 com as condições descritas no Teorema 12 e com $\lambda_j = 1$ para todo $j \in J$ e $a = 1$, consequentemente tem-se a expressão*

$$F_{\mathcal{N}}(x; \boldsymbol{\varphi}) = \prod_{j=1}^m F_{\mathcal{Z}}(x; \boldsymbol{\alpha}_j) \quad (4.9)$$

em que $\mathcal{N}$ é subfamília de $\mathcal{M}$ com $\varphi \in \rho$. A expressão 4.9 é não identificável pelo Teorema 12.

Portanto, pelo Corolário 7 o produto de fd 's de mesma família resulta em uma fd não identificável.

4.1.3 Método construtor de fd 's pela soma ponderada de produtórios de fd 's generalizadas

O método construtor de fd 's pela soma ponderada de produtórios de fd 's generalizadas é encontrado a partir do método construtor 4.2 depois de tomado alguns passos indicados pelos itens seguintes:

- $c_l = 0$ para todo $l \in L - \{l_0\}$ tal que $l_0 \in L$ é fixo e $c_{l_0} = 1$;
- $\lambda_{ij} = 1$ para todo $i \in I$ e para todo $j \in J$.

Sejam $I = \{1, 2, \dots, n\}$, $J = \{1, 2, \dots, m\}$ e $L = \{1, 2, \dots, k\}$ com $n, m, k \in \mathbb{N}$ e $n, m, k \geq 2$ e $\mathcal{G}$ sendo subfamília da família $\mathcal{F}$ e também com $\rho \in \alpha$, então agora se pode escrever o método construtor 4.2 como segue

$$F_{\mathcal{G}}(x; \rho) = \sum_{i=1}^n \left[a_i \prod_{j=1}^m F_{\mathcal{G}_{ij}}(x; \alpha_{ij}) \right] \quad (4.10)$$

em que $\rho = (a_1, \alpha_{11}, \dots, \alpha_{1m}, a_2, \alpha_{21}, \dots, \alpha_{2m}, \dots, a_n, \alpha_{n1}, \dots, \alpha_{nm})$.

Teorema 13. *Sejam $F_{\mathcal{G}_{ij}}(x; \alpha_{ij}) = F_{\mathcal{Z}}(x; \alpha_{ij})$ e $F_{\mathcal{Z}}$ uma fd identificável para todo $i \in I$ e $j \in J$ com $I = \{1, 2, \dots, n\}$ e $J = \{1, 2, \dots, m\}$, então fd expressa por (4.10) é não identificável.*

Demonstração. Para provar que o método construtor 4.1.3 com as condições relatadas no Teorema 13 é não identificável, considere o vp $\theta \in \Theta_{\mathcal{G}}$ tal que θ possua ordem inversa na configuração dos parâmetros com relação ao ρ .

Seja $\theta = (a_n, \alpha_{nm}, \dots, \alpha_{n1}, a_{n-1}, \alpha_{n-1m}, \dots, \alpha_{n-11}, \dots, \alpha_{11})$, considere que $a_i = a_u$ e $\alpha_{ij} \neq \alpha_{uv}$ sempre que $ij \neq uv$, exceto quando $i + u = n + 1$ em que $\alpha_{ij} = \alpha_{uj}$ para todo $i \neq u$ e para todo $j \neq v$ com $i, u \in I$ e com $j, v \in J$. Consequentemente $\theta \neq \rho$ e isto implica que

$$F_{\mathcal{G}}(x; \rho) = \sum_{i=1}^n \left[a_i \prod_{j=1}^m F(x; \alpha_j) \right] = \sum_{i=1}^n \left[a_i \prod_{j=1}^m F(x; \alpha_{m-j+1}) \right] = F_{\mathcal{G}}(x; \theta).$$

Logo, o método construtor 4.1.3 com as condições do Teorema 13 é não identificável. $\square$

4.1.4 Método construtor de fd 's pelo produto de potências de somatórios de fd 's generalizadas ponderadas

Seja $F_{\mathcal{G}}(x; \boldsymbol{\rho})$ uma fd fruto do método construtor 4.2, com $\mathcal{G}$ sendo subfamília da família $\mathcal{F}$ e também com $\boldsymbol{\rho} \in \boldsymbol{\alpha}$, que é obtida tomando as particularidades descritas no item a seguir:

- $a_i = 0$ para todo $i \in I - \{i_0\}$ tal que $i_0 \in I$ é fixo e $a_{i_0} = 1$.

Sejam $I = \{1, 2, \dots, n\}$, $J = \{1, 2, \dots, m\}$ e $L = \{1, 2, \dots, k\}$ com $n, m, k \in \mathbb{N}$ e $n, m, k \geq 2$. Então pode-se expressar o método construtor 4.2 a partir da condição expressa no item acima por

$$F_{\mathcal{G}}(x; \boldsymbol{\rho}) = \prod_{j=1}^m \left[\sum_{l=1}^k c_l F_{\mathcal{G}_{jl}}(x; \boldsymbol{\alpha}_{jl}) \right]^{\lambda_j}. \quad (4.11)$$

Teorema 14. *Sejam $F_{\mathcal{G}_{jl}}(x; \boldsymbol{\alpha}_{jl}) = F_{\mathcal{Z}}(x; \boldsymbol{\alpha}_{jl})$ e $F_{\mathcal{Z}}$ uma fd identificável, para todo $j \in J$ e $l \in L$ com $J = \{1, 2, \dots, m\}$ e $L = \{1, 2, \dots, k\}$. Consequentemente o método construtor 4.1.4 com as considerações explicitadas neste teorema é não identificável.*

Demonstração. Demonstrar-se-á aqui que o método construtor 4.1.4 é não identificável. A expressão seguinte

$$\sum_{l=1}^k c_l F(x; \boldsymbol{\alpha}_{jl})$$

é não identificável pelo teorema 11 e pelo Lema 4 a expressão

$$F_{\mathcal{G}}(x; \boldsymbol{\rho}) = \left[\sum_{l=1}^k c_l F(x; \boldsymbol{\alpha}_{jl}) \right]^{\lambda_j}$$

é não identificável. Logo, pelo Teorema 12 a expressão

$$F_{\mathcal{G}}(x; \boldsymbol{\rho}) = \prod_{j=1}^m \left[\sum_{l=1}^k c_l F(x; \boldsymbol{\alpha}_{jl}) \right]^{\lambda_j}$$

é não identificável.

