

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**Modelos de resposta discreta com funções de ligação da família gumbel**

**Jessica Suzana Barragan Alves**

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Jessica Suzana Barragan Alves**

## Modelos de resposta discreta com funções de ligação da família gumbel

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

Área de Concentração: Estatística

Orientador: Prof. Dr. Jorge Luis Bazán Guzmán

**USP – São Carlos**  
**Maio de 2024**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

A474m      Alves, Jessica  
              Modelos de resposta discreta com funções de  
              ligação da família Gumbel / Jessica Alves;  
              orientador Jorge Luis Bázan Guzmán. -- São Carlos,  
              2024.  
              101 p.

              Tese (Doutorado - Programa Interinstitucional de  
              Pós-graduação em Estatística) -- Instituto de Ciências  
              Matemáticas e de Computação, Universidade de São  
              Paulo, 2024.

              1. Curvas Características de Itens Assimétricas.  
              2. Dados desbalanceados. 3. Distribuição Gumbel. 4.  
              Distribuição de Valor Extremo Generalizado (GEV). 5.  
              Função de ligações flexíveis cloglog. I. Bázan Guzmán,  
              Jorge Luis , orient. II. Título.

**Jessica Suzana Barragan Alves**

## Discrete response models with gumbel family link functions

Thesis submitted to the Institute of Mathematics and Computer Science – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Doctor in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Jorge Luis Bazán Guzmán

**USP – São Carlos**  
**May 2024**



*Dedico este trabalho à minha família, ao meu noivo e a todos aqueles que acreditam no poder transformador da educação.*





# AGRADECIMENTOS

---

---

Neste momento de profunda gratidão, olho para trás e reconheço cada pessoa que foi essencial nesta jornada. Quero expressar minha sincera gratidão a todos que estiveram ao meu lado, tornando esta conquista possível.

Agradeço a Deus por atender minhas preces e por iluminar meu caminho ao longo desta trajetória desafiadora.

Aos meus pais, Olivio e Maria, a base sólida e minha fonte de inspiração. Seu incansável empenho na nossa educação e os exemplos de honestidade e resiliência traçaram o caminho para um futuro melhor, não apenas para mim mas também para minhas irmãs.

Às minhas queridas irmãs, Rafaela e Tayla, pela verdadeira amizade. Às minhas sobrinhas, Maria Fernanda e Laura, pelo constante carinho que sempre demonstraram, tornando os dias exaustivos mais leves.

Meu amor e profunda gratidão ao meu noivo, Gustavo Dartibale, pelo apoio inabalável, incentivo constante e por ser meu confidente e porto seguro. Com amor, parceria e companheirismo, você tem sido uma parte essencial desta jornada.

Ao meu orientador, Professor Jorge Bazán, que não apenas acreditou no meu potencial mas também me guiou com gentileza, conselhos e conhecimento. Sua paciência e apoio foram fundamentais para meu crescimento profissional e pessoal.

Aos meus companheiros de jornada, Patrícia, Naiara, Gustavo, Alex e todos os outros colegas, pelas inúmeras noites de estudo, pelas lágrimas e sorrisos compartilhados contribuindo com conhecimento e apoio mútuo.

Às amigas Eloisa, Vanessa, Talita e Laís, por estenderem suas mãos quando mais precisei, por serem ouvidos atentos às minhas preocupações e por incentivarem minha continuidade.

Enfim, expresso minha imensa gratidão a todos que confiaram em mim, partilharam alegrias e dificuldades, tornando esta jornada uma experiência inestimável e memorável. Cada um de vocês é parte indelével da minha história e, por isso, serei eternamente grata. Muito obrigada!



*“Não há sonho impossível para aqueles que realmente acreditam que o poder realizador reside no interior de cada ser humano.”*  
*(Eleanor Roosevelt)*



# RESUMO

ALVES, J.S.B. **Modelos de resposta discreta com funções de ligação da família gumbel.** 2024. 101 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

O presente estudo concentra-se na introdução e desenvolvimento de modelos estatísticos assimétricos para lidar com dados desbalanceados em regressões binomiais e na Teoria de Resposta ao Item (TRI). Inicialmente, é abordada a função de ligação *loglog* complementar, introduzida por Fisher em 1922, como uma alternativa assimétrica às funções de ligação *logit* e *probit*. Propõem-se variações flexíveis dessa função para modelar a regressão binomial, incluindo parâmetros adicionais que explicam o desbalanceamento nos resultados binomiais. Para a inferência dos modelos, desenvolve-se uma abordagem Bayesiana utilizando métodos de cadeias de Markov Monte Carlo. Além disso, explora-se a relação entre Curvas Características de Itens (CCI) assimétricas na TRI para dados de resposta binária desbalanceados. São propostos novos modelos de TRI com CCI assimétricas como característica principal, incluindo o modelo TRI *cloglog* como um caso especial. Destacam-se a importância desses modelos na análise de dados educacionais e compara sua eficácia com outros modelos propostos na literatura de TRI. Adicionalmente, apresenta-se dois novos modelos de teoria de resposta ao item baseados na distribuição Generalizada de Valores Extremos (GEV). Discuti-se a estimação Bayesiana desses modelos e são demonstradas sua aplicabilidade por meio de estudos de simulação e análise de dados reais de testes matemáticos em escolas públicas no Peru. Esses modelos mostram-se promissores para lidar com desbalanceamentos e assimetrias em dados binários, oferecendo uma abordagem robusta e flexível para análise estatística em diversos contextos, incluindo saúde, educação e avaliação de testes.

**Palavras-chave:** Curvas Características de Itens Assimétricas, Dados desbalanceados, Distribuição Gumbel, Distribuição de Valor Extremo Generalizado (GEV), Distribuição Logística Expoente positivo (LPE), Estimação Bayesiana, Função de ligações flexíveis *cloglog*, Teoria de Resposta ao Item (TRI).



# ABSTRACT

ALVES, J.S.B. **Discrete response models with gumbel family link functions**. 2024. 101 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

The present study focuses on the introduction and development of asymmetrical statistical models to address imbalanced data in binomial regressions and within Item Response Theory (IRT). Initially, we delve into the complementary log-log link function, introduced by Fisher in 1922, as an asymmetrical alternative to the logit and probit link functions. We propose flexible variations of this function to model binomial regression, incorporating additional parameters that account for imbalances in the binomial outcomes. For model inference, we develop a Bayesian approach employing Monte Carlo Markov chain methods. Furthermore, we investigate the relationship between asymmetrical Item Characteristic Curves (ICCs) within IRT for imbalanced binary response data. We propose new IRT models with asymmetrical ICCs as their primary feature, including the cloglog IRT model as a special case. We emphasize the significance of these models in educational data analysis and compare their efficacy against other models proposed in the IRT literature. Additionally, we introduce two new item response theory models based on the Generalized Extreme Value (GEV) distribution. We discuss Bayesian estimation methods for these models and demonstrate their applicability through simulation studies and analysis of real-world data from mathematical tests in public schools in Peru. These models show promise in handling imbalances and asymmetries in binary data, providing a robust and adaptable statistical approach across various domains, including healthcare, education, and test assessment.

**Keywords:** Skewed Item Characteristic Curves, Unbalanced Data, Gumbel Distribution, Generalized Extreme Value Distribution (GEV), Logistic Positive Exponent Distribution (LPE), Bayesian Estimation, Link Function *cloglog*, Item Response Theory (IRT).





# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Funções de densidade de probabilidade para as distribuições GEV-G, GEV-RG, PR-G, PR-RG, TS-G e TS-RG considerando diferentes valores de $\lambda \in \mathbb{R}$	30
Figura 2 – Funções de distribuição acumuladas para distribuições GEV-G, GEV-RG, P-RG, RP-RG, TS-G e TS-RG considerando diferentes valores de $\lambda \in \mathbb{R}$	32
Figura 3 – Resíduos de quantis randomizados normalizados para regressão binomial com as funções de ligação GEV-G, GEV -RG e RP-RG para dados de Warsaw Girls	42
Figura 4 – Curvas de probabilidade para $\lambda = -0,9, -0,5, 0, 0,5, 0,9$ nos modelos P-RG, RP-RG, TS-G e TS-RG considerando diferentes faixas para $\theta$ e $a = 1$ e $b = 0$	50
Figura 5 – Curvas características do item (CCIs) para os itens 14, 2 e 11 nos modelos TRI RP-RG e RLPE nos dados matemáticos	54
Figura 6 – Relação entre os parâmetros de discriminação, dificuldade e aceleração, para o modelo TRI RP-RG	55
Figura 7 – Curvas de densidade das distribuições GEV-G e GEV-RG para diferentes $\lambda$ s	61
Figura 8 – Curvas característica do item para diferentes $\lambda$ s no modelo GEV-G TRI considerando diferentes intervalos para $\theta, a$ e $b$	63
Figura 9 – Itens 14, 2 e 11 para o conjunto de dados do teste de matemática	73
Figura 10 – Curvas características do item (ICCs) para os itens 14, 2 e 11 sob os modelos de TRI GEV-RG 2P e RLPE 2P nos dados de matemática	74
Figura 11 – Relação entre os parâmetros $\hat{a}$ , $\hat{b}$ e $\hat{\lambda}$ para o modelo TRI GEV-RG 2P	75
Figura 12 – Distribuições das estimativas dos parâmetros ( $\hat{\theta}$ ) para os modelos TRI GEV-RG 2P e RLPE 2P	75



# LISTA DE TABELAS

---

---

Tabela 1 – Modelos binomiais flexíveis obtidos a partir de generalizações das distribuições G e RG . . . . .	31
Tabela 2 – Resultados da análise de sensibilidade dos modelos GEV-G, GEV-RG, P-RG, RP-RG, TS-G e TS-RG para as três distribuições <i>à Priori</i> . . . . .	36
Tabela 3 – Estimativas de parâmetros para os modelos propostos com dados desbalanceados simulados para cada modelo com $\lambda = 0,3$ usando o pacote rstan . . . . .	38
Tabela 4 – Estimativas de parâmetros para os modelos propostos com dados desbalanceados simulados para cada modelo com $\lambda = -0,3$ usando o pacote rstan . . . . .	39
Tabela 5 – Comparação de modelos para dados das meninas de Varsóvia . . . . .	41
Tabela 6 – Estimativas do Parâmetro para os três melhores modelos: GEV-G, GEV-RG e RP-RG usando o algoritmo NUTS nos dados das meninas de Varsóvia . . . . .	41
Tabela 7 – Estimativas do Parâmetro para os três melhores modelos: GEV-G, GEV-RG e RP-RG usando o algoritmo RWM nos dados das meninas de Varsóvia . . . . .	41
Tabela 8 – Comparação de modelos para dados de Infecção Pulmonar . . . . .	43
Tabela 9 – Estimativas de parâmetros para o modelo de ligação RP-RG para dados de infecção pulmonar. . . . .	43
Tabela 10 – Critérios de comparação de modelos para dados de teste de matemática . . . . .	53
Tabela 11 – Parâmetros de itens para modelos TRI alternativos para o item 14, item 2 e item 11 em dados matemáticos . . . . .	53
Tabela 12 – Resultados da análise de sensibilidade dos modelos GEV-G 1P, GEV-G 2P, GEV-RG 1P, GEV-RG 2P para as quatro distribuições de prioris . . . . .	66
Tabela 13 – Estimativas de parâmetros para o modelo TRI GEV-G 2P com dados simulados desbalanceados . . . . .	69
Tabela 14 – Estimativas de parâmetros para o modelo TRI GEV-RG 2P com dados simulados desbalanceados . . . . .	70
Tabela 15 – Desempenho dos critérios de ajuste para comparação de modelos . . . . .	71
Tabela 16 – Critérios de comparação de modelos para os dados do teste de matemática . . . . .	72
Tabela 17 – Parâmetros dos itens 14, 2 e 11 para modelos alternativos de TRI nos dados de matemática . . . . .	74
Tabela 18 – Critérios de comparação de modelos para os dados do LSAT . . . . .	76



# SUMÁRIO

---

---

1	INTRODUÇÃO . . . . .	21
1.1	Objetivos e Organização da Tese . . . . .	22
2	FUNÇÕES DE LIGAÇÃO <i>CLOGLOG</i> FLEXÍVEIS PARA MODELOS DE REGRESSÃO BINOMIAL . . . . .	23
2.1	Modelos de regressão binomial . . . . .	23
2.2	Distribuições flexíveis Reversa Gumbel . . . . .	26
2.3	Modelos de regressão binomial usando a função de ligação <i>cloglog</i> flexíveis . . . . .	29
2.4	Estimação Bayesiana . . . . .	31
2.4.1	<i>Especificação da distribuição à priori</i> . . . . .	31
2.4.2	<i>Critérios de comparação de modelos e análise de resíduos</i> . . . . .	34
2.5	Estudos de simulação . . . . .	35
2.5.1	<i>Análise de Sensibilidade da distribuição à priori para o parâmetro <math>\lambda</math></i> . . . . .	35
2.5.2	<i>Estudo de Recuperação de Parâmetros</i> . . . . .	36
2.6	Aplicações em Dados Médicos . . . . .	39
2.6.1	<i>Dados garotas de Varsóvia</i> . . . . .	40
2.6.2	<i>Dados de Infecção Pulmonar</i> . . . . .	41
2.7	Comentários Finais . . . . .	44
3	NOVOS MODELOS FLEXÍVEIS DE TRI PARA RESPOSTA DI- COTÔMICA . . . . .	45
3.1	A Teoria de Resposta ao item . . . . .	45
3.2	Modelos TRI tradicionais . . . . .	47
3.3	Modelos de TRI dicotômicos flexíveis . . . . .	48
3.4	Inferência . . . . .	49
3.4.1	<i>Função de Verossimilhança</i> . . . . .	51
3.4.2	<i>Abordagem Bayesiana</i> . . . . .	51
3.4.3	<i>Especificação das distribuição à priori</i> . . . . .	51
3.4.4	<i>Critérios de comparação de modelos</i> . . . . .	52
3.5	Aplicação . . . . .	52
3.6	Comentários finais . . . . .	55

4	UM NOVO MODELO DE TEORIA DE RESPOSTA AO ITEM COM A DISTRIBUIÇÃO GENERALIZADA DO VALOR EXTREMO . . .	57
4.1	O Modelo TRI GEV-G . . . . .	60
4.1.1	<i>A Distribuição GEV</i> . . . . .	60
4.1.2	<i>Os modelos propostos</i> . . . . .	61
4.2	Inferência . . . . .	62
4.3	Estudos de Simulação . . . . .	65
4.3.1	<i>Análise de Sensibilidade da distribuição à priori para o parâmetro <math>\lambda</math></i> . . . . .	65
4.3.2	<i>Estudo de recuperação dos parâmetros</i> . . . . .	68
4.3.3	<i>Estudos dos Critérios de Comparação de Modelos</i> . . . . .	70
4.4	Aplicações . . . . .	71
4.4.1	<i>Teste de Matemática</i> . . . . .	71
4.4.2	<i>LSAT</i> . . . . .	76
4.5	Comentários Finais . . . . .	77
5	COMENTÁRIOS FINAIS E DESENVOLVIMENTOS FUTUROS . .	79
5.1	Conclusões . . . . .	79
5.2	Produções . . . . .	80
5.3	Desenvolvimentos Futuros . . . . .	80
	REFERÊNCIAS . . . . .	83
APÊNDICE A	CÓDIGOS - CAPÍTULO 2 . . . . .	91
APÊNDICE B	CÓDIGOS - CAPÍTULO 3 . . . . .	93
APÊNDICE C	CÓDIGOS - CAPÍTULO 4 . . . . .	99

---

# INTRODUÇÃO

---

Dados binomiais são observações em experimentos repetidos, onde cada resposta se encaixa em uma de duas categorias, como sucesso ou fracasso, positivo ou negativo, acerto ou erro entre outras. Esses dados são comuns em diversas áreas, como ciências sociais, agricultura, genética e ciências do comportamento.

Muito se tem estudado a esses dados, principalmente no contexto de regressão binária onde sua probabilidade de sucesso está associada a atributos de interesse para explicar a variável resposta. Historicamente, funções de ligação simétricas, como probito e logito, foram frequentemente usadas para modelar dados desbalanceados. Porém alguns autores reconheceram que essa abordagem não era a mais apropriada em certos contexto (CHEN; DEY; SHAO, 1999). Foi a partir dessa constatação que começaram a aparecer na literatura novas funções de ligação assimétricas (Prentice (1976); Czado e Santner (1992); Guerrero e Johnson (1982); Stukel (1988)). Entretanto, nesses modelos propostos as funções de ligação *cloglog* e *loglog* foram pouco discutidas.

As novas funções de ligação assimétricas foram a fonte de inspiração para que se estendessem esses modelos de regressão binomial para a Teoria de Resposta ao Item (TRI) dicotômica. A TRI teve início com Lord (1952), inicialmente baseada na distribuição normal (ogiva normal). Posteriormente, Birnbaum (1968) contribuiu significativamente sugerindo substituir a função ogiva normal pela função logística, mais adequada matematicamente. Por muito tempo, a literatura da TRI manteve funções simétricas e suposições de normalidade. Entretanto, diversos autores questionaram essa suposição restritiva, apontando que a normalidade não representa adequadamente o comportamento humano (Samejima (1997); Micceri (1989); Bolfarine e Bazan (2010); Lee e Bolt (2018); Bazán, Bolfarine e Branco (2006)). Em seguida, Samejima (2000) propôs uma família de modelos com Curvas Características de Item (CCI) assimétricas, denominadas como Logística do expoente positivo (LPE), abrindo caminho para outros trabalhos nessa mesma linha de pesquisa (Bazán, Bolfarine e Branco (2006); Lee e Bolt (2018)).

Este estudo propõe uma família de funções de ligação assimétricas para a modelagem de dados binomiais, além de diferentes Curvas Características de Item assimétricas, considerando a flexibilização das funções de distribuição *cloglog* e *loglog* como base na Teoria de Resposta ao Item.

## 1.1 Objetivos e Organização da Tese

O objetivo principal desta tese é propor, apresentar e detalhar novas generalizações de funções de ligação aplicadas a dados binomiais, além de introduzir novas generalizações de Curvas Características assimétricas na teoria de resposta ao item. Essas generalizações são desenvolvidas a partir da utilização das funções *cloglog* e *loglog* como base. A metodologia de estimação Bayesiana é desenvolvida para os modelos obtidos e são apresentadas aplicações tanto na área da saúde quanto na avaliação educacional para os modelos propostos.

No Capítulo 2 é apresentada uma revisão da literatura sobre as generalizações existentes e sua importância para dados não balanceados. Nele também expõem-se as novas funções de ligação *cloglog* flexíveis no contexto da regressão binomial. Um estudo de simulação é desenvolvido usando as novas ligações para testar a recuperação de parâmetros dos algoritmos de estimativa. Além disso, são exibidas duas aplicações para ilustrar suas vantagens em relação a outros modelos, utilizando diferentes critérios de comparação de modelos.

No Capítulo 3, o objetivo é expandir a compreensão da Teoria de Resposta ao Item (TRI), introduzindo uma nova categoria de CCI. Além disso, apresenta-se, a metodologia de estimação bayesiana pelo método de Monte Carlo via Cadeias de Markov (MCMC) para os modelos assimétricos da TRI. Por fim, demonstra-se a aplicabilidade e a relevância desses novos modelos propostos por meio de sua implementação em conjuntos de dados gerais.

No Capítulo 4, o objetivo consiste em aprofundar a discussão de dois modelos propostos para a Teoria de Resposta ao Item (TRI). Realiza-se um estudo de sensibilidade das distribuições *a priori* para determinar a melhor abordagem para os modelos propostos. Além disso, conduz-se uma análise da recuperação dos parâmetros para garantir a eficácia e o bom funcionamento dos modelos propostos. Por fim, demonstra-se a eficiência e a aplicabilidade desses modelos por meio de duas aplicações distintas, destacando suas vantagens em relação a outros modelos, utilizando-se de diferentes critérios de comparação.

No Capítulo 5, apresenta-se algumas perspectivas e possíveis direções futuras, incluindo propostas complementares, derivadas dos temas abordados que servem como orientação para pesquisas subsequentes.



---

# FUNÇÕES DE LIGAÇÃO *CLOGLOG* FLEXÍVEIS PARA MODELOS DE REGRESSÃO BINOMIAL COMO ALTERNATIVA PARA DADOS MÉDICOS DESBALANCEADOS

---

Neste capítulo, são apresentadas novas funções de ligação flexíveis para dados desbalanceados na regressão binomial, adotando uma abordagem Bayesiana e utilizando o algoritmo NUTS (No-U-Turn Sampler), uma extensão do método HMC (Hamiltoniano Monte Carlo).

## 2.1 Modelos de regressão binomial

Modelos de regressão binomial são comuns em ciências sociais, ciências médicas, bem como em genética, agricultura e ciências comportamentais. Dados binários são a fonte de modelos binomiais, estudos epidemiológicos, modelagem multinível, análise de dados longitudinais, meta análise e Teoria de Resposta ao Item.

Sejam  $Y_1, Y_2, \dots, Y_m$  variáveis aleatórias discretas independentes (v.a), com  $Y_i \in \{0, \dots, n_i\}$  para  $i = 1, \dots, m$  e  $\sum_{i=1}^m n_i = N$ . Seja  $x_i = (x_{i1}, \dots, x_{ik})^T$  um vetor de covariáveis  $k \times 1$ , onde  $x_{i1} = 1$  corresponde ao intercepto, e seja  $\beta = (\beta_1, \dots, \beta_k)^T$  um vetor de coeficientes de regressão  $k \times 1$ . A regressão binomial geralmente assume que

$$Y_i | \beta \sim \text{Binomial}(n_i, p_i), \quad p_i = F(\eta_i), \quad \eta_i = x_i^T \beta, \quad i = 1, \dots, m, \quad (2.1)$$

onde  $F(\cdot)$  denota a função de distribuição acumulada (fda),  $\eta_i = x_i^T \beta$  é o  $i$ -ésimo preditor linear, e  $F^{-1}(\cdot)$  é chamada de função de ligação.

Quando  $F$  é a fda de uma distribuição simétrica, a função de ligação resultante é si-

métrica. Nesse caso, a curva de resposta para  $p_i$  é simétrica aproximadamente em  $p_i = 0,5$ . Exemplos desta situação são a fda de uma distribuição elíptica ou simétrica, como a Normal Padrão (*probit*), Logística (*logit*), *t*-Student, Exponencial (Laplace) e distribuições de Cauchy (*cauchit*) (Albert e Chib (1993); Naranjo, Martín e Pérez (2014)). No entanto, vários autores argumentam que funções de ligação simétricas podem ser inadequadas em casos nos quais os dados são desbalanceados. Isso ocorre quando uma das classes é muito menor que a outra, ou seja, dados resultantes da presença de valores iguais a 0 (ou 1) em proporção significativamente maior que os correspondentes valores reais de 1 (ou 0).

Por exemplo, Chen, Dey e Shao (1999) estabeleceram que se a probabilidade de uma dada resposta binária se aproximar de 0 a uma taxa diferente da qual se aproxima de 1, então as funções de ligação simétricas são inapropriadas, neste caso, as ligações assimétricas são recomendadas.

Um método para obter funções de ligação assimétrica em (2.1) é considerar a função  $F(\cdot)$  como sendo a fda de uma distribuição assimétrica. Nesse caso, a curva de resposta para a função  $p_i$  não tem uma forma simétrica em torno de  $p_i = 0,5$ , ou seja, o ponto de inflexão não é mais 0,5. Este é o caso da função de ligação *cloglog*, que é obtida quando  $F(\cdot)$  se torna a fda da distribuição Gumbel Reversa. Porém, neste caso específico, assim como no caso da função de ligação *loglog* (obtida quando  $F$  é a fda da distribuição Gumbel), não é possível modelar (ou controlar) a assimetria, além de não existir uma relação estabelecida entre essas funções de ligação. Outros exemplos mais recentes incluem Wang, Dey *et al.* (2010); Bazán, Bolfarine e Branco (2010); Li *et al.* (2016); Bazán *et al.* (2017) e Lemonte e Bazán (2018).

É importante notar que a assimetria pode ser introduzida mesmo quando a  $F$  é uma distribuição simétrica adicionando um parâmetro de assimetria  $\lambda$ . Isso implica substituir o preditor linear  $\eta_i$  por uma função contínua de  $\eta_i$ , denotada como  $k(\eta_i; \lambda)$ . Essa abordagem foi adotada, por exemplo, por Prentice (1976) e Czado e Santner (1992), que utilizaram a distribuição padrão de Burr. Nessa distribuição, a função  $k(\eta_i; \lambda)$  é definida como  $1 - 1 + \exp(\eta_i)^{-\lambda}$ , para  $\lambda > 0$  e  $\eta_i \in \mathbb{R}$ . Outros estudos, como os de Guerrero e Johnson (1982) e Stukel (1988), empregaram a distribuição Logística padrão, com  $k(\eta_i; \lambda) = \log \eta_i / (1 - \eta_i)$ , se  $\lambda \geq 0$ , e  $k(\eta_i; \lambda) = \lambda^{-1} [\eta_i / (1 - \eta_i)^\lambda - 1]$ , se  $\lambda \neq 0$ . Czado (1994), por sua vez, utilizou a distribuição Normal padrão, considerando  $k(\eta_i; \lambda) = (\eta_i + 1)^\lambda - 1 / \lambda$  se  $\eta_i \geq 0$ , e  $\eta_i$ , caso contrário.

Seguindo esta mesma ideia, Basu e Mukhopadhyay (2000) e Haro-López, Mallick e Smith (2000) propuseram uma classe geral obtida escolhendo  $F$  na classe de fda correspondente à família de misturas de escala elíptica dada por  $\mathcal{F} = \{F(\cdot) = \int_{[0, \infty]} H(\cdot | \lambda) dG(\lambda)\}$ , onde  $G$  é um fda em  $[0, \infty]$  e  $H$  é um fda elíptica. Usando uma representação estocástica aumentada e uma classe de distribuições de mistura, Chen, Dey e Shao (1999) e Chen, Dey e Shao (2001) consideraram um procedimento para modelar assimetria baseado em aumento de dados ou estrutura linear latente. Eles assumiram, especificamente, que  $Y_i = 1 \times I(W_i \geq 0)$ , com  $W_i = x_i' \beta + \delta Z_i + \varepsilon_i$ ,  $Z_i \sim G$ ,  $\varepsilon_i \sim H$ ,  $Z_i$  e  $\varepsilon_i$  independentes,  $G$  é a fda de uma distribuição assimétrica,

$H$  é a fda de uma distribuição simétrica e  $I(\cdot)$  é a função indicadora. Além disso, o parâmetro  $\delta \in (-\infty, \infty)$  é um parâmetro de assimetria. Uma representação semelhante foi usada por Kim, Chen e Dey (2008) para propor um  $t$ -link generalizado. Outra formulação da definição de funções de ligação assimétricas foi proposta por Jiang *et al.* (2013), substituindo  $F(\eta_i)$  em (2.1) por  $F(\eta_i|\delta) = F_0^\delta(\eta_i/\delta) \times I(0 < \delta < 1) + [1 - F_0^{1/\delta}(-\delta\eta_i)] \times I(\delta > 1)$  na qual  $F_0$  é a fda de uma distribuição de linha de base simétrica.

Apesar de já se ter na literatura funções de ligação assimétricas que generalizem ligações simétricas, como as funções de ligação *probit* (BLISS, 1935) e *logit* (BERKSON, 1944), de acordo com o conhecimento adquirido a respeito, generalizações da função de ligação *cloglog* que é uma distribuição assimétrica foi pouco discutidas na literatura. Isso é verdade, apesar do fato de que a função de ligação *cloglog* foi proposta, pela primeira vez, no trabalho de Fisher (1922).

Este capítulo é dedicado ao desenvolvimento de uma nova família paramétrica de ligações assimétricas. Mais especificamente, partindo da distribuição Reverse Gumbel como distribuição base, desenvolve-se a formulação de novas funções de ligação assimétricas com um parâmetro extra associado à forma. Essa escolha se deu pelo fato dessa função ser uma distribuição assimétrica e pouco utilizada na literatura para este tipo de abordagem.

Também é apresentado, neste trabalho, um tratamento unificado de várias generalizações da função de ligação *cloglog* para modelagem de resposta binomial, que possui as seguintes características: i)  $p_i$  é obtido substituindo  $F$  em (2.1) pela fda de uma função de distribuição avaliada no preditor linear, ii) o parâmetro de assimetria é indexado à distribuição de  $F$  e não está relacionado com o preditor linear, e iii) uma mistura ou estrutura linear latente não é necessária para formulação da função de ligação. A partir desta abordagem, são obtidas novas funções de ligação assimétricas que podem ser utilizadas para a modelagem de respostas binomiais desbalanceadas.

O capítulo está organizado da seguinte forma: a Seção 2.2 apresenta a distribuição Reverse Gumbel e versões flexíveis desta distribuição que podem ser consideradas como funções de ligação para conjuntos de dados binomiais; na Seção 2.3, é apresentado o modelo de regressão binomial usando ligações *cloglog* flexíveis; na Seção 2.4, os principais resultados referentes à análise bayesiana, critérios de comparação e análise de resíduos utilizados na aplicação; na Seção 2.5, um estudo de simulação é desenvolvido usando as novas funções de ligação flexíveis *cloglog* para testar a recuperação de parâmetros dos algoritmos de estimação; na Seção 2.6, a metodologia é exemplificada por meio de uma aplicação em dois conjuntos de dados médicos (Idade da menarca e Infecção pulmonar); por fim, na seção 2.7 comentários finais referente a esse primeiro capítulo.

## 2.2 Distribuições flexíveis Reversa Gumbel

Nesta seção, é apresentada, a classe de modelos flexíveis *loglog* e *cloglog* para respostas binomiais. Tem início com a definição de algumas distribuições importantes. A variável aleatória (v.a.)  $Y \in \mathbb{R}$  tem distribuição padrão *Gumbel* ( $G$ ) se sua correspondente função de densidade de probabilidade (fdp) e função de distribuição acumulada (fda) são dadas por

$$g(y) = \exp\{-y - \exp(-y)\} \quad \text{e} \quad G(y) = \exp\{-\exp(-y)\}, \quad (2.2)$$

respectivamente.

A v.a.  $Y \in \mathbb{R}$  é nomeada padrão *Gumbel Reversa* ( $RG$ ) se sua fdp e fda correspondentes forem dados por

$$f(y) = g(-y) = \exp\{y - \exp(y)\} \quad \text{e} \quad F(y) = 1 - G(-y) = 1 - \exp\{-\exp(y)\} \quad (2.3)$$

respectivamente. Note que se  $Y \sim G$ , então  $-Y \sim RG$ , e a recíproca também é verdadeira. Além disso, seguindo [Bazán et al. \(2017\)](#), ambas as distribuições satisfazem a propriedade de reversão, ou seja,  $G(y) + F(-y) = 1$ . Assim, a distribuição  $F$  é chamada de distribuição reversa de  $G$ .

Para a construção das novas funções de ligação, as fda das distribuições  $G$  e  $RG$  são consideradas como distribuições base. Em seguida, são introduzidas extensões que consideram três processos de generalização de distribuições.

O primeiro procedimento é baseado nas aproximações  $\lim_{\lambda \rightarrow 0} (1 + \lambda y)_+^{-1/\lambda} = \exp(-y)$  e  $\lim_{\lambda \rightarrow 0} (1 - \lambda y)_+^{-1/\lambda} = \exp(y)$ ,  $\lambda \in \mathbb{R}$ , que estão incluídas nas correspondentes distribuições  $G$  e  $RG$ , propostas por [Wang, Dey et al. \(2010\)](#).

O segundo procedimento é baseado na construção das distribuições de Potência e Potência Reversa, conforme consideradas em [Bazán et al. \(2017\)](#) e [Lemonte e Bazán \(2018\)](#). Nesse sentido, se  $H$  é uma distribuição base, então uma nova distribuição *Potência*  $H$  é definida a partir do procedimento de exponenciação, com fdp e fda dadas por:

$$h_P(y|\delta) = \delta h(y)H(y)^{\delta-1} \quad \text{e} \quad H_P(y|\delta) = H(y)^\delta, \quad (2.4)$$

respectivamente, na qual  $\delta > 0$ . Além disso, a partir do procedimento de exponenciação reversa correspondente, uma distribuição *Reversa Potência*  $H$  é obtida a partir da fdp e fda dadas por:

$$h_{RP}(y|\delta) = \delta h(-y)H(-y)^{\delta-1} \quad \text{e} \quad H_{RP}(y|\delta) = 1 - H(-y)^\delta, \quad (2.5)$$

respectivamente,  $\delta > 0$ . Note que se  $Y \sim H_P$ , então  $-Y \sim H_{RP}$ , a recíproca também é verdadeira. Além disso, seguindo [Bazán et al. \(2017\)](#), ambas as distribuições satisfazem a propriedade de

reversão, ou seja,  $H_P(y) + H_{RP}(-y) = 1$ . Assim, a distribuição  $H_{RP}$  é chamada de distribuição reversa de  $H_P$ .

Outro procedimento é baseado na construção de distribuições transmutada assimétrica seguindo [Shaw e Buckley \(2009\)](#). Neste caso, se  $H$  é uma distribuição base, então uma nova distribuição transmutada assimétrica  $H$  é definida considerando o procedimento definido por

$$h_{TS}(y|\delta) = h(y) [1 + \delta \{1 - 2H(y)\}] \quad \text{e} \quad H_{TS}(y|\delta) = H(y) [1 + \delta \{1 - H(y)\}], \quad (2.6)$$

sendo a fdp e fda, respectivamente, na qual  $|\delta| < 1$ . Essa formulação é conhecida como mapa de transmutação de postos quadráticos. Considerando o procedimento apresentado acima, introduzem-se as seguintes distribuições:

### ***i) Distribuição generalizada de valores extremos e sua correspondente distribuição reversa***

Diz-se que  $Y \in \mathbb{R}$  segue uma distribuição *Gumbel Generalizada de valores extremos* (GEV-G) e sua fdp e fda puderem ser escritas respectivamente, como

$$g_{GEV-G}(y|\lambda) = (1 + \lambda y)_+^{(-1/\lambda)-1} \exp\left\{- (1 + \lambda y)_+^{-1/\lambda}\right\} \quad \text{e} \quad G_{GEV-G}(y|\lambda) = \exp\left\{- (1 + \lambda y)_+^{-1/\lambda}\right\}, \quad (2.7)$$

na qual,  $x_+ = \max\{0, x\}$  e  $\lambda \in \mathbb{R}$ . Observe-se que o caso particular em que  $\lambda \rightarrow 0$ , ou seja,  $G_{GEV-G}(y|0) = G(y)$ , corresponde à distribuição de Gumbel, pois  $\lim_{\lambda \rightarrow 0} (1 + \lambda y)_+^{-1/\lambda} = \exp(-y)$ .

A v.a  $Y \in \mathbb{R}$  tem uma distribuição *Gumbel Reversa Generalizada de valores extremos* (GEV-RG) se a sua fdp e fda forem dadas respectivamente, por

$$f_{GEV-RG}(y;\lambda) = (1 - \lambda y)_+^{(-1/\lambda)-1} \exp\left\{- (1 - \lambda y)_+^{-1/\lambda}\right\} \quad \text{e} \quad F_{GEV-RG}(y;\lambda) = 1 - \exp\left\{- (1 - \lambda y)_+^{-1/\lambda}\right\}, \quad (2.8)$$

em que,  $x_+ = \max\{0, x\}$  e  $\lambda \in \mathbb{R}$ . Um caso particular é  $F_{GEV-RG}(y|0) = F(y)$ , ou seja, a distribuição RG como  $\lim_{\lambda \rightarrow 0} (1 - \lambda y)_+^{-1/\lambda} = \exp(y)$ . Se  $Y \sim GEV-G$ , então  $-Y \sim GEV-RG$ . Além disso, ambas as distribuições satisfazem a propriedade de reversão:  $G_{GEV-G}(y|\lambda) + F_{GEV-RG}(-y|\lambda) = 1$ .

### ***ii) Potência Reversa Gumbel e sua distribuição reversa correspondente***

Seguindo o procedimento introduzido por [Bazán et al. \(2017\)](#), optou-se por considerar apenas as distribuições RG como base. Essa escolha deu-se pelo fato de que quando se consideram

as distribuições  $G$  como base, a assimetria permaneceu constante, independentemente do valor de  $\lambda$  escolhido (HUAYANAY *et al.*, 2019). Assim, a v.a  $Y \in \mathbb{R}$  segue uma distribuição *Potência Reversa Gumbel* (P-RG) se sua fdp e fpa forem dadas, respectivamente, por

$$f_{P-RG}(y|\delta) = \delta \exp\{y - \exp(y)\} [1 - \exp\{-\exp(y)\}]^{\delta-1} \quad \text{e} \quad F_{P-RG}(y; \delta) = [1 - \exp\{-\exp(y)\}]^{\delta}, \quad (2.9)$$

$\delta > 0$ . Quando  $F_{P-RG}(y|\delta = 1) = F(y)$ , obtem-se a distribuição *RG*.

A v.a  $Y \in \mathbb{R}$  tem uma distribuição *Reversa da Potência Gumbel Reversa* (RP-RG) se a sua fdp e fda puderem ser escritas, respectivamente, como:

$$g_{RP-RG}(y|\delta) = \delta \exp\{-y - \exp(-y)\} [1 - \exp\{-\exp(-y)\}]^{\delta-1} \quad \text{e} \quad G_{RP-RG}(y|\delta) = 1 - [1 - \exp\{-\exp(-y)\}]^{\delta}, \quad (2.10)$$

com  $\delta > 0$ . Como caso particular, a distribuição  $G$  é obtida quando  $G_{RP-RG}(y|\delta = 1) = G(y)$ , ambas as distribuições satisfazem a propriedade de reversão. Além disso, em (2.9) e (2.10) uma forma conveniente de reparametrização é expressa por  $\delta = \exp(\lambda)$ . Essa abordagem foi adotada e resultou em melhorias computacionais observadas (HUAYANAY *et al.*, 2019). Os casos específicos mencionados anteriormente são obtidos quando  $\lambda = 0$ .