Portanto, o método construtor 4.1.4 descrito como no Teorema 14 é não identificável. $\square$

4.1.5 Método construtor de fd 's pelo somatório ponderado de potências de somatórios ponderados de fd 's generalizadas

Seja a fd $F_G(x; \boldsymbol{\rho})$ resultado do método construtor 4.2, quando aplicado as condições descritas no item a seguir:

- $\lambda_{ij} = 0$ para todo $j \in J - \{j_0\}$ tal que $j_0 \in J$ é fixo e $\lambda_{ij_0} = \lambda_i$.

Sejam $I = \{1, 2, \dots, n\}$, $J = \{1, 2, \dots, m\}$ e $L = \{1, 2, \dots, k\}$ com $n, m, k \in \mathbb{N}$ e $n, m, k \geq 2$, então com as considerações dos itens acima pode-se escrever o método construtor 4.2 tal como segue

$$F_G(x; \boldsymbol{\rho}) = \sum_{i=1}^n a_i \left[\sum_{l=1}^k c_l F_{G_{il}}(x; \boldsymbol{\alpha}_{il}) \right]^{\lambda_i}. \quad (4.12)$$

A expressão 4.12 é denominada de método construtor de fd 's pelo somatório ponderado de potências de somatórios ponderados de fd 's generalizadas.

Teorema 15. *Sejam $F_{G_{il}}(x; \boldsymbol{\alpha}_{il}) = F_Z(x; \boldsymbol{\alpha}_{il})$ e F_Z uma fd identificável para todo $i \in I$ e $l \in L$ com $I = \{1, 2, \dots, n\}$ e $L = \{1, 2, \dots, k\}$. A fd expressa em (4.12) com os termos mencionados neste teorema é não identificável.*

Demonstração. Provar-se-á que o método construtor 4.1.5 é não identificável. Seja a expressão seguinte

$$\sum_{l=1}^k c_l F(x; \boldsymbol{\alpha}_{jl})$$

é não identificável pelo teorema 11, pelo Lema 4 a expressão

$$F_G(x; \boldsymbol{\rho}) = \left[\sum_{l=1}^k c_l F(x; \boldsymbol{\alpha}_{jl}) \right]^{\lambda_j}$$

é não identificável e pelo Teorema 11 a expressão

$$F_G(x; \boldsymbol{\rho}) = \sum_{j=1}^m \left[\sum_{l=1}^k c_l F(x; \boldsymbol{\alpha}_{jl}) \right]^{\lambda_j}$$

é não identificável.

Portanto, o método construtor 4.1.5 acordando com as condições do Teorema 15 é não identificável. $\square$

4.1.6 Método construtor de fd 's pelo somatório ponderado de potências de fd 's generalizadas somado ao produtório de potências de fd 's generalizadas

Considere a fd $F_{\mathcal{G}}(x; \boldsymbol{\rho})$, sendo $\mathcal{G}$ subfamília da família $\mathcal{F}$ e também com $\boldsymbol{\rho} \in \boldsymbol{\alpha}$, considere o método construtor 4.2 com as aplicações descritas nos itens abaixo:

- $c_l = 0$ para todo $l \in L - \{l_0\}$ tal que $l_0 \in L$ é fixo e $c_{l_0} = 1$;
- $\lambda_{ij} = 0$ para todo $i \in I - \{i_0, n\}$ e $j \in J$ tal que $i_0 \in I$ é fixo e $\lambda_{i_0j} = 1$ para todo $j \in J$.

Sejam $I = \{1, 2, \dots, n\}$, $J = \{1, 2, \dots, m\}$ e $L = \{1, 2, \dots, k\}$ com $n, m, k \in \mathbb{N}$ e $n, m, k \geq 2$, então com as considerações dos itens acima anteriores se pode escrever

$$F_{\mathcal{G}}(x; \boldsymbol{\rho}) = \sum_{i=1}^{n-1} a_i \left[F_{\mathcal{G}_{ij}}(x; \boldsymbol{\alpha}_{ij}) \right]^{\lambda_{ij}} + a_n \prod_{j=1}^m \left[F_{\mathcal{G}_{nj}}(x; \boldsymbol{\alpha}_{nj}) \right]^{\lambda_{nj}}. \quad (4.13)$$

A expressão (4.2) é o Método construtor de fd 's pelo somatório ponderado de potências de fd 's generalizadas somado ao produtório de potências de fd 's generalizadas.

Teorema 16. *Sejam $F_{\mathcal{G}_{nj}}(x; \boldsymbol{\alpha}_{nj}) = F_{\mathcal{Z}}(x; \boldsymbol{\alpha}_{nj})$ e $F_{\mathcal{Z}}$ uma fd identificável para todo $i \in I$ e $j \in J$ com $I = \{1, 2, \dots, n\}$ e $J = \{1, 2, \dots, m\}$. Então, o método 4.1.6 considerando o descrito neste teorema é não identificável.*

Demonstração. Pelo Lema 4 e pelos Teoremas 12 e 13, tem-se que 4.13 obedecendo as restrições explicitadas no Teorema 16 é não identificável.

Portanto, o Teorema 16 está provado. $\square$

4.1.7 Método construtor de fd 's por sucessivas composições de funções f 's da fd generalizada

Seja a fd $F_{\mathcal{G}}(x; \boldsymbol{\rho})$ como resultado das sucessivas composições de funções f 's da fd generalizada tal como expresso a seguir:

$$F_{\mathcal{N}}(x; \boldsymbol{\varphi}) = f_1 \circ f_2 \circ f_3 \circ \dots \circ f_n[F_{\mathcal{G}}(x; \boldsymbol{\alpha})] = f_1 \{f_2 \{f_3 \{\dots \{f_n[F_{\mathcal{G}}(x; \boldsymbol{\alpha})]\} \dots \}\}\} \quad (4.14)$$

em que $f_i : [0, 1] \rightarrow [0, 1]$ é uma função paramétrica ou não para todo $1 \leq i \leq n$ com $n \in \mathbb{N}$.

A composição $F_{\mathcal{N}}(x; \boldsymbol{\varphi})$ é identificável ou não a depender da regra de correspondência da referida composição, este tema já foi discutido no Capítulo 3 desta tese. A

construção de métodos construtores de fd 's sem o emprego do crivo da identificabilidade pode ocasionar problemas na identificabilidade da medida de probabilidade encontrada e conduzir a sérios problemas de estimabilidade.

4.1.8 Conclusão do Capítulo 4

Nos artigos de [Teicher \(1960\)](#), [Teicher \(1961\)](#), [Teicher \(1963\)](#), que discutem misturas de fd 's, é de suma importância considerar restrições para garantir a identificabilidade. Essas restrições são minuciosamente detalhadas na subseção [2.3](#). Com base no conteúdo apresentado neste capítulo, para se obter misturas de fd 's não uniparamétricas, é prudente empregar distribuições pertencentes a famílias distintas.

5 Um Estudo sobre Identificabilidade da Família T-**G**-X de Distribuições

5.1 Um estudo da identificabilidade em alguns subcasos da família T-**G**-X de distribuições

Nas subseções subsequentes, serão apresentados e analisados, em termos de identificabilidade paramétrica, alguns subcasos da família T-**G**-X de distribuições de probabilidade, conforme descrita na revisão de literatura desta tese.

5.1.1 A classe de distribuições de probabilidade $F_{\mathcal{H}}-(G_1, G_2)$ e um estudo de identificabilidade

Considere o método T-**G**-X com $\mathcal{H} = \{F_{\mathcal{H}}(x) : x \in \mathbb{R}\}$ sendo uma família de *fd*'s não-paramétricas. Sejam $g[F_{\mathcal{G}}(x; \beta)]$ e $h[F_{\mathcal{M}}(x; \alpha)]$ funções tais que $g, h : \mathbb{R} \rightarrow \overline{\mathbb{R}}$. Ainda, g e h são expressas por

$$g[F_{\mathcal{G}}(x; \beta)] = \frac{F_{\mathcal{G}}(x; \beta)}{1 - F_{\mathcal{G}}(x; \beta)},$$

e

$$h[F_{\mathcal{M}}(x; \alpha)] = \log[1 - F_{\mathcal{M}}(x; \alpha)].$$

A partir do método T-**G**-X pode-se ser escrita a *fd*

$$F_{\mathcal{F}}(x; \theta) = \int_{h[F_{\mathcal{M}}(x; \alpha)]}^{g[F_{\mathcal{G}}(x; \beta)]} dF_{\mathcal{H}}(x) = F_{\mathcal{H}}\{g[F_{\mathcal{G}}(x; \beta)]\} - F_{\mathcal{H}}\{h[F_{\mathcal{M}}(x; \alpha)]\}. \quad (5.1)$$

Teorema 17 (A classe de distribuições de probabilidade $F_{\mathcal{H}}-(G_1, G_2)$ é identificável). *Se $F_{\mathcal{G}}(x; \cdot)$ e $F_{\mathcal{M}}(x; \cdot)$ pertencem a famílias identificáveis, então a classe de distribuições de probabilidade $F_{\mathcal{H}}-(G_1, G_2)$ é identificável.*

Demonstração. Sejam $\theta, \delta \in \Theta_{\mathcal{F}}$ tais que $\theta = (\alpha, \beta)$ e $\delta = (\psi, \xi)$. Por hipótese as *fd*'s $F_{\mathcal{G}}(x; \cdot)$ e $F_{\mathcal{M}}(x; \cdot)$ são identificáveis, então pela Definição 19 tem-se

$$F_{\mathcal{G}}(x; \beta) = F_{\mathcal{G}}(x; \xi), \quad \forall x \in \mathbb{R} \Rightarrow \beta = \xi \quad \text{e} \quad F_{\mathcal{M}}(x; \alpha) = F_{\mathcal{M}}(x; \psi), \quad \forall x \in \mathbb{R} \Rightarrow \alpha = \psi,$$

então

$$F_{\mathcal{H}}\{g[F_{\mathcal{G}}(x; \beta)]\} - F_{\mathcal{H}}\{g[F_{\mathcal{G}}(x; \xi)]\} = 0, \quad \forall x \in \mathbb{R} \Rightarrow \beta = \xi \quad (5.2)$$

e

$$F_{\mathcal{H}} \{h[F_{\mathcal{M}}(x; \alpha)]\} - F_{\mathcal{H}} \{h[F_{\mathcal{M}}(x; \psi)]\} = 0, \forall x \in \mathbb{R} \Rightarrow \alpha = \psi. \quad (5.3)$$

A partir das sentenças 5.2 e 5.3 pode-se escrever que

$$\begin{aligned} & F_{\mathcal{H}} \{g[F_{\mathcal{G}}(x; \beta)]\} - F_{\mathcal{H}} \{h[F_{\mathcal{M}}(x; \alpha)]\} = \\ & F_{\mathcal{H}} \{g[F_{\mathcal{G}}(x; \xi)]\} - F_{\mathcal{H}} \{h[F_{\mathcal{M}}(x; \psi)]\}, \forall x \in \mathbb{R} \Rightarrow \beta = \xi \quad \text{e} \quad \alpha = \psi. \end{aligned} \quad (5.4)$$

Note que a sentença 5.4 pode ser reescrita como

$$\int_{h[F_{\mathcal{M}}(x; \alpha)]}^{g[F_{\mathcal{G}}(x; \beta)]} dF_{\mathcal{H}}(x) = \int_{h[F_{\mathcal{M}}(x; \psi)]}^{g[F_{\mathcal{G}}(x; \xi)]} dF_{\mathcal{H}}(x) \Rightarrow \beta = \xi \quad \text{e} \quad \alpha = \psi.$$

Pela contrapositiva de 5.4 implica em

$$\theta \neq \delta \Rightarrow \int_{h[F_{\mathcal{M}}(x; \alpha)]}^{g[F_{\mathcal{G}}(x; \beta)]} dF_{\mathcal{H}}(x) \neq \int_{h[F_{\mathcal{M}}(x; \psi)]}^{g[F_{\mathcal{G}}(x; \xi)]} dF_{\mathcal{H}}(x).$$

Portanto, pela Definição 9 pode-se concluir que a *fd* $F_{\mathcal{F}}(x; \cdot)$ é identificável. $\square$

Exemplo 13 (A classe de distribuições de probabilidade Normal- (G_1, G_2) e um estudo de identificabilidade). *Considere a classe de distribuições de probabilidade $F_{\mathcal{H}}(G_1, G_2)$ com $\mathcal{H} = \{F_{\mathcal{H}}(x) : x \in \mathbb{R}\}$ sendo uma família de *fd's* não-paramétricas da distribuição normal padrão $(\Phi(x))$.*

A classe de distribuições de probabilidade Normal- (G_1, G_2) pode ser escrita como

$$F_{\mathcal{F}}(x; \theta) = \int_{h[F_{\mathcal{M}}(x; \alpha)]}^{g[F_{\mathcal{G}}(x; \beta)]} d\Phi(x) = \Phi \{g[F_{\mathcal{G}}(x; \beta)]\} - \Phi \{h[F_{\mathcal{M}}(x; \alpha)]\} \quad (5.5)$$

A expressão 5.5 é conhecida por “Normal- (G_1, G_2) Class of Probability Distributions” e foi apresentada por [Silveira et al. \(2023\)](#).