### iii) Distribuições Transmutadas Assimétricas Gumbel e Transmutadas Gumbel Reversa

A v.a  $Y \in \mathbb{R}$  segue uma distribuição *Gumbel Transmutada assimétrica* (TS-G) se a fdp e a fda forem dadas por:

$$g_{TS-G}(y|\delta) = \exp\{-y - \exp(-y)\} (1 + \delta [1 - 2 \exp\{-\exp(-y)\}]) \quad \text{e} \quad G_{TS-G}(y|\delta) = \exp\{-\exp(-y)\} (1 + \delta [1 - \exp\{-\exp(-y)\}]), \quad (2.11)$$

respectivamente, onde  $|\delta| < 1$ . O caso particular  $G_{TS-G}(y|\delta = 0) = G(y)$  é a distribuição Gumbel.

A v.a  $Y \in \mathbb{R}$  é distribuído como distribuição *Gumbel Reversa Transmutada assimétrica* (TS-RG) se a fdp e a fda forem dadas, respectivamente, por

$$f_{TS-RG}(y|\delta) = \exp\{y - \exp(y)\} (1 + \delta [1 - 2 \exp\{-\exp(y)\}]) \quad \text{e} \quad F_{TS-RG}(y|\delta) = [1 - \exp\{-\exp(y)\}] (1 + \delta \exp\{-\exp(y)\}), \quad (2.12)$$

onde  $|\delta| < 1$ . Para  $\delta = 0$ ,  $F_{TS-RG}(y|\delta = 0) = F(y)$  é a distribuição reversa de Gumbel.

Além disso, uma reparametrização conveniente para (2.11) e (2.12) é  $\delta = \frac{e^\lambda - 1}{e^\lambda + 1}$ ,  $\lambda \in \mathbb{R}$ , de onde se obtêm os casos particulares destas distribuições para  $\lambda = 0$ . Essa reparametrização foi utilizada por [Shaw e Buckley \(2009\)](#), e mostrou uma melhoria computacional.

A [Figura 1](#) descreve diferentes curvas de fdp que incluem as distribuições Gumbel (G) e Gumbel Reversa (RG) como linha de base. As fdps do GEV-G, GEV-RG, P-RG, RP-RG, TS-G e TS-RG são apresentadas para diferentes valores de  $\lambda \in \mathbb{R}$ . A linha curva sólida, em cada subfigura, corresponde à distribuição de Gumbel e à distribuição Reversa Gumbel, ou seja, as funções de ligação *loglog* e *cloglog*. Em todos os casos considerados, o efeito do parâmetro  $\lambda$  afeta a forma da fdp em relação às fdps G e RG.

## 2.3 Modelos de regressão binomial usando a função de ligação *cloglog* flexíveis

Nesta seção, são formulados novos modelos de regressão binomial usando ligações *cloglog* flexíveis com a fda das distribuições apresentadas na Seção 2.2. Assim, para a função  $F$  em (2.1), a inversa da função de ligação, utiliza-se o fda das distribuições descritas na seção anterior. Usando a notação aí introduzida, o modelo com abordagem bayesiana é dado por

$$Y_i | \beta, \lambda \stackrel{ind.}{\sim} \text{Binomial}(n_i, p_i), \quad p_i = F_\lambda(\eta_i), \forall i = 1, \dots, m, \quad \eta_i = x_i^T \beta \quad (2.13)$$

em que,  $\sum_{i=1}^m n_i = N$  e  $F_\lambda$  pode ser a fda das distribuições GEV-G, GEV-RG, P-RG, RP-RG, TS-G e TS-RG, introduzidas em (2.7) a (2.12), todas indexadas pelo parâmetro extra  $\lambda$  (consulte a [Tabela 1](#) para mais detalhes). Assim, se for utilizada a distribuição GEV-G e se forem consideradas as respostas binárias, obtém-se o modelo de regressão binária proposto por [Wang, Dey et al. \(2010\)](#); se forem utilizadas as distribuições P-RG e RP-RG, obtém-se o modelo de regressão binária proposto por [Bazán et al. \(2017\)](#). Entretanto, se forem consideradas as distribuições GEV-RG, TS-G e TS-RG, novos modelos de regressão binomial são obtidos. Além disso, de qualquer uma das distribuições flexíveis do Reverse Gumbel consideradas na seção anterior, tem-se que a função de ligação *cloglog* torna-se um caso particular das funções de ligação correspondentes a essas distribuições, quando  $\lambda = 0$ . De maneira semelhante, a função de ligação *loglog* torna-se um caso especial das distribuições Gumbel flexíveis, que é obtido quando considerado  $\lambda = 0$ .

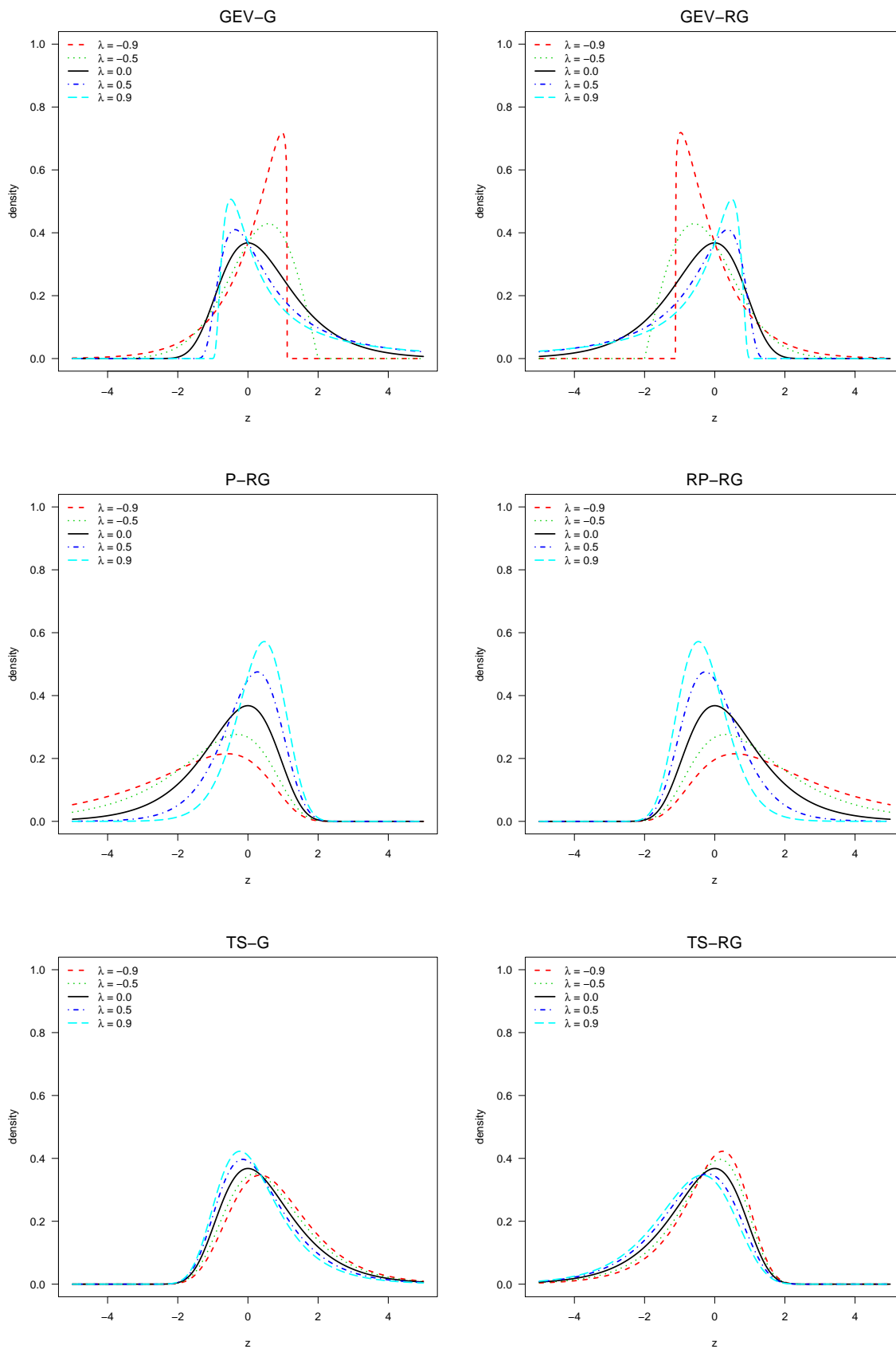


Figura 1 – Funções de densidade de probabilidade para as distribuições GEV-G, GEV-RG, PR-G, PR-RG, TS-G e TS-RG considerando diferentes valores de  $\lambda \in \mathbb{R}$

Fonte: Elaborada pelo autor.



Tabela 1 – Modelos binomiais flexíveis obtidos a partir de generalizações das distribuições G e RG

Funções de ligação	$p = F_\lambda(\eta)$	Espaço paramétrico
GEV-G	$p = \exp\left\{-\left(1 + \lambda\eta\right)_+^{-1/\lambda}\right\}$	$\lambda \in \mathbb{R}$
GEV-RG	$p = 1 - \exp\left\{-\left(1 - \lambda\eta\right)_+^{-1/\lambda}\right\}$	$\lambda \in \mathbb{R}$
P-RG	$p = [1 - \exp\{-\exp(\eta)\}]^\delta$	$\delta = \exp(\lambda)$ , $\lambda \in \mathbb{R}$ e $\delta > 0$
RP-RG	$p = 1 - [1 - \exp\{-\exp(-\eta)\}]^\delta$	$\delta = \exp(\lambda)$ , $\lambda \in \mathbb{R}$ e $\delta > 0$
TS-G	$p = \exp\{-\exp(-\eta)\} (1 + \delta [1 - \exp\{-\exp(-\eta)\}])$	$\delta = \frac{\exp(\lambda) - 1}{\exp(\lambda) + 1}$ , $\lambda \in \mathbb{R}$ e $-1 < \delta < 1$
TS-RG	$p = [1 - \exp\{-\exp(\eta)\}] [1 + \delta \exp\{-\exp(\eta)\}]$	$\delta = \frac{\exp(\lambda) - 1}{\exp(\lambda) + 1}$ , $\lambda \in \mathbb{R}$ and $-1 < \delta < 1$

Fonte: Elaborada pelo autor.

A [Figura 2](#) mostra diferentes curvas fda que incluem as distribuições Gumbel e Reversa Gumbel como linha de base. São mostradas as fda das distribuições GEV-G, GEV-RG, P-RG, RP-RG, TS-G e TS-RG para diferentes valores de  $\lambda$ . A curva sólida corresponde à distribuição G (*loglog*) e distribuição RG (*cloglog*). Observe-se que, em todos esses casos, o efeito do parâmetro de forma  $\lambda$  altera a probabilidade de sucesso em relação à função de ligação *loglog* e a função de ligação *cloglog*.

## 2.4 Estimação Bayesiana

Nesta seção, considera-se o modelo de regressão binomial flexível definido em (2.13), onde para cada  $i = 1, \dots, m$  a probabilidade de sucesso  $p_i = F_\lambda(\eta_i)$ , é substituída pela probabilidade flexível de sucesso dada pela fda de distribuições na [Tabela 1](#).

Para modelos *cloglog* flexíveis, a função de verossimilhança para o parâmetro de regressão  $\beta$  e o parâmetro de forma extra  $\lambda$  é dada por

$$L(\beta, \lambda | y, X) = \prod_{i=1}^m \binom{n_i}{y_i} \left\{ F_\lambda(x_i^\top \beta) \right\}^{y_i} \left\{ 1 - F_\lambda(x_i^\top \beta) \right\}^{n_i - y_i}, \quad (2.14)$$

onde  $y = (y_1, \dots, y_m)^\top$  é um vetor de respostas binomiais e  $X = (x_1, \dots, x_m)^\top$  é uma matriz de covariáveis de ordem  $m \times k$ .

### 2.4.1 Especificação da distribuição à priori

Além da função de verossimilhança, a análise bayesiana requer a especificação de uma distribuição à priori,  $\pi(\beta, \lambda)$ , para  $\beta$  e  $\lambda$ . Uma situação comum é considerar esses parâmetros independentes, de modo que

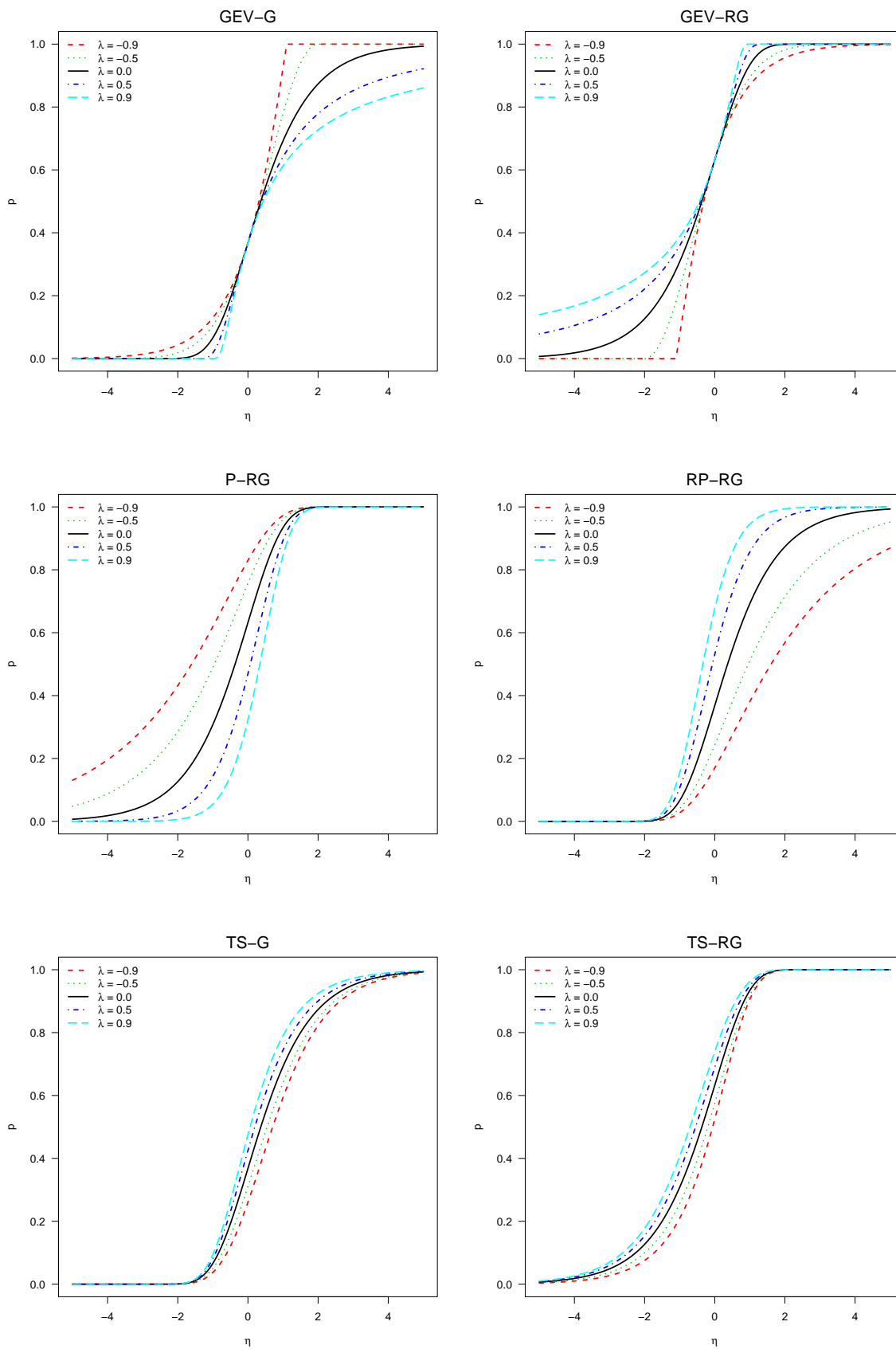


Figura 2 – Funções de distribuição acumuladas para distribuições GEV-G, GEV-RG, P-RG, RP-RG, TS-G e TS-RG considerando diferentes valores de  $\lambda \in \mathbb{R}$

Fonte: Elaborada pelo autor.

$$\pi(\beta, \lambda) = \pi_1(\beta)\pi_2(\lambda). \quad (2.15)$$

Para o vetor de parâmetros de regressão  $\beta = (\beta_1, \dots, \beta_k)^\top$ , podem ser usadas as distribuições *à priori* típicas consideradas no modelo *probit* ou *cloglog*, por exemplo, considerando que seus componentes  $\beta_1, \dots, \beta_k$  são independentes e normalmente distribuídos, com  $\beta_j \sim N(\mu_{0j}, \sigma_{0j}^2)$ . Para o parâmetro de forma  $\lambda$ , pode-se considerar uma distribuição *à priori* uniforme  $\lambda \sim U(-2, 2)$ , que foi encontrada na simulação do estudo de sensibilidade *à priori* da Seção 2.5.1 como sendo mais conveniente para os modelos propostos. Além disso, a priori  $\lambda \sim U(-2, 2)$  já havia sido considerada anteriormente por [Bazán et al. \(2014\)](#) e os resultados satisfatórios em outros estudos de simulação e aplicações reais. Considera-se que esta especificação define uma escala para este parâmetro que é conveniente para o modelo proposto.

Assumindo a função de verossimilhança binomial em (2.14) e uma distribuição a priori como em (2.15) é possível implementar um procedimento de estimação Bayesiana a partir das distribuições a posteriori dadas por

$$\pi(\beta, \lambda | y, X) \propto \prod_{i=1}^m \left\{ F_\lambda(x_i^\top \beta) \right\}^{y_i} \left\{ 1 - F_\lambda(x_i^\top \beta) \right\}^{n_i - y_i} \pi_1(\beta)\pi_2(\lambda), \quad (2.16)$$

onde  $F_\lambda$  é a fda de uma distribuição *cloglog* flexível.

Tal abordagem é complicada, pois as integrais envolvidas na obtenção das distribuições marginais a posteriori não são triviais. No entanto, a distribuição *à posteriori* em (2.16) permite a implementação de métodos Monte Carlo via Cadeias de Markov, que simplificam a amostragem eficiente das distribuições marginais *à posteriori* usando diferentes algoritmos MCMC.

Neste estudo são considerados dois algoritmos: o algoritmo No-U-Turn Sampler (NUTS) e Random Walk Metropolis (RWM). O algoritmo NUTS foi desenvolvido por [Hoffman e Gelman \(2014\)](#) e é uma extensão do algoritmo MCMC, chamado Hamiltonian Monte Carlo (HMC). Este algoritmo permite que a cadeia de Markov explore a distribuição objetiva com muito mais eficiência do que outros métodos MCMC, amplamente conhecidos como amostragem de Metropolis Hastings e Gibbs, usando a dinâmica hamiltoniana ao invés de uma distribuição de probabilidade, permitindo que as cadeias de Markov de saída converjam rapidamente ([NEAL et al., 2011](#)). O algoritmo NUTS pode ser executado sem qualquer ajuste manual e as amostras geradas são tão boas quanto as do HMC ([NEAL et al., 2011](#)). A revisão do método HMC e o desempenho do algoritmo NUTS podem ser vistos no trabalho de [Anyosa \(2017\)](#). Este algoritmo está implementado no software Stan.