Proposição 5 (A classe de distribuições de probabilidade Normal- (G_1, G_2) é identificável). *Se $F_{\mathcal{G}}(x; \cdot)$ e $F_{\mathcal{M}}(x; \cdot)$ pertencem a famílias identificáveis, então a classe de distribuições de probabilidade Normal- (G_1, G_2) é identificável.*

A Proposição 5 foi apresentada por [Silveira et al. \(2023\)](#); contudo, a prova fornecida no mencionado artigo carece de correção para ser considerada válida. No entanto, com base no Teorema 17, a validade da Proposição 5 é estabelecida.

5.1.2 A classe de distribuições de probabilidade $F_{\mathcal{H}}\text{-G}$ e um estudo de identificabilidade

Considere o método T- $\mathbf{G}$ -X com $\mathcal{H} = \{F_{\mathcal{H}}(x) : x \in \mathbb{R}\}$ sendo uma família de fd 's de probabilidade não-paramétrica. Sejam $g[F_{\mathcal{G}}(x; \beta)]$ uma função tal que $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ e $h : \mathbb{R} \rightarrow -\infty$ uma função constante, ainda g é expressa por

$$g[F_{\mathcal{G}}(x; \beta)] = \frac{2F_{\mathcal{G}}(x; \beta) - 1}{F_{\mathcal{G}}(x; \beta) [1 - F_{\mathcal{G}}(x; \beta)]},$$

agora o método gerador T- $\mathbf{G}$ -X pode ser escrito como

$$F_{\mathcal{F}}(x; \beta) = \int_{-\infty}^{g[F_{\mathcal{G}}(x; \beta)]} dF_{\mathcal{H}}(x) = F_{\mathcal{H}}[g(F_{\mathcal{G}}(x; \beta))]. \quad (5.6)$$

A expressão 5.6 será denominada por Classe de distribuições de probabilidade $F_{\mathcal{H}}\text{-G}$.

Teorema 18 (A classe de distribuições de probabilidade $F_{\mathcal{H}}\text{-G}$ é identificável). *Se $F_{\mathcal{G}}(x; \cdot)$ pertence a uma família identificável, então a classe de distribuições de probabilidade $F_{\mathcal{H}}\text{-G}$ é identificável.*

Demonstração. Sejam $\alpha, \beta \in \Theta_{\mathcal{G}}$ e por hipótese $F_{\mathcal{G}}$ é identificável, então se

$$F_{\mathcal{G}}(x; \alpha) = F_{\mathcal{G}}(x; \beta), \forall x \in \mathbb{R} \Rightarrow \alpha = \beta.$$

Por g ser injetiva e pela sentença anterior, tem-se que

$$g[F_{\mathcal{G}}(x; \alpha)] = g[F_{\mathcal{G}}(x; \beta)], \forall x \in \mathbb{R} \Rightarrow \alpha = \beta.$$

Por $F_{\mathcal{H}}$ ser identificável e pela sentença anterior, tem-se que

$$F_{\mathcal{H}}\{g[F_{\mathcal{G}}(x; \alpha)]\} = F_{\mathcal{H}}\{g[F_{\mathcal{G}}(x; \beta)]\}, \forall x \in \mathbb{R} \Rightarrow \alpha = \beta.$$

Pela contrapositiva da sentença anterior, tem-se que

$$\alpha \neq \beta \Rightarrow F_{\mathcal{H}}\{g[F_{\mathcal{G}}(x; \alpha)]\} \neq F_{\mathcal{H}}\{g[F_{\mathcal{G}}(x; \beta)]\},$$

para um conjunto de medida de Lebesgue não nula.

Portanto, a classe 5.6 é identificável. □

Exemplo 14 (A classe de distribuições de probabilidade Normal-G e um estudo de identificabilidade). *Considere a classe de distribuições de probabilidade $F_{\mathcal{H}}\text{-G}$ com $\mathcal{H} =$*

$\{F_{\mathcal{H}}(x) : x \in \mathbb{R}\}$ sendo uma família de fd's de probabilidade não-paramétrica da normal padrão $(\Phi(x))$. A classe de distribuições de probabilidade $F_{\mathcal{H}}-G$ pode ser escrita como

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = \int_{-\infty}^{g[F_{\mathcal{G}}(x; \boldsymbol{\beta})]} d\Phi(x) = \Phi[g(F_{\mathcal{G}}(x; \boldsymbol{\beta}))]. \quad (5.7)$$

A expressão 5.7 é conhecida por “Normal-G Class of Probability Distributions” e foi apresentada por [Silveira et al. \(2019\)](#).

Proposição 6 (A classe de distribuições de probabilidade Normal-G é identificável). *Se $F_{\mathcal{G}}(x; \cdot)$ pertence a uma família identificável, então a classe de distribuições de probabilidade Normal-G é identificável.*

A Proposição 6 foi apresentada em [Silveira et al. \(2019\)](#) e pode ser tratada como um corolário do Teorema 18.

5.1.3 A classe $F_{\mathcal{H}}(x)$ -Tangente-G de distribuições de probabilidade e um estudo da identificabilidade

Seja o método T-**G**-X com $\mathcal{H} = \{F_{\mathcal{H}}(x) : x \in \mathbb{R}\}$ sendo uma família de fd's de probabilidade não-paramétrica. Sejam $g[F_{\mathcal{G}}(x; \boldsymbol{\beta})]$ uma função tal que $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ e $h : \mathbb{R} \rightarrow -\infty$ uma função constante, ainda g é expressa por

$$g[F_{\mathcal{G}}(x; \boldsymbol{\beta})] = \tan[\pi(F_{\mathcal{G}}(x; \boldsymbol{\beta}) - 1/2)]$$

agora o método gerador T-**G**-X pode ser expresso por

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = \int_{-\infty}^{g[F_{\mathcal{G}}(x; \boldsymbol{\beta})]} dF_{\mathcal{H}}(x) = F_{\mathcal{H}}[g(F_{\mathcal{G}}(x; \boldsymbol{\beta}))]. \quad (5.8)$$

A expressão 5.8 pertence a classe de distribuições $F_{\mathcal{H}}$ -tangente-G, que um subcaso da T-**G**-X ([BRITO et al., 2019](#)).