O esquema de atualização do algoritmo RWM foi introduzido por [Metropolis et al. \(1953\)](#). Este algoritmo é um esquema Monte Carlo de cadeia de Markov simples e pode ser usado para obter amostras aleatórias de qualquer distribuição alvo arbitrariamente complicada de qualquer dimensão conhecida até uma constante de normalização. O algoritmo RWM define

uma cadeia de variáveis aleatórias cuja distribuição converge para a distribuição desejada e, a partir de um determinado ponto, torna a cadeia amostral uma amostra da distribuição de interesse (GELMAN; GILKS; ROBERTS, 1997). O procedimento MCMC no SAS usa o algoritmo RWM para gerar uma sequência de sorteios a partir da distribuição posterior conjunta de parâmetros.

Todos os modelos aqui considerados foram implementados usando `proc MCMC` do software SAS 9.4 (SAS Institute Inc., 2009) e Stan através do software RStudio versão 4.2.1, usando o pacote `rstan` (Stan Development Team, 2020).

### 2.4.2 Critérios de comparação de modelos e análise de resíduos

Existem diversas abordagens metodológicas para a comparação de modelos alternativos dentro de uma estrutura bayesiana. Neste estudo, são considerados diferentes critérios, tais como o Critério de Informação de Deviance (DIC), proposto por Gelman, Hwang e Vehtari (2014), o Critério de Informação Esperada de Akaike (EAIC), os Critérios de Informação de Schwarz e Bayesianos (EBIC), discutidos em Bazán *et al.* (2014) e Brooks (2002), o Critério de Informação (IC), bem como o Critério de Informação de Hannan-Quinn (HQIC), conforme abordado por Anderson e Burnham (2004).

Todos esses critérios são baseados na média posterior do desvio, dadas por  $DIC = \bar{D} + p_D$ ,  $EAIC = \bar{D} + 2p$ ,  $EBIC = \bar{D} + p \times \log(N)$ ,  $IC = \bar{D} + 2p_D$  e  $HQIC = \bar{D} + 2p \log(\log(N))$ , em que  $\bar{D}$  é a média posteriori do desvio do modelo,  $p_D$  é o número efetivo de parâmetros no modelo,  $N$  é o número total de observações e  $p$  é o número de parâmetros no modelo.

Além disso, usa-se o critério de informação Watanabe-Akaike (WAIC), definido como  $WAIC = -2l_{ppd} + 2p_D$ , onde  $l_{ppd}$  é o logaritmo da densidade preditiva pontual e  $p_D$  é o número efetivo de parâmetros (WATANABE; OPPER, 2010). Também considera-se outra medida chamada ‘leave-one-out cross-validation’ (LOO), proposta por Geisser e Eddy (1979).

Além dos critérios mencionados acima, também utiliza-se a análise de resíduos para decidir pelo o melhor modelo, para tal, considera-se o resíduo quantil randomizado normalizado, proposto por Dunn e Smyth (1996). De acordo com Rigby e Stasinopoulos (2005), quando as suposições do modelo são satisfeitas, então esses resíduos devem ser normalmente distribuídos em caso de adequação da previsão ou ajuste. A vantagem de usar esses resíduos é que eles têm uma distribuição conhecida, independente da distribuição da variável resposta, e facilmente verificável por meio de testes de hipóteses e gráficos de probabilidade normal (QQ-plot) (SILVA; ANYOSA; BAZAN, 2020).

Para todos os critérios acima, valores menores indicam melhor ajuste do modelo aos dados observados.

## 2.5 Estudos de simulação

### 2.5.1 Análise de Sensibilidade da distribuição à priori para o parâmetro $\lambda$

Uma análise de sensibilidade foi realizada para explorar o uso de três distribuições à priori diferentes para o parâmetro  $\lambda$ , considerando os modelos propostos por meio deste estudo: GEV-G, GEV-RG, P-RG, RP-RG, TS-G e TS -RG. As prioris estudadas foram:

- Prior 1:  $\lambda \sim U(-2, 2)$ , proposto por [Bazán et al. \(2014\)](#);
- Prior 2:  $\lambda \sim N(0, 1, 5)$ , proposto por [Wang, Dey et al. \(2010\)](#);
- Prior 3:  $\lambda \sim Ca(0, 2, 5)$ , proposto por [Gelman et al. \(2008\)](#).

Para isso, geram-se dados desbalanceados de cada distribuição proposta com coeficientes de regressão  $\beta = (\beta_0, \beta_1, \beta_2)^T$ , associados a um vetor de covariáveis  $\mathbf{x}_i = (x_{0i}, x_{1i}, x_{2i})^T$  e um parâmetro de forma  $\lambda$  ou  $\delta$ , conforme descrito na Tabela 1.

As covariáveis foram geradas usando  $x_{1i} \sim N(0, 1)$  e  $x_{2i} \sim N(0, 1)$ ,  $i = 1, \dots, 100$ , e os valores considerados para os coeficientes de regressão foram  $(\beta_0, \beta_1, \beta_2, \lambda) = (-1, 114, -0, 5, 0, 009, -1, 0)$ . Estes valores foram obtidos ajustando a regressão *cloglog* aos dados utilizados no trabalho de [Devidas, George e Zelterman \(1993\)](#). Assim, foram gerados  $Y_i \stackrel{ind.}{\sim} Binomial(n_i, p_i)$ , com  $n_i = 15$ ,  $p_i = F_\lambda(\eta_i)$  e  $\eta_i = x_i^\top \beta$ , para  $i = 1, \dots, 100$  ( $m = 100$ ).

Neste estudo, considera-se um total de  $R = 100$  réplicas, e para cada conjunto de dados gerados, ajusta-se seu modelo correspondente considerando as três distribuições à priori mencionadas. Usa-se o procedimento de estimação descrito na Seção 2.4 e o pacote `rstan` para encontrar as estimativas com as seguintes especificações: 40.000 iterações executadas com 20.000 iterações de descarte, usando um intervalo de espaçamento de 10 iterações e três cadeias. A raiz do erro quadrático médio (RMSE) foi usada para comparar essas prioris a qual é dada por:  $RMSE(\hat{\theta}_j) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_j^{(r)} - \theta_j)^2}$ , onde  $R$  é o número de réplicas na simulação,  $\hat{\theta}_j^{(r)}$  é a média posterior do parâmetro  $\theta_j$  na  $r$ -ésima réplica, e  $\theta_j$  é o  $j$ -ésimo componente de  $\theta = (\beta^\top, \lambda)^\top$ . Para cada réplica, considera-se o DIC, apresentado na Seção 2.4.2 e o tempo médio de execução do algoritmo em cada réplica.

A Tabela 2 informa os valores de RMSE e a estatística  $\hat{R}$  para cada um dos parâmetros, junto com o critério de comparação do modelo (DIC) e o tempo para executar os algoritmos MCMC (Tempo).

Constata-se que as três distribuições à priori se comportam de maneira semelhante para os modelos GEV-G e GEV-RG, ou seja, não há sensibilidade das distribuições à priori aos parâmetros quando considerado o RMSE e o critério de comparação DIC. Porém, para os

Tabela 2 – Resultados da análise de sensibilidade dos modelos GEV-G, GEV-RG, P-RG, RP-RG, TS-G e TS-RG para as três distribuições *à Priori*

		$\beta_0$		$\beta_1$		$\beta_2$		$\lambda$		DIC	Tempo
		RMSE	$\hat{R}$	RMSE	$\hat{R}$	RMSE	$\hat{R}$	RMSE	$\hat{R}$		
GEV-G	Priori 1	0,033	1,003	0,042	1,013	0,030	1,003	0,178	1,016	389,905	139,662
	Priori 2	0,033	1,002	0,042	1,011	0,030	1,002	0,184	1,015	389,945	118,429
	Priori 3	0,033	1,003	0,042	1,014	0,030	1,002	0,181	1,016	390,001	129,378
GEV-RG	Priori 1	0,034	1,000	0,053	1,000	0,027	1,000	0,122	1,000	381,092	83,640
	Priori 2	0,033	1,000	0,053	1,000	0,027	1,000	0,121	1,000	381,112	81,928
	Priori 3	0,034	1,000	0,053	1,000	0,027	1,000	0,122	1,000	381,119	84,987
P-RG	Priori 1	0,490	1,004	0,206	1,003	0,074	1,000	0,397	1,006	350,036	97,160
	Priori 2	7,517	1,026	3,142	1,025	0,392	1,006	2,254	1,059	351,012	101,397
	Priori 3	11,055	1,021	4,578	1,020	0,559	1,006	2,803	1,041	351,794	84,391
RP-RG	Priori 1	0,380	1,002	0,177	1,002	0,071	1,001	0,371	1,002	311,224	135,915
	Priori 2	5,940	1,031	2,585	1,029	0,379	1,009	2,160	1,078	312,206	156,593
	Priori 3	9,907	1,042	4,261	1,040	0,593	1,010	2,878	1,086	313,261	137,430
TS-G	Priori 1	0,170	1,000	0,054	1,000	0,036	1,000	0,646	1,000	352,909	82,203
	Priori 2	0,158	1,000	0,056	1,000	0,035	1,000	0,583	1,000	352,898	96,003
	Priori 3	0,248	2,129	0,073	1,242	0,035	1,061	25,202	1,163	352,777	424,390
TS-RG	Priori 1	0,418	1,001	0,089	1,000	0,040	1,000	1,301	1,001	403,653	119,968
	Priori 2	0,433	1,000	0,088	1,000	0,039	1,000	1,394	1,000	403,639	146,415
	Priori 3	0,700	2,128	0,086	1,348	0,039	1,055	40,881	1,773	403,486	437,947

Fonte: Elaborada pelo autor.

modelos P-RG e RP-RG, priori 1(Uniforme) apresenta a melhor recuperação entre todos os parâmetros considerados nos modelos, segundo o RMSE e DIC. Por fim, para os modelos TS-G e TS-RG, temos um RMSE pequeno para as distribuições *à priori* 1 e 2 (Uniforme e Normal) e DIC comportaram-se de maneira semelhante, mas descarta-se o uso da terceira distribuição *à priori* (Cauchy) porque o RMSE é maior para o parâmetro  $\lambda$ . Além disso, na estatística  $\hat{R}$  pôde-se observar a convergência das cadeias MCMC quando se considera as distribuições *à priori* 1 e 2 (Uniforme e Normal), porém com a priori 3 vê-se que as cadeias MCMC para TS-G e TS-RG não convergiram. Com relação ao tempo computacional, observa-se uma grande diferença para os modelos TS-G e TS-RG considerando o pior caso a priori 3 (Cauchy) e, com relação aos demais modelos essa diferença foi sutil. Nota-se que a priori 1 (Uniforme) funciona bem para todos os modelos. Desta forma, para dar continuidade aos estudos, opta-se por utilizar a distribuição *à priori* Uniforme para o parâmetro  $\lambda$ .

## 2.5.2 Estudo de Recuperação de Parâmetros

Um estudo de simulação foi desenvolvido para avaliar a recuperação de parâmetros para cada generalização das funções de ligação *cloglog* e *loglog* propostas neste trabalho. Dados desbalanceados são gerados a partir de cada distribuição proposta com coeficientes de regressão  $\beta = (\beta_0, \beta_1, \beta_2)^T$  associados a um vetor de covariáveis  $\mathbf{x}_i = (x_{0i}, x_{1i}, x_{2i})^T$  e um parâmetro de forma  $\lambda$  ou  $\delta$ , conforme descrito na Tabela 1.

O cenário considerado foi o mesmo do estudo de simulação de sensibilidade da distribuição *à priori* descrito acima. Em resumo, as covariáveis foram geradas usando  $x_{1i}, x_{2i} \sim N(0, 1)$

para  $i = 1, \dots, 100$ . Os valores considerados para os parâmetros nos coeficientes de regressão são especificados nas Tabelas 3 e 4, escolhidas por conveniência para obter dados desbalanceados de forma oposta. Para o parâmetro de forma, utiliza-se  $\lambda = 0,3$  e  $\lambda = -0,3$ . Assim, foram geradas v.a  $Y_i \stackrel{ind.}{\sim} \text{Binomial}(n_i, p_i)$ , com  $n_i = 15$ ,  $p_i = F_\lambda(\eta_i)$  e  $\eta_i = x_i^\top \beta$ , para  $i = 1, \dots, 100$  ( $m = 100$ ), considerando a especificação a priori discutida na Seção 2.5.1.

Considerando os dados simulados, há as seguintes proporções de 1 em cada uma das  $n$  amostras: quando  $\lambda = 0,3$  as proporções foram  $p = (0,308, 0,587, 0,546, 0,456, 0,346, 0,610)$ , e para  $\lambda = -0,3$  as proporções foram  $p = (0,325, 0,567, 0,670, 0,329, 0,288, 0,545)$ , para GEV-G, GEV-RG, P-RG, RP-RG, TS-G e modelos TS-RG, respectivamente.

Para os modelos P-RG e RP-RG considera-se o intercepto igual a zero. Esta escolha se deu pelo fato de se pontuar problemas de convergência em  $\lambda$  e  $\beta_0$  ao considerar  $\beta_0$  diferente de 0. Esse tipo de problema também foi relatado em Lee e Sinha (2019) para o modelo assimétrico-probit, e a solução encontrada pelos autores para contornar esse problema foi modelar sem o intercepto.

Para cada conjunto de dados gerado, ajusta-se todos os modelos na Seção 2.2. Para cada amostra simulada, usamos o procedimento de estimação descrito na Seção 2.4. Utiliza-se três cadeias MCMC, 40.000 iterações são executadas com 20.000 iterações de descarte, usando um intervalo de espaçamento de 10 iterações. Foram consideradas 100 repetições desse procedimento.

A raiz quadrada média do erro (RMSE) foi usada para comparar os métodos,  $RMSE(\hat{\theta}_j) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_j^{(r)} - \theta_j)^2}$ , onde  $R$  é o número de réplicas na simulação,  $\hat{\theta}_j^{(r)}$  é a média posterior do parâmetro  $\theta_j$  na  $r$ ésima réplica, e  $\theta_j$  é o  $j$ ésimo componente de  $\theta = (\beta^\top, \lambda)^\top$ . Também foi calculada a probabilidade de cobertura de 95% para os intervalos de credibilidade das cadeias MCMC; pode-se pensar neste valor como a probabilidade de que o intervalo de credibilidade contenha o valor do verdadeiro parâmetro em 95% dos casos (Rubin *et al.* (2017); Trikalinos, Hoaglin e Schmid (2013)). Outra medida utilizada para avaliar a recuperação dos parâmetros foi o tempo computacional, dado em segundos, que foi incorporado por meio da função `system.time` do programa R.

Os códigos para todos os modelos foram desenvolvidos e implementados no Stan através do RStudio usando o pacote `rstan`. Os resultados da simulação para  $\lambda = 0,3$  e  $\lambda = -0,3$  são mostrados na Tabela 3 e Tabela 4, respectivamente.

Pode-se observar, na Tabela 3, que quando  $\lambda = 0,3$  todos os modelos possuem: a) uma boa recuperação dos parâmetros; b) valores pequenos de RMSE próximos de zero; c) uma probabilidade de cobertura em torno de 1 para cada parâmetro estimado indicando que todos estão dentro do erro de simulação de 95%, confirmando a precisão dos cálculos numéricos e d) a estatística  $\hat{R}$  próxima de 1 em todos modelos, o que permite concluir que as cadeias MCMC estão convergindo. No entanto, observa-se que o RMSE para o parâmetro  $\lambda$  para o modelo TS-G

Tabela 3 – Estimativas de parâmetros para os modelos propostos com dados desbalanceados simulados para cada modelo com  $\lambda = 0,3$  usando o pacote rstan

Função de ligação	Medidas	$\beta_0$	$\beta_1$	$\beta_2$	$\lambda$	Tempo médio	prop média	SD prop
GEV-G	Parâmetro	-0,114	-0,500	0,009	0,300	39,865	0,308	0,154
	Estimativa	-0,125	-0,509	0,009	0,244			
	RMSE	0,041	0,051	0,030	0,289			
	Cobertura	0,960	0,970	0,980	0,910			
	$\hat{R}$	1,000	1,000	1,000	1,000			
GEV-RG	Parâmetro	-0,114	-0,500	0,009	0,300	36,784	0,587	0,149
	Estimativa	-0,110	-0,499	0,011	0,278			
	RMSE	0,041	0,045	0,036	0,251			
	Cobertura	0,950	0,940	0,940	0,970			
	$\hat{R}$	1,000	1,000	1,000	1,000			
P-RG	Parâmetro	0,000	1,000	1,000	0,300	87,298	0,546	0,341
	Estimativa	-0,111	1,092	1,090	0,247			
	RMSE	0,304	0,198	0,185	0,332			
	Cobertura	0,950	0,940	0,930	0,960			
	$\hat{R}$	1,001	1,001	1,001	1,001			
RP-RG	Parâmetro	0,000	1,000	1,000	0,300	81,274	0,456	0,340
	Estimativa	0,107	1,082	1,081	0,253			
	RMSE	0,297	0,177	0,177	0,330			
	Cobertura	0,980	0,990	1,000	0,980			
	$\hat{R}$	1,001	1,001	1,001	1,001			
TS-G	Parâmetro	-0,114	-0,500	0,009	0,300	84,688	0,346	0,158
	Estimativa	0,086	-0,526	0,010	-0,321			
	RMSE	0,204	0,048	0,032	0,625			
	Cobertura	1,000	1,000	0,990	1,000			
	$\hat{R}$	1,000	1,000	1,000	1,000			
TS-RG	Parâmetro	-0,114	-0,500	0,009	0,300	76,342	0,610	0,150
	Estimativa	-0,128	-0,485	0,010	0,329			
	RMSE	0,045	0,041	0,038	0,064			
	Cobertura	1,000	1,000	0,890	1,000			
	$\hat{R}$	1,000	1,000	1,000	1,000			

Fonte: Elaborada pelo autor.

é um pouco maior em comparação aos outros modelos estudados.

Ao considerar os resultados para  $\lambda = -0,3$  na Tabela 4, chega-se a uma conclusão semelhante à apresentada na Tabela 3, pois: a) apresentam uma boa recuperação dos parâmetros; b) apresentam valores de RMSE pequenos e próximos de zero; c) uma probabilidade de cobertura próxima de 1 para cada parâmetro estimado; d) a estatística  $\hat{R}$  é próxima de 1 em todos os modelos.

Com esses pontos, pode-se concluir que as cadeias MCMC estão convergindo. Observa-se que há um RMSE um pouco maior nas estimativas do parâmetro  $\lambda$  para o modelo TS-RG, mas enquanto o RMSE é um pouco maior em relação à estimativa desse parâmetro, indicando que o parâmetro não foi muito bem recuperado, a probabilidade de cobertura ficou em torno de 100% indicando que as estimativas estão dentro do erro de simulação de 0,95, confirmando a precisão dos cálculos numéricos.



Tabela 4 – Estimativas de parâmetros para os modelos propostos com dados desbalanceados simulados para cada modelo com  $\lambda = -0,3$  usando o pacote rstan

Função de ligação	Medidas	$\beta_0$	$\beta_1$	$\beta_2$	$\lambda$	Tempo médio	prop média	SD prop
GEV-G	Parâmetro	-0,114	-0,500	0,009	-0,300	32,581	0,325	0,148
	Estimativa	-0,119	-0,496	0,008	-0,280			
	RMSE	0,037	0,042	0,032	0,257			
	Cobertura	0,980	0,960	0,980	0,960			
	$\hat{R}$	1,000	1,000	1,000	1,000			
GEV-RG	Parâmetro	-0,114	-0,500	0,009	-0,300	37,668	0,567	0,162
	Estimativa	-0,118	-0,506	0,014	-0,229			
	RMSE	0,043	0,043	0,036	0,260			
	Cobertura	0,930	0,960	0,930	0,960			
	$\hat{R}$	1,000	1,000	1,000	1,000			
P-RG	Parâmetro	0,000	1,000	1,000	-0,300	70,169	0,670	0,279
	Estimativa	-0,114	1,093	1,089	-0,357			
	RMSE	0,301	0,188	0,180	0,326			
	Cobertura	0,970	0,980	0,970	0,970			
	$\hat{R}$	1,001	1,001	1,001	1,001			
RP-RG	Parâmetro	0,000	1,000	1,000	-0,300	76,799	0,329	0,273
	Estimativa	0,094	1,069	1,072	-0,337			
	RMSE	0,267	0,148	0,160	0,274			
	Cobertura	1,000	1,000	1,000	1,000			
	$\hat{R}$	1,001	1,001	1,001	1,001			
TS-G	Parâmetro	-0,114	-0,500	0,009	-0,300	77,263	0,288	0,142
	Estimativa	-0,088	-0,486	0,010	-0,341			
	RMSE	0,050	0,046	0,035	0,075			
	Cobertura	1,000	0,970	0,930	1,000			
	$\hat{R}$	1,000	1,000	1,000	1,000			
TS-RG	Parâmetro	-0,114	-0,500	0,009	-0,300	75,248	0,545	0,157
	Estimativa	-0,318	-0,520	0,012	0,311			
	RMSE	0,208	0,045	0,037	0,614			
	Cobertura	1,000	1,000	0,940	1,000			
	$\hat{R}$	1,000	1,000	1,000	1,000			

Fonte: Elaborada pelo autor.

## 2.6 Aplicações em Dados Médicos

Nesta seção, serão realizadas duas aplicações práticas para demonstrar o desempenho das novas funções de ligação na análise de dados em contextos reais. Uma aplicação será abordada no contexto da regressão binomial, utilizando dados consolidados na literatura, enquanto a outra estará no contexto da regressão binária. O objetivo principal é apresentar e evidenciar o bom desempenho das funções de ligação considerando ambas as abordagens. Ao explorar essas duas situações distintas, buscamos ilustrar a eficácia e a utilidade dessas funções no contexto da análise de dados, destacando sua aplicabilidade e capacidade de oferecer resultados consistentes e confiáveis.

### 2.6.1 Dados garotas de Varsóvia

Os dados foram previamente analisados por [Finney \(1971\)](#), [Morgan \(1992\)](#), [Sprent \(2019\)](#) e [Bura e Gastwirth \(2001\)](#), originados de um estudo realizado com 3.918 pré-adolescentes e adolescentes do sexo feminino na cidade de Varsóvia (Milicer, 1966). As participantes foram classificadas em 25 categorias de idade, onde a idade de cada categoria representa o ponto médio da faixa etária ( $x$ ), e o número de meninas que atingiram a menarca no momento do estudo ( $y$ ) a variável resposta.

Além de ajustar os modelos propostos, também são ajustadas uma regressão binomial com as funções de ligação *logit*, *cloglog* e *loglog* com o propósito de compara-las. Em todos os casos, as distribuições a priores usadas foram  $\beta_j \sim N(0; 100)$ ,  $j = 0, 1$ , e  $\lambda \sim U(-2, 2)$ .

Nesta aplicação, considera-se dois algoritmos de estimação usando os softwares R e SAS. Por meio do pacote `rstan` da linguagem R, foram realizadas 40.000 iterações com 20.000 iterações de descarte, juntamente com um intervalo de espaçamento de 10 iterações para obter convergência usando três cadeias MCMC. Usando o SAS com o `proc MCMC` 4.000 iterações de descarte e 400.000 iterações MCMC com um intervalo de espaçamento de 100 são utilizadas.

Para comparar as estimativas e os critérios de seleção obtidos no ajuste dos modelos, foram consideradas condições semelhantes nos dois softwares, respeitando a diferença entre esses algoritmos, uma vez que o pacote `rstan` usa o algoritmo NUTS e a `proc MCMC` usa o algoritmo RWM.

Inicialmente, foi realizado uma padronização das convvariáveis para melhorar sua convergência. Isso se deve ao fato de que, em alguns casos, a estrutura usual do modelo convergiu apenas considerando um valor inicial específico. Este problema ocorreu tanto com o uso do `rstan` no modelo GEV-RG, quanto no `proc MCMC` com o modelo TS-G. Portanto, considera-se a seguinte correção no preditor linear:  $\eta_i = \beta_0^* + \beta_1(x - \bar{x})$ , onde  $\beta_0^* = \beta_0 + \beta_1\bar{x}$ . Considera-se a verificação de convergência baseada na estatística potencial de redução de escala ( $\hat{R}$ ) de ([GELMAN; RUBIN \*et al.\*, 1992](#)) para todos os modelos. Neste caso, descobre-se que  $\hat{R} \approx 1$  em todos os casos.

Para comparar os modelos, faz-se uso dos diferentes critérios de seleção apresentados na Seção 2.4.2. O WAIC e o LOO foram considerados apenas para o pacote `rstan`. Os resultados obtidos são apresentados na Tabela 5.

Os resultados obtidos em ambos os softwares são comparáveis e próximos, conforme indicado na Tabela 5. Os modelos que se destacaram para ambos os algoritmos são GEV-G, GEV-RG e RP-RG, de acordo com os critérios de comparação na Seção 2.4.2. As estimativas, padrão (St.Dev) e os intervalos de credibilidade 95% obtidos para estes modelos são apresentados nas Tabelas 6 e 7, para NUTS e para RWM, respectivamente.

Foi determinado que todos os parâmetros são significativos devido ao fato de nenhum dos intervalos de credibilidade 95% incluir o valor zero. Os resultados dos algoritmos NUTS

Tabela 5 – Comparação de modelos para dados das meninas de Varsóvia

Função de ligação	NUTS							RWM				
	Dic	Eaic	Ebic	Ic	Hqic	Waic	Loo	Dic	Eaic	Ebic	Ic	Hqic
LOGIT	114,71	116,78	129,33	116,64	121,23	114,84	114,88	114,70	116,73	129,28	116,67	121,18
CLOGLOG	207,81	208,90	221,45	210,72	213,35	214,27	214,34	206,85	208,86	221,41	208,84	213,32
LOGLOG	122,83	124,69	137,23	124,97	129,14	124,00	124,07	122,72	124,71	137,25	124,72	129,16
GEV-G	103,83	107,11	125,93	106,54	113,79	103,00	103,08	104,13	107,12	125,94	107,15	113,80
GEV-RG	104,29	107,44	126,26	107,15	114,12	103,58	103,75	104,34	107,43	126,25	107,26	114,11
P-RG	121,04	124,95	143,77	123,14	131,63	121,63	121,66	121,07	124,96	143,78	123,18	131,64
RP-RG	104,29	107,68	126,50	106,89	114,36	103,68	103,76	103,68	107,58	126,40	105,79	114,26
TS-G	116,16	120,05	138,87	118,26	126,73	116,82	116,94	116,47	119,96	138,78	118,97	126,64
TS-RG	185,88	189,26	208,08	188,49	195,94	190,88	191,04	185,26	188,82	207,65	187,70	195,50

Fonte: Elaborada pelo autor.

Tabela 6 – Estimativas do Parâmetro para os três melhores modelos: GEV-G, GEV-RG e RP-RG usando o algoritmo NUTS nos dados das meninas de Varsóvia

Parâmetro	GEV-G					GEV-RG					RP-RG				
	Media	St.Dev.	2,50%	97,50%	$\hat{R}$	Media	St.Dev.	2,50%	97,50%	$\hat{R}$	Media	St.Dev.	2,50%	97,50%	$\hat{R}$
$\beta_0$	-12,37	0,49	-13,34	-11,45	1,00	-11,48	0,42	-12,33	-10,66	1,00	-8,03	0,97	-10,21	-6,47	1,00
$\beta_1$	0,99	0,04	0,91	1,06	1,00	0,86	0,03	0,80	0,92	1,00	0,58	0,09	0,44	0,78	1,00
$\lambda$	-0,16	0,04	-0,23	-0,09	1,00	-0,39	0,03	-0,45	-0,32	1,00	1,23	0,33	0,58	1,89	1,00

Fonte: Elaborada pelo autor.

Tabela 7 – Estimativas do Parâmetro para os três melhores modelos: GEV-G, GEV-RG e RP-RG usando o algoritmo RWM nos dados das meninas de Varsóvia

Parâmetro	GEV-G				GEV-RG				RP-RG			
	Media	St.Dev.	2,50%	97,50%	Media	St.Dev.	2,50%	97,50%	Media	St.Dev.	2,50%	97,50%
$\beta_0$	-12,38	0,50	-13,40	-11,42	-11,47	0,43	-12,35	-10,69	-8,07	0,91	-9,84	-6,43
$\beta_1$	0,99	0,04	0,91	1,06	0,86	0,03	0,80	0,92	0,59	0,08	0,43	0,75
$\lambda$	-0,16	0,04	-0,23	-0,09	-0,39	0,03	-0,45	-0,32	3,53	1,11	1,62	5,85

Fonte: Elaborada pelo autor.

(rstan) e RWM (proc MCMC) são semelhantes e próximos, tanto nas estimativas dos parâmetros quanto no ranking dos melhores modelos nos critérios de comparação.

Para verificar qual seria o melhor modelo entre GEV-G, GEV-RG e RP-RG, considera-se uma análise diagnóstica proposta por [Dunn e Smyth \(1996\)](#), através da análise de resíduos quantílicos randomizados normalizados com envelopes gerados a partir de cada modelo. A figura 3 mostram os gráficos de resíduos de quantis aleatórios normalizados para os modelos de função de ligação GEV-G, GEV-RG e RP-RG.

Observa-se que os pontos (resíduos) estão dispersos aleatoriamente entre os limites do envelope, o que permite verificar a adequação dos modelos de regressão binomial ajustados usando as funções de ligação GEV-G e RP-RG.

### 2.6.2 Dados de Infecção Pulmonar

Um conjunto de dados de infecções pulmonares também foi usado para ilustrar as abordagens bayesianas desenvolvidas neste capítulo para as respostas binárias. Utiliza-se dados apresentados por [Paula e Tuder \(1986\)](#), que correspondem a um estudo caso-controle desenvol-

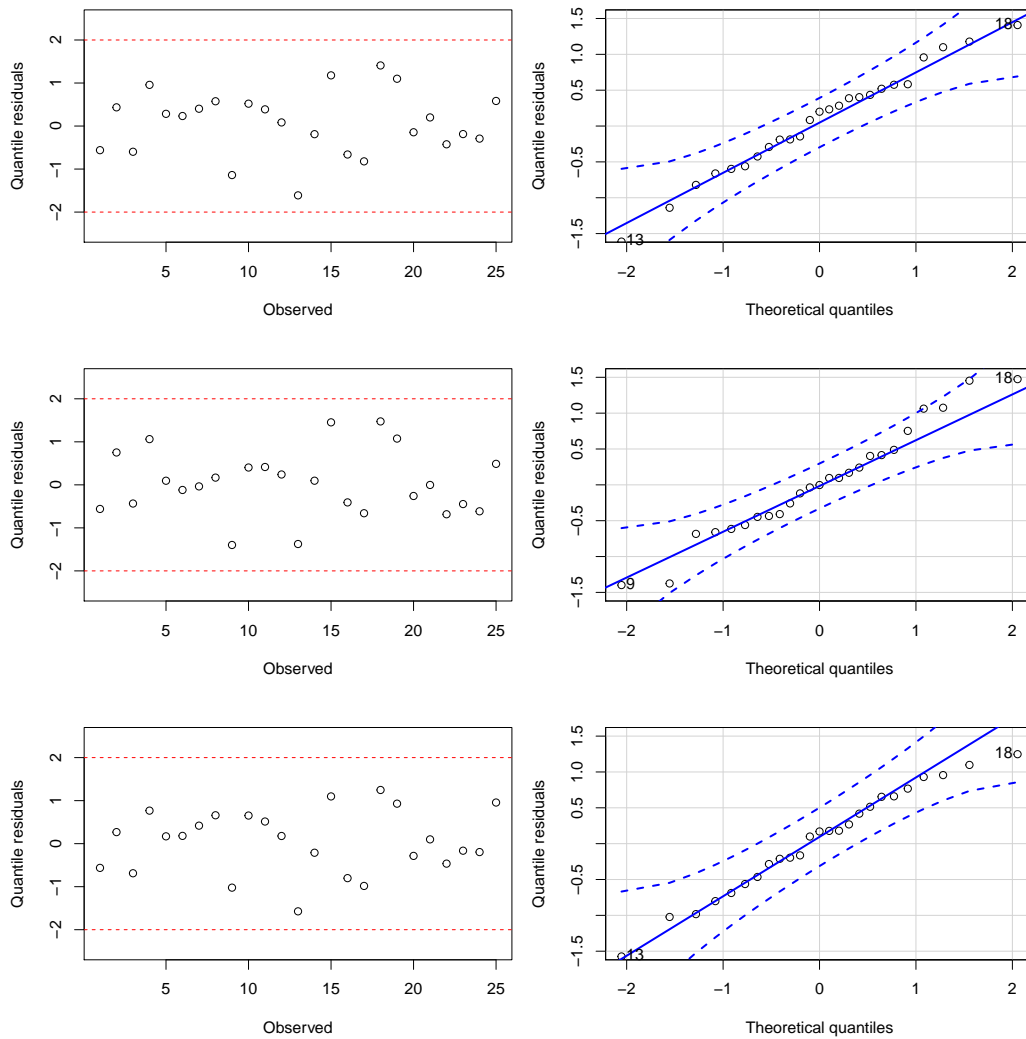


Figura 3 – Resíduos de quantis randomizados normalizados para regressão binomial com as funções de ligação GEV-G, GEV -RG e RP-RG para dados de Warsaw Girls

Fonte: Elaborada pelo autor.

vido no Hospital Heliópolis, em São Paulo - Brasil, de 1970 a 1982. Um total de 175 pacientes com processos infecciosos pulmonares atendidos no hospital, no período anteriormente mencionado, foram classificados de acordo com as seguintes variáveis: Y, tipo de tumor (1: maligno, 0: benigno); IDADE, idade em anos; SEXO (1: masculino, 0: feminino); HL, intensidade de linfócitos celulares histiocíticos (1: ausente, 2: baixo, 3: médio, 4: alto) e SF, intensidade celular fibrose frouxa (1: ausente, 2: baixo, 3: médio, 4: alto). A tabela completa de dados pode ser encontrada em Paula (1995). Nesse caso, a proporção de indivíduos com tumor maligno é 0,4057.

Assim, considera-se  $Y_i \stackrel{ind.}{\sim} Bernoulli(p_i)$  sendo  $p_i = F_\lambda(\eta)$ ,  $i = 1, \dots, 175$ , e  $\eta = \beta_0 + \beta_1 * IDADE1 + \beta_2 * SEX + \sum_{l=1}^4 \beta_{3l} * HL_l + \sum_{l=1}^4 \beta_{4l} * SF_l$ , com HL e SF sendo variáveis binárias correspondentes aos níveis de HL e SF, respectivamente. Assume-se  $\beta_{31} = \beta_{41} = 0$ , como em Paula (1995), assim como  $IDADE1 = IDADE - media(IDADE)$ . Finalmente,  $F_\lambda(\cdot)$  denota o fda de cada uma das distribuições introduzidas em (2.7) a (2.12).

Para esta aplicação, utiliza-se o software SAS. Assim como na aplicação anterior, fez-se uso da `proc MCMC` e consideram-se 4.000 iterações de descarte e 400.000 iterações MCMC com intervalo de espaçamento de 100. Também ajusta-se uma regressão binária com as funções de ligação *logit*, *cloglog* e *loglog* para comparar com os modelos propostos. Em todos os casos, as distribuições a priori  $\beta_j \sim N(0; 100)$ ,  $j = 0, 1, 2$ ;  $\beta_{ik} \sim N(0; 100)$ ,  $i = 3, 4$  e  $k = 1, \dots, 4$  e  $\lambda \sim U(-2, 2)$  são usadas.

Os resultados obtidos pelo cálculo dos diferentes critérios de seleção apresentados na Seção 2.4.2 são apresentados na Tabela 8.

Tabela 8 – Comparação de modelos para dados de Infecção Pulmonar

Função de ligação	Dic	Eaic	Ebic	Ic	Hqic
LOGIT	175,65	184,67	213,15	184,63	196,22
CLOGLOG	181,66	190,89	219,38	190,43	202,45
LOGLOG	170,03	179,27	207,75	178,78	190,82
GEV-G	167,90	179,56	211,21	176,25	192,40
GEV-RG	167,28	178,59	210,24	175,98	191,43
P-RG	186,49	197,07	225,55	193,91	208,62
<b>RP-RG</b>	<b>165,84</b>	<b>178,01</b>	<b>209,66</b>	<b>173,68</b>	<b>190,85</b>
TS-G	170,58	181,60	213,25	179,56	194,44
TS-RG	181,46	192,30	223,95	190,61	205,14

Fonte: Elaborada pelo autor.

A tabela 8 mostra que o melhor modelo para os dados de infecção pulmonar foi o RP-RG, de acordo com os critérios de comparação da Seção 2.4.2. Além disso, os modelos GEV-G, GEV-RG e TS-G apresentaram melhores resultados em comparação ao modelo proposto por Paula (1995). A Tabela 9 apresenta as estimativas de parâmetros, erros padrão assintóticos (SE) e ICs assintóticos de 95% para o modelo RP-RG.

Tabela 9 – Estimativas de parâmetros para o modelo de ligação RP-RG para dados de infecção pulmonar.

Variável	Parâmetro	Média	St.Dev.	2,50%	97,50%
Intercepto	$\beta_0$	12,50	4,16	5,50	21,11
Age	$\beta_1$	0,20	0,04	0,12	0,27
Sex	$\beta_2$	-2,34	1,30	-4,96	0,19
HL(2)	$\beta_{32}$	-2,68	3,96	-10,68	4,43
HL(3)	$\beta_{33}$	-7,33	3,99	-15,25	-0,23
HL(4)	$\beta_{34}$	-9,47	4,46	-18,96	-1,70
SF(2)	$\beta_{42}$	-2,30	1,50	-5,30	0,60
SF(3)	$\beta_{43}$	-2,75	1,54	-5,67	0,37
SF(4)	$\beta_{44}$	1,74	4,99	-6,26	10,99
Parâmetro de forma	$\lambda$	-1,87	0,13	-2,00	-1,62

De acordo com os intervalos de credibilidade, algumas variáveis como sexo e intensidade celular de fibrose (SF), não foram significativas. Portanto, um modelo reduzido poderia ser proposto, embora a intenção deste trabalho seja apenas ilustrar que os modelos propostos também

funcionam com regressão binária. Adicionalmente, encontramos o parâmetro de forma  $\lambda = -1,87$ , o que significa, de acordo com a Figura, que este é um valor consistente com a proporção de 1's na variável independente (visto que  $\lambda < 1$  e inferior a 0,5), o que conseqüentemente explica o desequilíbrio observado nos dados.