Teorema 19 (A classe de distribuições de probabilidade $F_{\mathcal{H}}$ -Tangente-G é identificável). *Se $F_{\mathcal{G}}(x; \cdot)$ pertence a uma família identificável, então a classe de distribuições de probabilidade $F_{\mathcal{H}}$ -Tangente-G é identificável.*

Demonstração. Sejam $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \Theta_{\mathcal{G}}$ e como $F_{\mathcal{H}}[g(F_{\mathcal{G}}(x; \cdot))]$ é injetiva, pode-se inferir de 5.8 que

$$\tan[\pi(F_{\mathcal{G}}(x; \boldsymbol{\alpha}) - 1/2)] = \tan[\pi(F_{\mathcal{G}}(x; \boldsymbol{\beta}) - 1/2)] \Rightarrow F_{\mathcal{G}}(x; \boldsymbol{\alpha}) = F_{\mathcal{G}}(x; \boldsymbol{\beta}) \Rightarrow \boldsymbol{\alpha} = \boldsymbol{\beta},$$

pela Definição 19.

Portanto, se $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$ implica em $F_{\mathcal{G}}(x; \boldsymbol{\alpha}) \neq F_{\mathcal{G}}(x; \boldsymbol{\beta})$ pode-se concluir que 5.8 é identificável. $\square$

Exemplo 15 (A classe Normal-Tangente-G de distribuições de probabilidade e um estudo da identificabilidade). *Considere a classe de distribuições de probabilidade $F_{\mathcal{H}}$ -Tangente-G com $\mathcal{H} = \{F_{\mathcal{H}}(x) : x \in \mathbb{R}\}$ sendo uma família de fd's de probabilidade não-paramétrica da normal padrão ($\Phi(x)$). O método 19, com $F_{\mathcal{H}}$ sendo a distribuição normal padrão, pode ser expresso por*

$$F_{\mathcal{F}}(x; \boldsymbol{\theta}) = \int_{-\infty}^{g[F_{\mathcal{G}}(x; \boldsymbol{\beta})]} d\Phi(x) = \Phi[g(F_{\mathcal{G}}(x; \boldsymbol{\beta}))]. \quad (5.9)$$

A expressão 5.9 é conhecida por “The normal-tangent-G class of probabilistic distributions” e foi introduzida por [Silveira et al. \(2020\)](#).

Proposição 7 (A classe de distribuições de probabilidade Normal-Tangente-G é identificável). *Se $F_{\mathcal{G}}(x; \cdot)$ pertence a uma família identificável, então a classe de distribuições de probabilidade Normal-Tangente-G é identificável.*

A Proposição 7 foi apresentada em [Silveira et al. \(2020\)](#) e também precisa do Teorema 19 para ser validada.

5.1.4 Estudo de identificabilidade em casos particulares da Família T-X de distribuições

Considere o método gerador distribuições e classes de distribuições de probabilidade no caso univariado T- $\mathcal{G}$ -X. Considere $h(x) = a$ uma função constante e não-paramétrica com $a \in \mathbb{R}$ e seja g uma função diferenciável com domínio em uma única $\mathcal{G}_{i_0}$ em que $i_0 \in [1, m]$ com i_0 é fixo. Se $F_{\mathcal{H}}(x; \boldsymbol{\rho})$ for contínua e diferenciável, então a família T-X de distribuições pode ser escrita como:

$$F_{\mathcal{T}-\mathcal{X}}(x; \boldsymbol{\theta}) = \int_a^{g[F_{\mathcal{G}_{i_0}}(x; \boldsymbol{\theta}_{i_0})]} dF_{\mathcal{H}}(x; \boldsymbol{\rho}) = F_{\mathcal{H}}[g(x; \boldsymbol{\theta}_{i_0}); \boldsymbol{\rho}] - F_{\mathcal{H}}[a] = F_{\mathcal{N}}(x; \boldsymbol{\delta}) - F_{\mathcal{H}}(a; \boldsymbol{\rho}) \quad (5.10)$$

em que $F_{\mathcal{H}}(a; \boldsymbol{\rho}) = 0$, $\boldsymbol{\theta} = (\boldsymbol{\delta}, \boldsymbol{\rho})$ e $\boldsymbol{\delta} = (\boldsymbol{\theta}_{i_0}, \boldsymbol{\rho})$.

Exemplo 16. *Considere o membro da família T-X de distribuições de probabilidade abaixo*

$$F_{\mathcal{T}-\mathcal{X}}(x; \boldsymbol{\theta}) = \int_n^{g[F_{\mathcal{G}_{i_0}}(x; \boldsymbol{\theta}_{i_0})]} dF_{\mathcal{H}}(x; \boldsymbol{\rho}) \quad (5.11)$$

em que $F_{\mathcal{H}}(x; \alpha, \beta) = 1 - \exp[-(\alpha x)^{\beta}]$ com $\boldsymbol{\rho} = (\alpha, \beta)$, $F_{\mathcal{G}_{i_0}}(x; a, b) = 1 - \exp[-(ax)^b]$. Seja g uma função tal como segue $g[F_{\mathcal{G}_{i_0}}(x; \boldsymbol{\theta}_{i_0})] = -\log\{1 - F_{\mathcal{G}_{i_0}}(x; \boldsymbol{\theta}_{i_0})\}$ e considere $n = 0$; levando em consideração as condições descritas neste parágrafo a expressão 5.11 pode ser escrita como:

$$F_{\mathcal{W}-\mathcal{W}}(x; \boldsymbol{\theta}) = \int_0^{-\log\{1 - \{1 - \exp[-(ax)^b]\}\}} d\{1 - \exp[-(\alpha y)^{\beta}]\}. \quad (5.12)$$

A expressão (5.12) é um caso particular da família introduzida por [Alzaatreh e Ghosh \(2015\)](#) e denominada de “On the Weibull-X family of distributions”.

Será provado que a distribuição de probabilidade Weibull-Weibull ($F_{\mathcal{W}-\mathcal{W}}(x; \cdot)$) é não identificável. Sejam $\boldsymbol{\alpha} = (a_1, b, \alpha_1, \beta)$, $\boldsymbol{\beta} = (a_2, b, \alpha_2, \beta) \in \Theta_{\mathcal{W}-\mathcal{W}}$ tais que $a_1 \neq a_2$ e $\alpha_1 \neq \alpha_2$. Usando a hipótese se tem

$$(a_1 x)^b \neq (a_2 x)^b.$$

Entretanto,

$$\exists \quad \alpha_1 \neq \alpha_2 : \alpha_1 (a_1 x)^b = \alpha_2 (a_2 x)^b \Rightarrow F_{\mathcal{W}-\mathcal{W}}(x; \boldsymbol{\alpha}) = F_{\mathcal{W}-\mathcal{W}}(x; \boldsymbol{\beta}).$$

Portanto, $\Theta_{\mathcal{W}-\mathcal{W}}$ é não identificável.