## 2.7 Comentários Finais

Neste capítulo, propomos funções de ligação flexíveis baseadas na função de ligação *cloglog* como uma alternativa para modelar dados de resposta binomial na presença de dados médicos desbalanceados. Um procedimento de estimação bayesiana foi desenvolvido para modelos de regressão binomial. Neste estudo, observa-se a importância do uso das funções de ligação flexíveis quanto à assimetria dos dados, que é controlada por meio de um parâmetro de forma extra ( $\lambda$ ).

Descobrimos que funções de ligação comuns em regressão binomial, como *cloglog*, nem sempre fornecem o melhor ajuste na presença de dados desbalanceados. Especificamente, o estudo da análise de sensibilidade a priori (Seção 2.5.1), mostrou que utilizar distribuição Uniforme como priori é o mais conveniente para todos os modelos propostos. Observa-se também que para valores de  $\lambda$  mais altos não se obteve uma boa recuperação para esse parâmetro. Além disso, o estudo de simulação na Seção 2.5.2 mostrou que as funções de ligação propostas recuperam bem os parâmetros e fornecem um bom ajuste na presença de dados desbalanceados.

Na aplicação (Seção 2.6.1) referente aos dados das meninas de Varsóvia, a estimação usando a correção é mais recomendável para a convergência dos modelos estudados. Os resultados, tanto no algoritmo RWM (`proc MCMC`) quanto no algoritmo NUTS (`rstan`) são semelhantes e próximos nas estimativas de parâmetros, bem como no ranking dos melhores modelos usando critérios de comparação.

Na segunda aplicação (Seção 2.6.2) aos dados de infecção pulmonar, a maioria dos modelos de regressão binária propostos com generalizações das funções de ligação *cloglog* obtiveram melhores ajustes em comparação os modelos de regressão binária usando funções de ligação comuns, conforme proposto em Paula (1995). Com esta aplicação, as generalizações das funções de ligação *cloglog* propostas neste artigo vão além da regressão binomial flexível, trazendo também bons resultados para regressões binárias.

É possível considerar algumas extensões dos métodos desenvolvidos neste artigo em pesquisas futuras, por exemplo, uma extensão para modelos mistos pode ser desenvolvida, ou mesmo modelos de teoria de resposta ao item. Além disso, estudos adicionais podem ser sugeridos como, por exemplo, estudos de simetria e curtose das distribuições propostas neste trabalho.

---

# NOVOS MODELOS FLEXÍVEIS DE TRI PARA RESPOSTA DICOTÔMICA

---

Neste capítulo, expande-se a aplicação de certos modelos abordados no Capítulo 1, para o âmbito da Teoria de Resposta ao Item. Além disso, abordaremos a estimação Bayesiana desses novos modelos no contexto da TRI e são ilustradas as vantagens das novas Curvas Características do Item.

## 3.1 A Teoria de Resposta ao item

A Teoria de Resposta ao Item (TRI) preocupa-se em modelar a relação entre a probabilidade de um indivíduo selecionar uma certa resposta de um determinado item e os traços latentes do indivíduo que são características do indivíduo que não podem ser diretamente observadas ou medidas (HAMBLETON; SWAMINATHAN; ROGERS, 1991). Especificamente, nesta tese, há o interesse em trabalhar com respostas de itens dicotômicos, modelando a probabilidade de selecionar a resposta correta, ou seja,  $p_{ij}$ , como  $p_{ij} = F(m_{ij})$ , onde  $F$  é chamada de Curva Característica do Item (CCI),  $m_{ij} = a_j(\theta_i - b_j)$ , para  $i = 1, \dots, n$  indivíduos com,  $j = 1, \dots, k$  itens, na qual  $a_j$  e  $b_j$  são parâmetros associados aos itens e  $\theta_i$  é a variável latente associada aos indivíduos. Os parâmetros  $a_j$  e  $b_j$  são chamados de parâmetros de discriminação e dificuldade, respectivamente.

Os modelos TRI tradicionais, como os modelos logísticos de dois e três parâmetros ou ogiva normal, têm a CCI simétrica. Isso significa que o comportamento das CCIs observados à direita do ponto de inflexão é uma imagem espelhada do que ocorre à esquerda do ponto de inflexão (AYALA, 2013).

As primeiras CCIs assimétricas apresentadas na literatura de TRI foram introduzidas por Samejima (2000), em seu estudo, ela introduziu um parâmetro extra no modelo logístico,

que originou o modelo Logístico de Expoente Positivo (LPE). Esse novo parâmetro do item, chamado de complexidade do item, foi incorporado ao expoente e resultou na assimetria da CCI. Outra contribuição foi feita por [Bazán, Bolfarine e Branco \(2006\)](#), que propuseram uma família de modelos com CCI assimétricas usando a função de distribuição cumulativa da normal assimétrica. Além disso, [Bolfarine e Bazan \(2010\)](#) propuseram um novo modelo, chamado RLPE, Reverso Logístico do Expoente Positivo, que é uma variação do modelo LPE de Samejima ([SAMEJIMA, 2000](#)). [Gonçalves, Silva e Loschi \(2023\)](#) oferecem um novo modelo bayesiano TRI que considera de, forma flexível, itens simétricos e assimétricos, obtidos por meio da atribuição de uma distribuição *a priori* de mistura de pontos para o parâmetro de assimetria do item. Por outro lado, [Zhang et al. \(2022\)](#) descrevem uma família de modelos de TRI com ligações de *logit* generalizadas, que incluem os tradicionais modelos ogiva logística e normal como casos especiais.

Embora a importância de se propor novos modelos que possuam CCI assimétrica na TRI já esteja bem estabelecida, até onde se sabe, generalizações da CCI *cloglog* não foram introduzidas na literatura. A proposta de novos links no contexto da regressão binomial visto no capítulo 2, motiva a proposição de algumas CCIs no contexto da Teoria de Resposta ao Item. Todos os modelos podem ser considerados modelos TRI, e possuem como principal característica uma CCI assimétrica. A estimação será desenvolvida usando uma abordagem bayesiana, especificamente o algoritmo NUTS do software Stan ([Stan Development Team, 2020](#)), que pode ser usado para simular distribuições posteriores de parâmetros de itens e variáveis latentes. O algoritmo NUTS é uma extensão do algoritmo de Monte Carlo proposto por [Hoffman e Gelman \(2014\)](#). Este algoritmo permite que a cadeia de Markov explore a distribuição objetiva com mais eficiência do que outros métodos MCMC amplamente conhecidos, como Metropolis Hastings e amostragem de Gibbs, usando a dinâmica hamiltoniana em vez de uma distribuição de probabilidade, permitindo que as cadeias de Markov converjam rapidamente ([NEAL et al., 2011](#)).

O objetivo deste estudo é investigar e propor novos modelos de Teoria de Resposta ao Item (TRI) que apresentem Curvas Características dos Itens (ICC) assimétricas como sua característica principal. O estudo visa utilizar técnicas de estimação Bayesianas para os modelos propostos e comparar o desempenho desses modelos com outros modelos de TRI encontrados na literatura.

O capítulo está organizado da seguinte forma: a Seção 3.2 descreve modelos tradicionais de TRI; a Seção 3.3 apresenta modelos dicotômicos flexíveis da distribuição de Gumbel, que podem ser considerados como CCIs para modelos de TRI; na Seção 3.4 são apresentadas a análise bayesiana, os critérios de comparação utilizados na aplicação e, por fim, a seção 3.5 exemplifica a metodologia através de uma aplicação a dados reais.



## 3.2 Modelos TRI tradicionais

Seja  $Y_{ij}$  a variável aleatória que denota a resposta do  $i$ -ésimo indivíduo ao item  $j$ , sendo  $i = 1, \dots, n$  e  $j = 1, \dots, k$ . O padrão de resposta do  $i$ -ésimo indivíduo é escrito como  $Y_i = (Y_{i1}, \dots, Y_{ik})$ . Quando os itens são pontuados dicotomicamente (correto ou incorreto), os dados observados assumirem os valores  $Y_{ij} = 1$ , para uma resposta correta, e  $Y_{ij} = 0$ , caso contrário. Supõe-se, também, que a probabilidade do evento  $[Y_{ij} = 1]$ , isto é, resposta correta é denotado por  $p_{ij}$  e pode ser escrito como

$$p_{ij} = P[Y_{ij} = 1 \mid \theta_i, a_j, b_j] = F(m_{ij}), \quad (3.1)$$

onde  $F$  é chamado de CCI, e

$$m_{ij} = a_j(\theta_i - b_j), \quad i = 1, \dots, n, \quad j = 1, \dots, k. \quad (3.2)$$

É um preditor linear latente, onde  $a_j$  e  $b_j$  são parâmetros associados aos itens e  $\theta_i$  é a variável latente associada à habilidade ou traço latente do  $i$ -ésimo indivíduo. As variáveis aleatórias  $Y_{ij}$  associadas aos itens são condicionalmente independentes, dado  $\theta_i$ . Conforme indicado anteriormente, os parâmetros  $a_j$  e  $b_j$  são chamados de parâmetros de discriminação e dificuldade, respectivamente.

O primeiro modelo binário de TRI foi introduzido por [Lord \(1952\)](#). A equação (3.1) foi a função de distribuição acumulada (fda) da distribuição normal padrão. Além disso, [Birnbaum \(1968\)](#) considerou a fda da distribuição logística padrão. Esses modelos são frequentemente referidos como o modelo TRI de ogiva normal e o modelo TRI logístico, respectivamente, denotados aqui como modelos TRI 2P e 2L. Quando  $a_j = 1$ , para  $j = 1, \dots, k$  na Equação (3.2), obtêm-se os modelos 1P e 1L, considerando apenas os parâmetros de dificuldade do item. Além disso, podem-se considerar os modelos com três parâmetros como 3P e 3L, propostos por [Sahu \(2002\)](#) que são obtidos considerando  $p_{ij} = c_j + (1 - c_j)F(m_{ij})$  na Equação (3.1), onde  $c_j$  é o parâmetro de adivinhação, indicando que a probabilidade de uma resposta correta é maior que zero.

Os modelos tradicionais de TRI apresentados acima possuem uma CCI simétrica. Recentemente, a função de ligação *cloglog* foi usado por [Robitzsch \(2022\)](#), como uma CCI assimétrica. No entanto, foi notado por [Samejima \(2000\)](#), [Bazán, Bolfarine e Branco \(2006\)](#) assim como [Bolfarine e Bazan \(2010\)](#) que CCIs assimétricas podem ser incorporados usando um novo parâmetro de item que controla a forma da curva.

Os modelos LPE e RLPE assumem que:  $Y_{ij} \mid \theta_i, a_j, b_j \sim \text{Bernoulli}(p_{ij})$ , onde  $p_{ij} = P(Y_{ij} = 1 \mid \theta_i, a_j, b_j) = F_{\lambda_j}(m_{ij})$ ,  $i = 1, \dots, n$  e  $j = 1, \dots, k$ , onde  $\lambda_j > 0$  é o parâmetro de forma e  $F_{\lambda_j}$  é uma função de distribuição acumulada indexada por  $\lambda_j$ . O LPE e o RLPE correspondem a  $F(m_{ij}) = (1 + \exp(-m_{ij}))^{-\lambda_j}$  e  $F(m_{ij}) = 1 - (1 + \exp(m_{ij}))^{-\lambda_j}$ , respectivamente. Observa-se

que, quando  $\lambda_j = 0$ , no modelo LPE, é um caso especial do modelo TRI 2L (Samejima (2000) e Bolfarine e Bazan (2010)).

Assim como na regressão binária, essa assimetria é necessária em situações em que se observam respostas com proporções baixas ou altas de 1s. Alguns autores observam que a especificação incorreta da função de ligação pode resultar em um viés considerável nas estimativas de resposta média (CZADO; SANTNER, 1992). Conforme enfatizado em Chen, Dey e Shao (2001), funções de ligação simétricas nem sempre fornecem um bom ajuste para alguns conjuntos de dados. Alves, Bazán e Arellano-Valle (2022) enfatizam que algumas generalizações da função de ligação *cloglog* acabam sendo interessantes existem dados desbalanceados no contexto da regressão binária. Desta forma, estendem-se as funções de ligação apresentados em Alves, Bazán e Arellano-Valle (2022) no contexto dos modelos de TRI.

### 3.3 Modelos de TRI dicotômicos flexíveis

Nesta seção, são propostos novos modelos TRI dicotômicos modificando a equação (3.1). A proposta começa com a construção das distribuições de Expoente Positivo teorizadas por Bazán *et al.* (2017) e Lemonte e Bazán (2018):

$$p_{ij} = F_{P-RG}(m_{ij}) = [1 - \exp\{-\exp(m_{ij})\}]^{\delta_j}, \quad (3.3)$$

e

$$p_{ij} = F_{RP-RG}(m_{ij}) = 1 - [1 - \exp\{-\exp(-m_{ij})\}]^{\delta_j}, \quad (3.4)$$

com  $\delta_j > 0$ . Além disso, em (3.3) e (3.4) uma reparametrização conveniente é obtida assumindo-se  $\delta_j = \exp(\lambda_j)$ , para  $\lambda_j \in \mathbb{R}^n$ . Essa reparametrização resulta em uma melhoria substancial na eficiência computacional durante o ajuste do modelo. Quando  $\lambda_j = 0 \forall j = 1, \dots, k$ , tem-se o modelo TRI Gumbel e o Modelo TRI Gumbel Reversa como casos particulares, respectivamente. Designam-se esses modelos de modelo TRI *Gumbel Reversa de expoente positivo* (P-RG) e Modelo TRI *Reversa da Gumbel Reversa de expoente positivo* (RP-RG), respectivamente, pelo fato de utilizar-se a fda correspondendo dessas distribuições.

Além disso, também considera-se o processo baseado na construção de distribuições transmutadas assimétricas, seguindo Shaw e Buckley (2009). Os modelos resultantes são dados por:

$$p_{ij} = F_{TS-G}(m_{ij}) = \exp\{-\exp(-m_{ij})\} (1 + \delta_j [1 - \exp\{-\exp(-m_{ij})\}]), \quad (3.5)$$

e

$$p_{ij} = F_{TS-RG}(m_{ij}) = [1 - \exp\{-\exp(m_{ij})\}] (1 + \delta_j \exp\{-\exp(m_{ij})\}), \quad (3.6)$$

onde  $|\delta_j| < 1$ . Considera-se uma reparametrização para (3.5) e (3.6) dada por  $\delta_j = \frac{e^{\lambda_j} - 1}{e^{\lambda_j} + 1}$ , para  $\lambda_j \in \mathbb{R}^n$ . Assim como mencionado anteriormente, essa reparametrização conduz a uma notável otimização da eficiência computacional durante o processo de ajuste do modelo, conforme evidenciado anteriormente no capítulo 3. Observe-se que se  $\lambda_j = 0$ , obtêm-se os modelos TRI Gumbel e Gumbel Reversa, respectivamente, como casos individuais. Esses modelos são chamados de Modelo TRI *Gumbel Transmutada Assimétrica* (TS-G) e Modelo TRI *Gumbel Reversa Transmutada Assimétrica* (TS-RG), respectivamente, porque faz-se uso da fda correspondente dessas distribuições.

A Figura 4 representa as CCIs, na qual a probabilidade de sucesso é dada em função da probabilidade do individuo  $\theta$ , fixando-se  $a$  (parâmetro de discriminação) e  $b$  (parâmetro de dificuldade) em 1 e 0, respectivamente. A partir disso, pode-se dizer que se um item possui uma CCI mais acentuada então pode ser considerada como tendo alto poder de discriminação.

A partir da Figura 6 é possível perceber que, para  $0 < \lambda < 1$ , a CCI do modelo P-RG TRI está abaixo da CCI do modelo com a CCI Gumbel, e se  $\lambda_j < 0$  as CCIs de ambos os modelos estão acima da CCI do modelo com a função de ligação Gumbel. O contrário acontece quando considera-se o modelo RP-RG TRI, ou seja, se  $0 < \lambda_j < 1$  resulta na CCI do modelo RP-RG TRI acima do modelo com a CCI Reversa Gumbel, e se  $0 < \lambda_j$ , temos a CCI do modelo TRI RP-RG abaixo da CCI do modelo Reversa Gumbel TRI.

Os modelos TS-G e TS-RG TRI comportam-se de forma semelhante ao  $\lambda_j$ , ou seja, se  $0 < \lambda_j < 1$  a CCI desses modelos ficará acima da CCI do modelo com a CCI Gumbel, e se  $\lambda_j < 0$  a CCI ficará abaixo da CCI do modelo Gumbel TRI.

O parâmetro  $\lambda_j$  nos modelos de TRI propostos está associado à forma assimétrica da CCI ou, equivalentemente, à assimetria considerada no erro latente da regressão da variável auxiliar latente em relação ao traço latente subjacente da resposta correta ou incorreta do item. Este parâmetro extra é interpretado de duas formas na literatura, sendo parâmetro de penalidade [Bolfarine e Bazan \(2010\)](#) ou como parâmetro de aceleração ([SAMEJIMA, 2000](#)). Ambos os termos serão utilizados neste trabalho sem distinção, pois fazem parte de um parâmetro de complexidade ([BOLT; LIAO, 2022](#)).

## 3.4 Inferência

Nesta seção, será apresentada a inferência estatística. Primeiramente, será feita a introdução da função de verossimilhança e de como ela expressa a probabilidade dos dados em relação aos parâmetros do modelo. Na especificação das distribuições *a priori*, será apresentadas as distribuições *a priori* utilizadas neste capítulo. Por fim, serão discutidos os critérios de comparação dos modelos, ferramentas essenciais para selecionar o modelo mais adequado aos dados observados.