Exemplo 17. Considere o membro da família T-X de distribuições de probabilidade a seguir

$$F_{\mathcal{T}-\mathcal{X}}(x; \boldsymbol{\theta}) = \int_n^g [F_{\mathcal{G}_{i_0}}(x; \boldsymbol{\theta}_{i_0})] dF_{\mathcal{H}}(x; \boldsymbol{\rho}) \quad (5.13)$$

em que $F_{\mathcal{H}}(x; \alpha, \beta) = 1 - \exp[-\alpha(x)^\beta]$ com $\boldsymbol{\rho} = (\alpha, \beta)$, $F_{\mathcal{G}_{i_0}}(x; a, b) = 1 - \exp[-(a/x)^b]$. Seja g uma função tal como segue $g[F_{\mathcal{G}_{i_0}}(x; \boldsymbol{\theta}_{i_0})] = -\log[F_{\mathcal{G}_{i_0}}(x; \boldsymbol{\theta}_{i_0})]$ e seja $n = 0$; considerando as condições de contorno deste parágrafo a expressão 5.13 pode ser escrita como:

$$F_{\mathcal{W}-\mathcal{IW}}(x; \boldsymbol{\theta}) = \int_0^{-\log\{\{1 - \exp[-(a/x)^b]\}\}} d\{1 - \exp[-\alpha(y)^\beta]\} = \{1 - \exp[-\alpha(a/x)^{b\beta}]\}. \quad (5.14)$$

A expressão (5.14) corresponde a um caso específico da família de distribuições introduzida e analisada por [Tahir et al. \(2016\)](#), denominada "A new Weibull-G family of distributions". De acordo com [Cole \(2020\)](#), quando dois parâmetros presentes no modelo aparecem como um produto, não é possível estimá-los individualmente; somente o produto dos parâmetros pode ser estimado. Esse tipo de modelo é classificado como redundante em relação aos parâmetros, ou seja, os parâmetros são considerados não identificáveis. Um fenômeno análogo foi observado no Exemplo 2.

Portanto, $\Theta_{\mathcal{W}-\mathcal{IW}}$ é não identificável.

5.1.5 Conclusão do Capítulo 5

Neste capítulo, foram apresentados teoremas que garantem a identificabilidade de três subclasses da família T-**G**-X de distribuições. Além disso, foram analisados alguns

subconjuntos da família T-X de distribuições que enfrentaram problemas em termos de identificabilidade.

Conclui-se, portanto, que a introdução de uma nova classe ou família de fd 's requer a garantia da identificabilidade para que se torne uma ferramenta útil em futuros projetos de pesquisa para os pesquisadores interessados no tema.

6 Conclusão

No Capítulo 3 desta tese de doutorado, foram delineadas definições que formalizam conceitos muitas vezes implícitos e não explicitados na teoria da identificabilidade paramétrica. Essas definições abarcam termos como subfamília, vetores encaixados, vetores gerados por outros vetores, entre outros. Além disso, os teoremas, lemas, corolários e proposições apresentados neste capítulo servem como diretrizes para a classificação da identificabilidade de uma distribuição ou classe de distribuições de probabilidades. Essa classificação prática é exemplificada nos diversos exemplos e discussões ao longo da tese.

No Capítulo 4, o autor desta tese introduz um método construtivo de fd 's por meio da combinação, seja por soma ou multiplicação, de distribuições generalizadas. Diversos subcasos desse método foram analisados em termos de identificabilidade. Tal investigação foi conduzida por meio de teoremas, lemas, proposições e corolários introduzidos neste capítulo, evidenciando que algumas misturas de distribuições são identificáveis, enquanto outras, frequentemente empregadas em pesquisas diversas, não o são.

No Capítulo 5, é conduzido um estudo detalhado da classe de distribuições $T\text{-}\mathbf{G}\text{-}X$, a qual constitui um submétodo gerador de distribuições e classes de probabilidade apresentado por Brito et al. (2019). O método $T\text{-}\mathbf{G}\text{-}X$ é delineado, e subcasos (do mesmo) são examinados em termos da identificabilidade paramétrica. Teoremas e proposições pertinentes aos subcasos mencionados são apresentados para orientar a classificação deles como identificáveis ou não.

Como contribuição significativa desta tese de doutorado, destaca-se o enriquecimento da teoria da identificabilidade paramétrica com formalizações e propriedades inovadoras, previamente não exploradas na pesquisa e revisão da literatura estatística. Espera-se que este trabalho represente uma modesta, porém valiosa, contribuição para o avanço da ciência neste campo da inferência estatística paramétrica.

7 Trabalhos Futuros

Para futuras investigações, várias direções de pesquisa podem ser exploradas com base nos resultados e contribuições apresentados nesta tese. Algumas sugestões incluem:

- Extensão das formalizações teóricas: Continuar a explorar as formalizações introduzidas no Capítulo 3, investigando outras facetas da teoria da identificabilidade paramétrica e desenvolvendo novas definições e teoremas que possam aprofundar o conhecimento sobre o assunto;
- Explorar aplicações práticas do método construtivo de fd 's introduzido no Capítulo 4 em diferentes domínios, como modelagem estatística e análise de dados, a fim de avaliar sua eficácia e aplicabilidade em cenários do mundo real;
- Investigar outras classes de distribuições além da $T\text{-}\mathbf{G}\text{-}X$ discutida no Capítulo 5, analisando sua identificabilidade paramétrica e buscando possíveis extensões ou generalizações que possam ampliar o escopo de aplicação dessas distribuições;
- Realizar aplicações das teorias desenvolvidas nesta tese, empregando conjuntos de dados reais para testar a identificabilidade de distribuições em diversos contextos e cenários.

Referências Bibliográficas

- AHMAD, K. E. Identifiability of finite mixtures using a new transform. **Annals of the Institute of Statistical Mathematics**, v. 40, p. 261–265, 1988.
- AHMAD, K. E.; AL-HUSSAINI, E. K. Remarks on the non-identifiability of mixtures of distributions. **Annals of the Institute of Statistical Mathematics**, v. 34, p. 543–544, 1982.
- AL-HUSSAINI, E.; AHMAD, K. On the identifiability of finite mixtures of distributions. **IEEE Transactions on Information Theory**, v. 27, n. 5, p. 664–668, 1981.
- ALEXANDER, C. et al. Generalized beta-generated distributions. **Computational Statistics and Data Analysis**, v. 56, p. 1880–1897, 2012.
- ALZAATREH, A.; GHOSH, I. A on the Weibull-x family of distributions. **Journal of statistical theory and applications**, v. 14, p. 169–183, 2015.
- ALZAATREH, A.; LEE, C.; FAMOYE, F. A new method for generating families of continuous distributions. **Metron**, v. 71, p. 63–79, 2013.
- ANDRADE, T. et al. **Journal of Data Science**, v. 17, p. 299 –330, 2019.
- ANDRADE, T.; GOMES-SILVA, F.; ZEA, L. Mathematical properties, application and simulation for the exponentiated generalized standardized gumbel distribution. **Acta Scientiarum Technology**, p. 1807 – 8664, 2018.
- ANDRADE, T. et al. The exponentiated generalized gumbel distribution. **JRevista Colombiana de Estadística**, v. 38, n. 1, p. 123 – 143, 2015.
- ANDRADE, T.; ZEA, L. The exponentiated generalized extended pareto distribution. **Journal of Data Science**, v. 16, p. 781 – 800, 2018.
- ANDRADE, T. A. N. et al. The exponentiated generalized extended exponential distribution. **Journal of Data Science**, v. 14, p. 393 – 414, 2016.
- BARNDORFF-NIELSEN, O. Identifiability of mixtures of exponential families. *Journal of Mathematical Analysis and Applications*, v. 12, n. 1, p. 115–121, 1965.
- BASSE, G.; BOJINOV, I. A general theory of identification. 2020.
- BRITO, C. R. et al. Method for generating distributions and classes of probability distributions: the univariate case. **Hacettepe Journal of Mathematics and Statistics**, v. 48, n. 3, p. 897–930, 2019.
- CARRASCO, J. M.; ORTEGA, E. M.; CORDEIRO, G. M. A generalized modified Weibull distribution for lifetime modeling. **Computational Statistics and Data Analysis**, v. 53, n. 2, p. 450–462, 2008.

- CASELLA, G.; BERGER, R. L. **Statistical Inference**. Pacific Grove, California, USA: Thomson Learning, 2001. (Duxbury Advanced Series).
- CATCHPOLE, E. A.; MORGAN, B. J. T. Detecting parameter redundancy. **Biometrika**, v. 84, p. 187–196, 1997.
- CHANDRA, S. On the mixtures of probability distributions. **Scandinavian Journal of Statistics**, v. 4, n. 3, p. 105–112, 1977.
- COLE, D. J. **Parameter redundancy and identifiability**. Boca Raton, Florida, USA: Taylor and Francis Group, CRC Press, 2020. (Chapman and Hall interdisciplinary statistics).
- CORDEIRO, G. M.; CASTRO, M. A new family of generalized distributions. **Journal of Statistical Computation and Simulation**, v. 81, p. 883–898, 2011.
- CORDEIRO, G. M.; LEMONTE, A.; CAMPELO, A. K. Extended arcsine distribution to proportional data: properties and applications. **Studia Scientiarum Mathematicarum Hungarica**, v. 53, p. 440 – 466, 2017.
- CORDEIRO, G. M.; LEMONTE, A. J. The exponentiated generalized birnbaum-saunders distribution. **Applied Mathematics and Computation**, v. 247, p. 762–779, 2014.
- CORDEIRO, G. M.; ORTEGA, E. M. M.; CUNHA, D. C. C. da. The exponentiated generalized class of distributions. **Journal of Data Science**, v. 11, p. 1–27, 2013.
- DOBSON, A. J. **An introduction to generalized linear models**. [S.l.]: CRC Press, 2018. (Texts in statistical science).
- ELGARHY, M.; HAQ, M. A. U.; AIN, Q. ul. Exponentiated generalized Kumaswamy distribution with applications. **Annals of Data Science**, v. 5, p. 273–292, 2018.
- EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. **Communications in Statistics - Theory and Methods**, v. 31, p. 497–512, 2002.
- EVERITT, B. S.; HAND, D. J. **Finite mixture distributions**. 733 Third Avenue, New York. NY 10017, USA: Chapman and Hall, 1981. (Monographs on applied probability and statistics).
- FRASER, M. D.; HSU, Y. H.; WALKER, J. J. Identifiability of finite mixtures of von Mises distributions. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 9, n. 5, p. 1130 – 1131, 1981. Disponível em: <<https://doi.org/10.1214/aos/1176345595>>.
- FREEDMAN, R. S. Some new results on binary relations. **CoRR**, abs/1501.01914, 2015. Disponível em: <<http://arxiv.org/abs/1501.01914>>.
- FRISCH, R. **Statistical confluence analysis by means of complete regression systems**. Oslo, Noruega: University Institute of Economics, 1934.
- GOMES-SILVA, F. et al. An extended dagum distribution: Properties and applications. **International Journal of Applied Mathematics and Statistics**, v. 56, p. 35 –56, 2017.