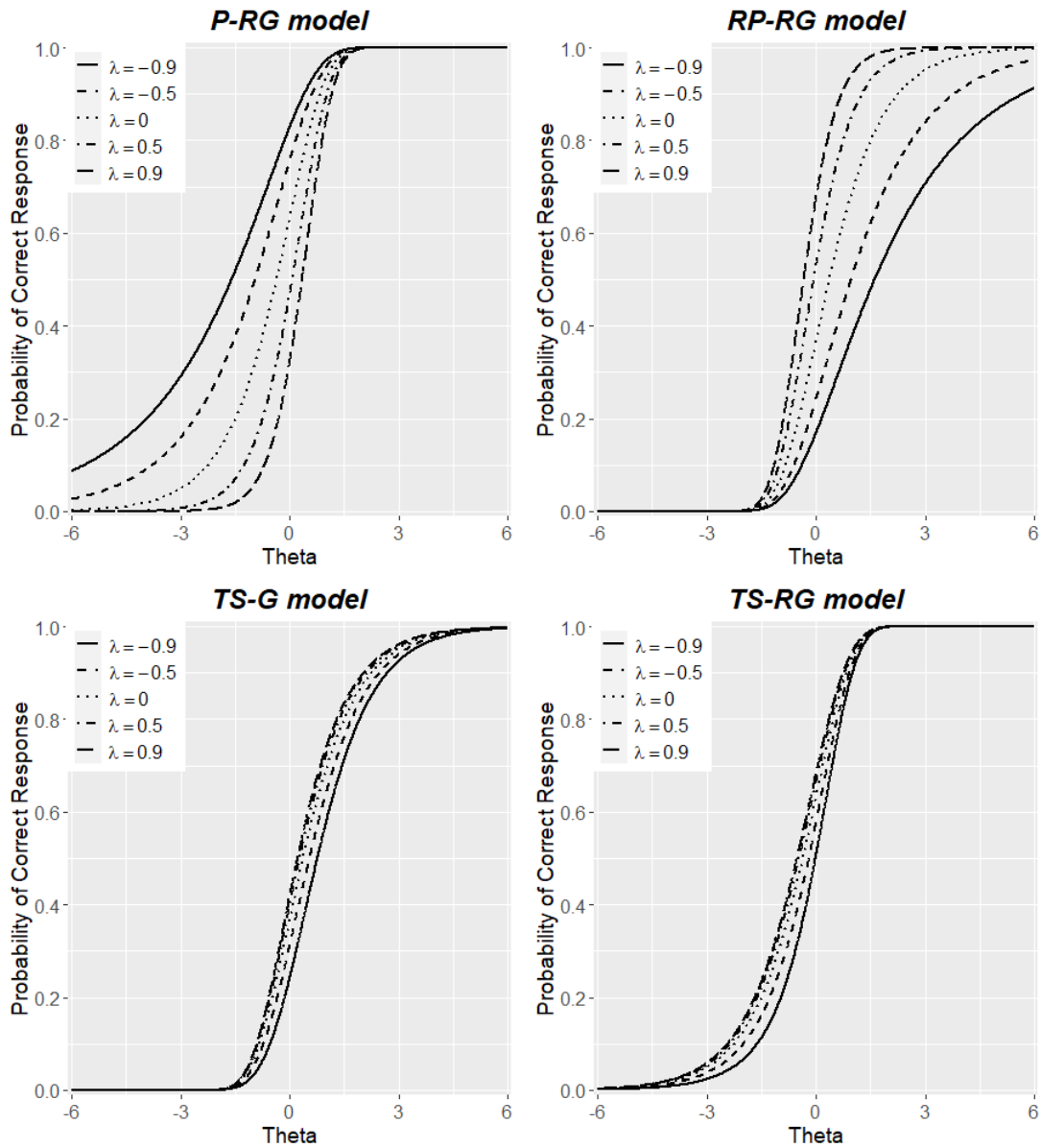


Figura 4 – Curvas de probabilidade para  $\lambda = -0,9, -0,5, 0, 0,5, 0,9$  nos modelos P-RG, RP-RG, TS-G e TS-RG considerando diferentes faixas para  $\theta$  e  $a = 1$  e  $b = 0$

Fonte: Elaborada pelo autor.

### 3.4.1 Função de Verossimilhança

Considerando as equações de 3.3 a 3.6, utiliza-se  $F_{\lambda_j}(m_{ij})$  como notação geral para a fda. Para os modelos *cloglog* flexíveis indexados por  $\lambda_j$ , a função de verossimilhança é dada por

$$L(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \prod_{j=1}^k \left\{ F_{\lambda_j}(m_{ij}) \right\}^{y_{ij}} \left\{ 1 - F_{\lambda_j}(m_{ij}) \right\}^{1-y_{ij}}, \quad (3.7)$$

onde  $\boldsymbol{\beta} = (\mathbf{a}^T, \mathbf{b}^T)$ ,  $\mathbf{a} = (a_1, \dots, a_k)^T$ ,  $\mathbf{b} = (b_1, \dots, b_k)^T$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$ ,  $m_{ij}$  é o preditor linear latente na equação (3.2), para  $i = 1, \dots, n$  e  $j = 1, \dots, k$ , e  $F_{\lambda_j}$  é o fda P-RG, RP-RG, TS-G e TS-RG nas equações (3.3-3.6), indexado com  $\lambda_j$  e avaliado em  $m_{ij}$ .

### 3.4.2 Abordagem Bayesiana

Neste estudo, consideramos especificamente a estimação bayesiana com métodos de Monte Carlo via cadeia de Markov que facilitam a amostragem eficiente da distribuição marginal posterior dos parâmetros de interesse. Essa escolha pela estimação bayesiana foi feita considerando os trabalhos de [Swaminathan et al. \(2003\)](#), que demonstraram que a estimativa precisa dos parâmetros de itens em amostras pequenas é obtida usando a abordagem bayesiana. Para tal, usa-se o pacote `rstan` no software [Team \(2009\)](#), que usa o algoritmo No-U-Turn Sampler (NUTS) ([HOFFMAN; GELMAN, 2014](#)). [Luo e Jiao \(2018\)](#) demonstrou, em um estudo comparativo com outros softwares BUGS, que o STAN é consideravelmente mais rápido na estimativa de modelos de TRI.

### 3.4.3 Especificação das distribuição à priori

A especificação da distribuição à priori é um aspecto importante da análise bayesiana, especialmente no caso de amostra pequena ([BOLFARINE; BAZAN, 2010](#)). A escolha de uma distribuição à priori adequada no traço latente resolve problemas particulares de identificação ([ALBERT; GHOSH, 2000](#)). Na literatura da TRI, em geral utilizada para distribuição à priori para o traço latente  $\theta_i \sim N(0, 1)$ , para  $i = 1, \dots, n$ .

Com relação às distribuições à priori dos parâmetros  $a_j$  e  $b_j$  dos itens, não há um consenso, ([JOHNSON; ALBERT, 1999](#); [CONGDON, 2007](#); [ALBERT; GHOSH, 2000](#); [SAHU, 2002](#); [SPIEGELHALTER et al., 1996](#); [PATZ; JUNKER, 1999](#)). Portanto, usa-se a distribuição à priori independentes para  $a$ ,  $b$  e  $\lambda$  e deixa-se tais correlações dependentes de dados, como visto em [Bolfarine e Bazan \(2010\)](#). Ou seja, a distribuição à priori que se considera é escrita como

$$\pi(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \boldsymbol{\lambda}) = \prod_{i=1}^n \Phi(\theta_i) \prod_{j=1}^k \pi_1(a_j) \pi_2(b_j) \pi_3(\lambda_j) \quad (3.8)$$

onde  $\Phi(\cdot)$  é a fdp da distribuição normal padrão e  $\pi_1(\cdot)$ ,  $\pi_2(\cdot)$ ,  $\pi_3(\cdot)$  são as fdp à priori para os parâmetros  $a_j$ ,  $b_j$  e  $\lambda_j$ , respectivamente. Assim,  $a_j \sim LN(0, 1)$  e  $b_j \sim N(0, 1)$  (ver, (BOLFARINE; BAZAN, 2010)).

A distribuição à priori considerada para  $\lambda_j$  é a  $U(-2, 2)$  baseado em Alves, Bazán e Arellano-Valle (2022) com dados de regressão binária.

### 3.4.4 Critérios de comparação de modelos

Existem várias metodologias para comparar modelos alternativos em uma estrutura bayesiana. Aqui, considera-se o critério de informação de Watanabe-Akaike (WAIC) baseado no parâmetro de complexidade ( $p_{WAIC}$ ), proposto por Watanabe e Opper (2010), e definido como:  $WAIC = -2lppd + 2p_D$ , onde  $lppd$  é o logaritmo da densidade preditiva pontual e  $p_D$  é o número efetivo de parâmetros. Outra medida chamada ‘leave-one-out cross-validation’ (LOO), proposta por Geisser e Eddy (1979) foi utilizado. Finalmente, o Deviance Information Criterion (DIC) conforme a definição encontrada em Brooks (2002):  $DIC = \overline{D(\theta)} + 0,5var(D(\theta))$ , onde  $\overline{D(\theta)}$  é o desvio médio posterior, foi usado para comparação do modelo. Para todos os critérios acima, valores menores indicam melhor ajuste.

## 3.5 Aplicação

Uma aplicação foi utilizada para ilustrar as abordagens Bayesianas desenvolvidas neste trabalho para modelos TRI com respostas binárias. Foram usados dados do Bolfarine e Bazan (2010), que correspondem a um teste de matemática para alunos da quarta série em escolas primárias rurais peruanas. Existem 974 respostas dos alunos para 18 itens que se qualificam como respostas binárias (corretas ou incorretas). Esses dados são desbalanceados e têm proporções de 1s (respostas corretas) para cada item, respectivamente, dadas por 0,72, 0,61, 0,43, 0,37, 0,50, 0,08, 0,65, 0,27, 0,53, 0,80, 0,48, 0,67, 0,47, 0,57, 0,30, 0,28, 0,12, 0,42.

Assim sendo, é mostrado um estudo sobre o ajuste dos modelos paramétricos da TRI, discutidos anteriormente na Seção 3.3, usando os dados do teste matemático. Modelos logísticos TRI com dois parâmetros (2L), modelos LPE e RLPE propostos por (BOLFARINE; BAZAN, 2010) são considerados nessa comparação. Em todos os casos, as distribuições à priori usadas são  $\theta_i \sim N(0, 1)$ ,  $b_j \sim N(0, 1)$ ,  $a_j \sim LN(0, 1)$  para  $i = 1, \dots, 974$  e  $j = 1, \dots, 18$ . Para  $\lambda_j$ , considera-se à priori  $U(-2, 2)$ , para  $j = 1, \dots, 18$ .

Nesta aplicação, usa-se o algoritmo de estimação NUTS, em R. Com o rstan, são considerados os seguintes parâmetros MCMC: 20,000 iterações com 10,000 iterações de descarte, juntamente com um intervalo de espaçamento de 10 iterações para alcançar a convergência usando 3 cadeias MCMC. Consideramos a verificação de convergência baseada na estatística  $\hat{R}$  (GELMAN; RUBIN *et al.*, 1992). Os códigos stan utilizado pode ser encontrado no Apêndice B.

Para comparar os modelos, utiliza-se os critérios de seleção DIC, WAIC e LOO, discutidos na Seção 4.2. Esses valores são mostrados na Tabela 10.

Tabela 10 – Critérios de comparação de modelos para dados de teste de matemática

	DIC	WAIC	LOO	$\hat{R}_a$	$\hat{R}_b$	$\hat{R}_\lambda$	$\hat{R}_\theta$
2L	16332,96	17754,66	17769,34	1,00	1,00	-	1,00
LPE	16281,30	17731,34	17753,75	1,00	1,00	1,00	1,00
RLPE	16262,84	17713,35	17737,99	1,01	1,00	1,01	1,00
P-RG	16230,57	17723,74	17744,28	1,12	1,14	1,14	1,11
<b>RP-RG</b>	<b>16258,68</b>	<b>17694,58</b>	<b>17726,59</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
TS-G	16273,34	17701,97	17726,57	1,05	1,06	1,06	1,02
TS-RG	16371,22	17795,55	17817,49	1,11	1,15	1,15	1,03

Fonte: Elaborada pelo autor.

Pode-se observar na Tabela 10, que o melhor modelo proposto considerando os critérios estudados é o modelo TRI RP-RG. Observa-se que, para todos os parâmetros nos modelos propostos o  $\hat{R}$  ficou em torno de 1, o que indica que não foram detectados problemas de convergência.

Agora, comparam-se os modelos TRI RP-RG e RLPE. Assim, segue a análise descrevendo três itens: o 14, o 2 e o 11, que são os mesmos itens analisados por [Bolfarine e Bazan \(2010\)](#). Para o modelo RLPE, o item 14 é considerado um item assimétrico, o item 2 possui algum grau de assimetria e o item 11 não possui assimetria ([BOLFARINE; BAZAN, 2010](#)).

A Tabela 11 mostra os parâmetros dos itens para o modelo alternativo da TRI, que teve o melhor ajuste aos conjunto de dados trabalhado. Vale ressaltar que os resultados obtidos neste estudo consideram a mudança na variável  $\delta = \exp(\lambda)$ , na qual  $\lambda \sim U(-2, 2)$ , que é denominada de priori indireta, diferente da encontrada em [Bolfarine e Bazan \(2010\)](#), cuja distribuição à priori  $\lambda \sim \text{Gamma}(0.25, 0.25)$  foi utilizada. A distribuição à priori  $U(-2, 2)$  foi escolhida porque o trabalho de [Alves, Bazán e Arellano-Valle \(2022\)](#) mostra que melhores resultados foram obtidos usando a priori indireta.

Tabela 11 – Parâmetros de itens para modelos TRI alternativos para o item 14, item 2 e item 11 em dados matemáticos

Itens	Modelos	Itens Parametros					
		Discriminação		Dificuldade		Aceleração	
		$a$	$\hat{R}_a$	$b$	$\hat{R}_b$	$\lambda$	$\hat{R}_\lambda$
Item 14 (Assimétrico)	RPLE	2.69	1.00	-1.51	1.01	-1.41	1.01
	RP-RG	1.95	1.00	-1.41	1.00	-0.98	1.00
Item 02 (Algum grau de assimetria)	RPLE	1.05	1.00	0.48	1.00	0.69	1.00
	RP-RG	0.54	1.00	0.36	1.00	1.00	1.00
Item 11 (Assimetria não evidente)	RPLE	1.98	1.00	0.15	1.00	0.01	1.00
	RP-RG	0.95	1.00	0.33	1.00	0.67	1.00

Fonte: Elaborada pelo autor.



Ao comparar os modelos TRI RP-RG e RLPE em termos do parâmetro  $\lambda$ , observa-se magnitudes diferentes no caso assimétrico, mas com sinais iguais. No caso assimétrico, as magnitudes são semelhantes. Também observa-se  $\lambda = 0,67$  no Item 11 (caso em que a assimetria não é evidente) para o modelo RP-RG TRI. Isso é esperado para este modelo, pois sabe-se que ele não detecta itens simétricos. Quanto aos parâmetros de dificuldade e discriminação, observou-se igualdade de sinais. Quanto à magnitude, nota-se uma diferença entre os modelos TRI RP-RG e RLPE para o parâmetro de discriminação, mas esta magnitude é semelhante para os parâmetros de dificuldade. Veja que para todos os parâmetros nos modelos propostos encontra-se  $\hat{R} \approx 1$ .

Pode-se observar na Figura 5 que para o item 14, a CCI do modelo RP-RG mostra que se o indivíduo não possui habilidade, ou esta habilidade é pequena ( $\theta < -3$ ), a probabilidade de acertar o item é praticamente zero, mas se ele tiver algum nível de habilidade ( $-2.5 < \theta < 0$ ) a probabilidade de acertar o item aumenta significativamente. E essa habilidade for maior que 0, a probabilidade de acertar o item é praticamente 1. Observe que o modelo TRI RP-RG e o RLPE mostram que esse item é assimétrico.

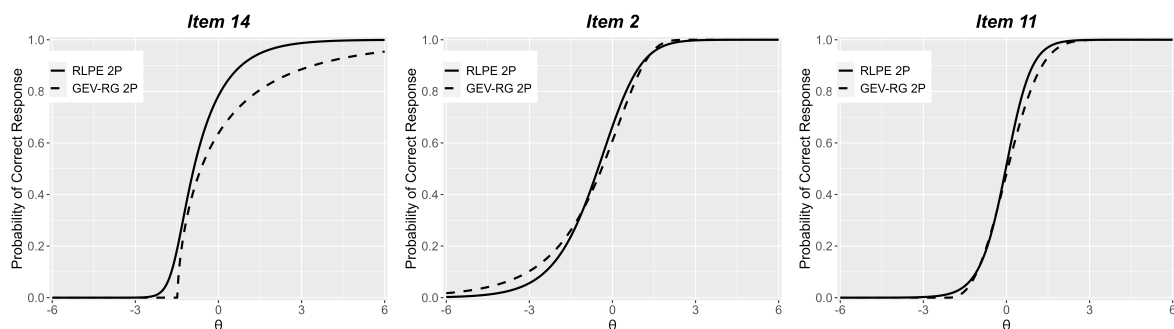


Figura 5 – Curvas características do item (CCIs) para os itens 14, 2 e 11 nos modelos TRI RP-RG e RLPE nos dados matemáticos

Fonte: Elaborada pelo autor.

O item 2, a probabilidade de acertar o item, mesmo que o indivíduo tenha uma habilidade extremamente baixa ( $\theta < -6$ ), é diferente de zero para o modelo TRI RP-RG. Já se a habilidade for maior que 0, a probabilidade de acertar é bem próxima de 1. Isso significa que esse item não é um item extremamente complexo. Nota-se que, no item 11,  $\lambda = 0,01$  para o modelo RPLE, que é um caso particular onde a curva é simétrica. Quanto ao modelo TRI RP-RG,  $\lambda = 0,67$  que também se aproxima de zero, então, a curva é próxima ao modelo TRI Reverso Gumbel. Pode-se assim dizer, em relação ao item 11, que o indivíduo não precisa ter subprocessos sequenciais para selecionar uma resposta correta, portanto, é um item simétrico.

A Figura 6 mostra a relação entre os parâmetros de discriminação, dificuldade e aceleração para o modelo TRI RP-RG. Pode-se notar que existe alguma relação entre os parâmetros de discriminação e o parâmetro  $\lambda$ , ou seja, conforme o valor do parâmetro de discriminação ( $a$ ) aumenta, o valor do parâmetro de precisão ( $\lambda$ ) diminui. Os outros parâmetros são, aparentemente, não relacionados.



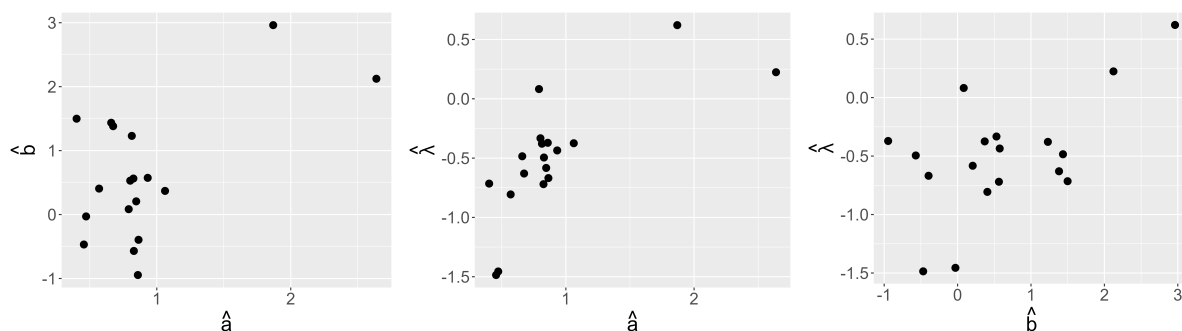


Figura 6 – Relação entre os parâmetros de discriminação, dificuldade e aceleração, para o modelo TRI RP-RG

Fonte: Elaborada pelo autor.

## 3.6 Comentários finais

No âmbito deste capítulo, apresentamos generalizações das Curvas Características Item (CCI) fundamentadas na CCI *cloglog* como uma inovadora alternativa aos modelos convencionais da Teoria de Resposta ao Item (TRI). Vale destacar que o *cloglog* foi previamente adotado por [Robitzsch \(2022\)](#) e representa um caso específico do modelo aqui proposto.