- GUSMÃO, F. R. S. et al. Analyzing and solving the identifiability problem in the exponentiated generalized Weibull distribution. **Journal of the Egyptian Mathematical Society**, Springer, v. 29, n. 21, p. 2090–9128, 2021.
- GUSMÃO, F. R. S.; ORTEGA, E. M. M.; CORDEIRO, G. M. The generalized inverse Weibull distribution. **Statistical Papers**, Springer, v. 52, n. 3, p. 591–619, 2011.
- HALMOS, P. R. **Naive set theory**. [S.l.]: Courier Dover Publications, 2017.
- HAQ, M. A. u. Kumaswamy exponentiated inverse Rayleigh distribution. **Mathematical Theory and Modeling**, v. 6, n. 3, p. 2224–5804, 2016.
- HOLZMANN, H.; MUNK, A.; GNEITING, T. Identifiability of finite mixtures of elliptical distributions. **Scandinavian Journal of Statistics**, v. 33, p. 753–763, 2006.
- HOLZMANN, H.; MUNK, A.; STRATMANN, B. Identifiability of finite mixtures - with applications to circular distributions. **Sankhya**, v. 66, p. 440–449, 2004. Disponível em: <https://api.semanticscholar.org/CorpusID:126017826>.
- HSIAO, C. **Identification**. [S.l.]: Elsevier, 1983. v. 1. 223-283 p. (Handbook of econometrics 1, v. 1).
- HURWICZ, L. **Generalization of the concept of identification**. [S.l.]: New York John Wiley and Sons, Inc., 1950. v. 10. 245-257 p. (Statistical Inference in Dynamic Economic Models, Number 10 in Cowles Commission for Research in Economics, v. 10).
- JOHNSON, N. L. N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous univariate distributions**. [S.l.]: Wiley-Interscience, 1994. v. 1. (Wiley Series in Probability and Statistics, v. 1).
- JONES, G. et al. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. **Biometrics**, v. 66, n. 3, p. 855–863, 2010.
- KELLER, A. Z.; KAMATH, A. R. R.; PERERA, U. D. Reliability analysis of CNC machine tools. **Reliability engineering**, v. 3, n. 6, p. 449–473, 1982.
- KENT, J. T. Identifiability of finite mixtures for directional data. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 11, n. 3, p. 984 – 988, 1983. Disponível em: <https://doi.org/10.1214/aos/1176346264>.
- KOOPMANS, T. C. Identification problems in economic model construction. **Econometrica, Journal of the Econometric Society**, p. 125–144, 1949.
- KOOPMANS, T. C.; REIERSOL, O. The identification of structural characteristics. **The Annals of Mathematical Statistics**, v. 21, n. 2, p. 165–181, 1950.
- KUMAR, D. et al. Parameter estimation for the exponentiated Kumaswamy-power function distribution based on order statistics with application. **Annals of Data Science**, v. 8, n. 4, p. 785–811, 2021.
- LEHMANN, E. L.; CASELLA, G. **Theory of point estimation**. [S.l.]: Springer Science and Business Media, 2006.

- LIMA, E. L. **Curso de Análise - Vol. 1.** [S.l.]: Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2008. v. 1. (Projeto Euclides, v. 1).
- LINK, W. A. Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. **Biometrics**, v. 59, n. 4, p. 1123–1130, 2003.
- MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; FREIMER, M. The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. **Technometrics**, Taylor and Francis, v. 37, n. 4, p. 436–445, 1995.
- MUHAMMED, H. Z. A class of bivariate modified weighted distributions: Properties and applications. **Annals of Data Science**, Springer, p. 2198–5812, 2021.
- NADARAJAH, S.; CORDEIRO, G. M.; ORTEGA, E. M. M. The exponentiated Weibull distribution: a survey. **Statistical Papers**, v. 54, p. pages 839–877, 2013.
- OGUNTUNDE, P. E.; ODETUNMIBI, O. A.; ADEJUMO, A. O. On the exponentiated generalized Weibull distribution: A generalization of the Weibull distribution. **Indian Journal of Science and Technology**, v. 8, n. 35, p. 0974–6846, 2015.
- PAULINO, C. D. M.; PEREIRA, C. A. B. On identifiability of parametric statistical models. **Journal of the Italian Statistical Society**, Giardini Editori, Pisa, v. 3, n. 1, p. 125–151, 1994. ISSN 1121-9130.
- PESCIM, R. R. et al. The beta generalized half-normal distribution. **Computational Statistics and Data Analysis**, v. 54, n. 4, p. 945–957, 2010.
- ROTHENBERG, T. J. Identification in parametric models. **Journal of the Econometric Society**, p. 577–591, 1971.
- SANTOS, A. L. P. d. **Métodos geradores de modelos de crescimento e decrescimento aplicados às ciências agrárias.** Tese (Doutorado) — Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Pernambuco, BR, 2019.
- SHAHBAZ, M. Q.; SHAHBAZ, S.; BUTT, N. S. The Kumaswamy inverse Weibull distribution. **Pakistan Journal of Statistics and Operation Research**, v. 8, n. 3, p. 479–489, 2012.
- SHAHBAZ, S. H.; ISMAIL, M.; BUTT, N. S. The mcdonald's inverse Weibull distribution. **Pakistan Journal of Statistics and Operation Research**, v. 12, n. 4, 2016.
- SILVA, R. et al. A new extended gamma generalized model. **International Journal of Pure and Applied Mathematics**, v. 100, n. 2, p. 309 – 335, 2015.
- SILVEIRA, F. V. J. et al. Normal-g class of probability distributions: Properties and applications. **Symmetry**, v. 11, 2019.
- SILVEIRA, F. V. J. et al. The normal-tangent-g class of probabilistic distributions: properties and real data modelling. **Pakistan Journal of Statistics and Operation Research**, v. 16, p. pages 827–838, 2020.

SILVEIRA, F. V. J. et al. Modelling wind speed with a univariate probability distribution depending on two baseline functions. **Hacettepe Journal of Mathematics and Statistics**, v. 52, n. 3, p. pages 808 – 827, 2023.

SILVEY, S. D. **Statistical Inference**. London, UK: Chapman and Hall CRC, 1975. (Monographs on Statistics and Applied Probability).

TAHIR, M. H. et al. A new Weibull-g family of distributions. **Hacettepe Journal of Mathematics and Statistics**, v. 45, n. 2, p. 629 – 647, 2016.

TEICHER, H. On the convolution of distributions. **The Annals of Mathematical Statistics**, v. 25, n. 4, p. 775 – 778, 1954.

TEICHER, H. On the mixture of distributions. **The Annals of Mathematical Statistics**, v. 31, p. 55 – 73, 1960.

TEICHER, H. Identifiability of mixtures. **The Annals of Mathematical Statistics**, v. 32, p. 244 – 248, 1961.

TEICHER, H. Identifiability of finite mixtures. **The Annals of Mathematical Statistics**, v. 34, n. 4, p. 1265 – 1269, 1963.

TEICHER, H. Identifiability of mixtures of product of measures. **The Annals of Mathematical Statistics**, v. 38, n. 4, p. 1300 – 1302, 1967.

XIE, M.; LAI, C. D. Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. **Reliability Engineering and System Safety**, v. 52, n. 1, p. 87–93, 1996.

YAKOWITZ, S. J.; SPRAGINS, J. D. On the identifiability of finite mixtures. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 39, n. 1, p. 209 – 214, 1968. Disponível em: <<https://doi.org/10.1214/aoms/1177698520>>.

YASSMEN, Y. A. Exponentiated transmuted generalized inverse Weibull distribution a generalization of the generalized inverse Weibull distribution. **Asian Research Journal of Mathematics**, v. 8, n. 2, p. 1–12, 2018.