Durante este estudo, foi evidenciada a relevância do emprego de CCIs flexíveis para capturar a assimetria da variável latente, o que se revelou crucial para uma modelagem mais precisa e interpretativa dos resultados.

Na aplicação dos modelos aos dados provenientes da prova de matemática, observou-se que o modelo TRI RP-RG demonstrou um ajuste superior em comparação com os modelos TRI propostos por [Bolfarine e Bazan \(2010\)](#). Este resultado ressalta a eficácia e a utilidade dos modelos propostos, especialmente em contextos que envolvem dados dicotômicos não balanceados na TRI.

Os modelos propostos não apenas representam uma alternativa viável aos métodos convencionais, mas também oferecem uma ferramenta valiosa e eficiente para a análise de dados, permitindo a captura efetiva da assimetria presente nos dados. Esta capacidade de lidar com a assimetria é fundamental em diversos campos de estudo, pois muitas vezes os dados apresentam distribuições não simétricas que podem influenciar significativamente a interpretação e a análise dos resultados.

Ao proporcionar uma abordagem flexível para modelar a assimetria da variável latente, os modelos propostos têm o potencial de melhorar a precisão e a interpretabilidade das análises estatísticas. Isso não só beneficia a área de avaliação educacional, mas também tem aplicações em uma variedade de outros domínios, incluindo ciências sociais, economia, medicina e muitos outros.



# UM NOVO MODELO DE TEORIA DE RESPOSTA AO ITEM COM A DISTRIBUIÇÃO GENERALIZADA DO VALOR EXTREMO

Este capítulo é semelhante ao anterior, porém, nele é apresentada uma investigação mais detalhada de dois modelos específicos, utilizando a distribuição do valor extremo GEV-G na Teoria de Resposta ao Item (TRI).

A Teoria da Resposta ao Item (TRI) é uma técnica estatística amplamente utilizada em diversas áreas, como educação e psicométrica, para modelar dados de testes de múltipla escolha e estimar as habilidades dos participantes. A TRI inclui uma grande família de modelos, na qual é descrita a relação matemática entre uma característica observada, medida por um teste, e a característica subjacente, latente, que a origina, associada ao indivíduo e aos itens do teste.

Formalmente, seja  $Y_{ij}$  ( $i = 1, \dots, n; j = 1, \dots, k$ ) a variável aleatória que denota a resposta binária do indivíduo  $i$  para o item  $j$ , com  $Y_{ij} = 1$ , para uma resposta correta e  $Y_{ij} = 0$ , para uma resposta incorreta. O modelo estatístico é especificado como:

$$Y_{ij} \stackrel{ind.}{\sim} \text{Bernoulli}(p_{ij}), \quad p_{ij} = P[Y_{ij} = 1] = F(m_{ij}), \quad (4.1)$$

onde  $F$  é uma função estritamente crescente geralmente chamada de Função de Resposta ao Item (FRI), e  $m_{ij}$  é um parâmetro que caracteriza os indivíduos, e também os itens. Uma representação gráfica de  $F$  é chamada de Curva Característica do Item (CCI). Aqui, usam-se ambos sem perda de generalidade.

Duas escolhas comuns para  $F$  são as funções de distribuição acumulada (fda) logística padrão e normal padrão, ou seja, modelos TRI logístico e ogiva normal, respectivamente. O parâmetro  $m_{ij}$  pode ser estruturado de diferentes maneiras, sendo a escolha  $m_{ij} = a_j(\theta_i - b_j)$  uma das mais utilizadas na prática. Aqui,  $a_j > 0$ ,  $b_j \in \mathbb{R}$  e  $\theta_i \in \mathbb{R}$  são comumente conhecidos

como os parâmetros de discriminação, dificuldade e habilidade associados ao  $j$ -ésimo item e ao  $i$ -ésimo indivíduo, respectivamente.

A relação representada pela CCI ou FRI mostra como a probabilidade de responder corretamente muda com o traço latente (para itens dicotômicos). Como a probabilidade não pode ser inferior a 0 ou maior do que 1, a CCI geralmente possui uma forma de  $S$  alongada e simétrica nos modelos mais populares da TRI, conhecidos como de um parâmetro (1L), dois parâmetros (2L) e três parâmetros (3L), que utilizam a função de distribuição acumulada (fda) da distribuição logística (HAND, 1998).

A forma explícita de  $F$  pode ser derivada a partir do que é conhecido como representação latente do modelo, ou interpretação da resposta latente para uma resposta dicotômica, segundo Skrandal e Rabe-Hesketh (2004), na qual uma variável latente de resposta contínua  $Z_{ij}$  é assumida para gerar as respostas binárias, considerando o preditor linear latente  $m_{ij}$  e  $\varepsilon_{ij}$  um erro aleatório latente seguindo alguma distribuição  $F$  considerando:

$$Y_{ij} = \begin{cases} 1, & Z_{ij} > 0 \\ 0, & Z_{ij} \leq 0 \end{cases}, \quad (4.2)$$

$$Z_{ij} = m_{ij} + \varepsilon_{ij} \quad (4.3)$$

É facilmente verificável que, sob a representação latente,  $p_{ij} = P[Z_{ij} > 0] = 1 - F(-m_{ij})$ , e quando a distribuição de erro latente é a logística padrão ou a normal padrão, então, os modelos ogiva logística e ogiva normal comumente utilizados são obtidos por  $p_{ij} = 1 - F(-m_{ij}) = F(m_{ij})$ . O fato da Função de Resposta ao Item (FRI) coincidir, exatamente, com a fda da distribuição de erro se deve à natureza simétrica das distribuições logística e normal.

No entanto, essa característica não se generaliza quando são consideradas distribuições assimétricas como CCI. Em vez disso, se forem levadas em conta distribuições assimétricas como  $F(\cdot)$ , a FRI resultante corresponde à variável de resposta latente  $Z_{ij}$ , que é diferente da distribuição de erro latente  $\varepsilon_{ij}$ . Na verdade, a distribuição de  $Z_{ij}$  é uma versão invertida da distribuição de  $\varepsilon_{ij}$ , a recíproca também é válida. Esse fenômeno é denominado, no contexto da regressão binária, como a propriedade de inversão por Bazán *et al.* (2017). Formalmente, afirma-se que se  $S$  e  $W = -S$  são duas variáveis aleatórias tal que  $S \sim F_S$  e  $W \sim F_W$ ; diz-se que  $F_S$  e  $F_W$  satisfazem a *propriedade de inversão* se  $F_W(t) \neq F_S(t)$ , essa informação pode ser representada por  $F_W(t) = 1 - F_S(-t)$  (BAZÁN *et al.*, 2017). Nesse caso,  $F_S$  e  $F_W$  não têm a mesma distribuição, mas podem ser expressas de maneira semelhante como na equação 4.2 em termos da distribuição de erro.

A utilidade das CCIs assimétricas tem sido amplamente documentada nos trabalhos citados anteriormente e, também, por Lee e Bolt (2018). Além disso, Bazán *et al.* (2023) demonstram que é possível estender esse tipo de modelo para acomodar outras estruturas, como o teste de conjunto, onde os itens não são independentes. A inclusão de novas CCIs na TRI é

importante para acomodar diferentes padrões de resposta observados, sendo equivalente ao caso de inclusão de novas ligações na regressão binária, o que pode ser justificado pelo desequilíbrio entre 0 e 1 (BAZÁN *et al.*, 2017; LEMONTE; BAZÁN, 2018; BOLFARINE; BAZAN, 2010).

Este capítulo propõe a introdução de novas FRI em modelos TRI para dados dicotômicos, com base na representação latente do modelo, considerando a distribuição Generalizada do Valor Extremo (GEV). A utilização da distribuição GEV para propor uma função de ligação foi proposta na regressão binária por Wang, Dey *et al.* (2010) e Li *et al.* (2016). Os resultados desses autores mostram que essa ligação é flexível para ajustar a assimetria na curva de resposta com um parâmetro de forma livre, o que torna essa distribuição atraente para uso na TRI. Assim sendo, são consideradas essa ligação para formular uma CCI, assim como em Bolfarine e Bazan (2010), utilizando a propriedade de inversão para propor duas distribuições para o erro latente  $\varepsilon_{ij}$ . Caso seja utilizada a distribuição GEV como distribuição para esse erro, obtêm-se uma CCI considerando a distribuição Reversa da GEV, e vice-versa; se a Reversa da GEV for considerada como distribuição para esse erro, resulta uma CCI considerando a GEV. Ambas as CCIs abrangem situações com proporções mais baixas ou mais altas de 1s observadas nos dados. Portanto, ao considerar essa característica, a abordagem estabelecida para propor CCIs assimétricas difere das outras formulações de CCIs assimétricas citadas acima, que utilizam diretamente a distribuição  $F$ .

A estimativa dos modelos de TRI propostos é realizada utilizando uma abordagem Bayesiana, mais especificamente, o algoritmo NUTS do software Stan (Stan Development Team, 2020). Este algoritmo é capaz de simular as distribuições a posteriori dos parâmetros do item e variáveis latentes. Ele foi desenvolvido como uma extensão do algoritmo de Monte Carlo e permite que a cadeia de Markov explore a distribuição objetiva de forma mais eficiente que outros métodos MCMC, pois utiliza dinâmica hamiltoniana em vez de uma distribuição de probabilidade, o que possibilita uma convergência mais rápida das cadeias de Markov (HOFFMAN; GELMAN, 2014).

Este capítulo está organizado da seguinte forma: a Seção 4.1 descreve modelos de TRI dicotômicos, introduz brevemente a distribuição GEV e apresenta os modelos propostos com base na distribuição do valor extremo e sua reversa; a Seção 4.2 detalha a abordagem Bayesiana adotada para inferência, especificando as prioris e os critérios de comparação de modelos utilizados no estudo. Na Seção 4.3, é apresentado um estudo de sensibilidade das prioris para o novo parâmetro, além de um estudo de recuperação de parâmetros e, por fim, um estudo de desempenho dos critérios de comparação de modelos; na Seção 4.4, os novos modelos são ajustados e exemplificados por meio de uma aplicação em dois conjuntos de dados. o capítulo é concluído na Seção 4.5 com algumas conclusões e discussões.

## 4.1 O Modelo TRI GEV-G

Nesta seção, descreve-se a especificação geral dos modelos TRI, que serão úteis para introduzir os modelos propostos GEV-TRI. Também é apresentada uma breve revisão da distribuição GEV.

### 4.1.1 A Distribuição GEV

Seja  $Y \sim GEV$  uma variável aleatória (v.a), e  $X = \frac{Y-\mu}{\sigma}$  a versão padronizada. A fda da distribuição padronizada GEV é escrita como (KOTZ; NADARAJAH, 2000):

$$F_X(x) = \begin{cases} \exp\{-(1 + \lambda x)^{-1/\lambda}\} & -\infty < x \leq -\frac{1}{\lambda} \quad \text{para } \lambda < 0; \\ & -\frac{1}{\lambda} < x < \infty \quad \text{para } \lambda > 0; \\ \exp\{-\exp\{-x\}\} & -\infty < x < \infty \quad \text{para } \lambda = 0 \end{cases} \quad (4.4)$$

ou, em uma forma mais compacta:

$$F_X(x) = \exp\{-(1 + \lambda x)_+^{-1/\lambda}\},$$

onde  $x_+ = \max\{0, x\}$ . Utiliza-se a notação  $X \sim GEV-G(0, 1, \lambda)$  ou  $F_X(x) \equiv GEV-G(x; \lambda)$ . Observe que quando  $\lambda = 0$ , obtêm-se a distribuição Gumbel padrão (KOTZ; NADARAJAH, 2000). Também observe que a distribuição GEV satisfaz a propriedade de reversão, portanto, a fda de sua reversa, que se chama de GEV-RG, pode ser escrita como  $1 - \exp\{-(1 - \lambda x)_+^{-1/\lambda}\}$ , pois resulta na distribuição Gumbel Reversa padrão. Para mais detalhes sobre a distribuição GEV (KOTZ; NADARAJAH, 2000).

A Figura 7 mostra a função de densidade das distribuições GEV-G e GEV-RG para diferentes valores de  $\lambda$ , quando este parâmetro assume valores positivos e negativos. A partir da Figura 9, pode-se observar que a distribuição GEV-G permite tanto assimetria para a esquerda quanto para a direita para valores de  $\lambda$ , que são negativos ou positivos, respectivamente. O oposto ocorre para a distribuição GEV-RG, na qual valores negativos e positivos de  $\lambda$  produzem assimetria para a direita e para a esquerda, respectivamente. Como mencionado anteriormente, o caso  $\lambda = 0$  corresponde à distribuição Gumbel. Outro ponto a ser observado é que os valores de  $x$  dependem de  $\lambda$ , tendo um limite superior em  $1/|\lambda| > 0$  quando  $\lambda < 0$  e um limite inferior em  $-1/\lambda < 0$  quando  $\lambda > 0$ . Por exemplo, quando  $\lambda = -0,9$ , o limite superior para  $x$  na função de densidade de probabilidade (fdp) da GEV é 1,111, e quando  $\lambda = 0,9$ , o limite inferior para  $x$  na fdp da GEV é -1,111.

O parâmetro  $\lambda$  está associado à assimetria das distribuições GEV-G e GEV-RG. Essa assimetria induzirá CCI's assimétricas quando se assume que o termo de erro segue uma distri-

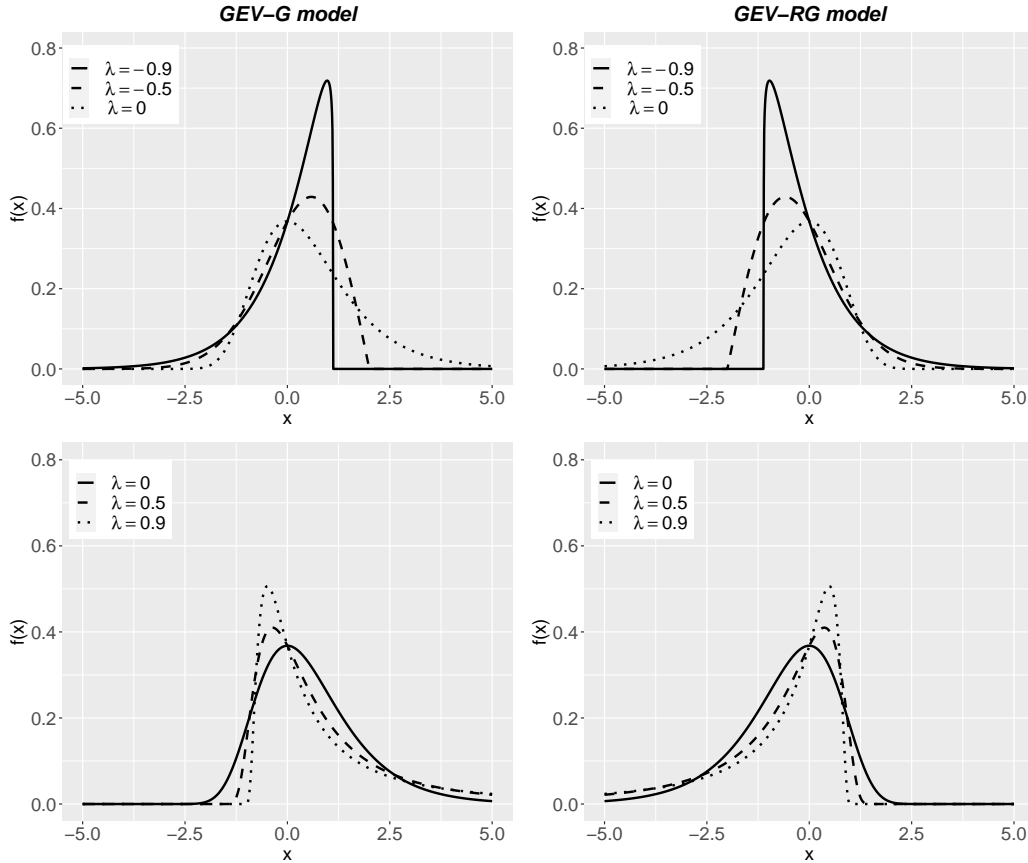


Figura 7 – Curvas de densidade das distribuições GEV-G e GEV-RG para diferentes  $\lambda$ s

Fonte: Elaborada pelo autor.

buição GEV-G na representação latente do modelo 4.1. Essas ICCs assimétricas são a principal característica dos modelos que propostos a seguir.

### 4.1.2 Os modelos propostos

Para a representação latente do modelo (4.1), assume-se que o termo de erro segue uma distribuição padrão GEV-G. Então, é facilmente demonstrado que:

$$p_{ij} = 1 - \exp \left\{ - \left( 1 - m_{ij} \lambda_j \right)_+ \right\}^{-1/\lambda_j},$$

e, portanto, a FRI corresponde à fda de uma distribuição que satisfaz a condição de reversão, neste caso, a distribuição GEV-G. Utiliza-se tanto a distribuição GEV-G quanto a distribuição GEV-RG para propor dois modelos TRI, conforme especificado em (4.1), com

$$p_{ij} = F_{GEV-G}(m_{ij}) = \exp \left\{ - \left( 1 + m_{ij} \lambda_j \right)_+^{-1/\lambda_j} \right\}, \quad (4.5)$$

e

$$p_{ij} = F_{GEV-RG}(m_{ij}) = 1 - \exp \left\{ - (1 - m_{ij}\lambda_j)_+^{-1/\lambda_j} \right\}, \quad (4.6)$$

no qual  $\lambda_j \in \mathbb{R}^n$ . Os dois modelos serão chamados de GEV-G TRI e GEV-RG TRI, respectivamente. No modelo GEV-G TRI, para  $\lambda_j = 0$ , a CCI corresponde ao modelo TRI quando se considera o modelo TRI de Gumbel. Quando  $\lambda_j = 0$  no modelo GEV-RG TRI, surge o modelo TRI de Gumbel Reversa.

Note que se  $\theta = b$  ( $m = 0$ ), obtêm-se que  $p = \exp(-1) \approx 0,368$  para o GEV-G TRI e  $p = 1 - \exp(-1) \approx 0,632$  para o GEV-RG TRI, o que difere de  $p = 0,5$  se a distribuição logística ou normal forem consideradas como FRI. O GEV-G TRI está mais associado a probabilidades menores de respostas corretas, enquanto o GEV-RG TRI está mais associado a probabilidades maiores de respostas corretas, mostrando que essas FRIs são assimétricas.

A Figura 8 mostra as CCIs para o modelo GEV-G, estabelecendo diferentes valores de  $a$ ,  $b$  e  $\lambda$ . Para todos os valores de  $\lambda$ , as CCIs do modelo GEV-G está abaixo da CCI do modelo Gumbel. Além disso, quando o parâmetro  $b$  aumenta de  $b = 0$  para  $b = 1$ , para um mesmo valor de  $\theta$ , a probabilidade de resposta correta diminui para qualquer valor de  $\lambda$ . Assim, pode-se manter a interpretação de  $b$  como um parâmetro de dificuldade. Da mesma forma, quando o parâmetro  $a$  aumenta de  $a = 1$  para  $a = 1,5$ , a inclinação da ICC se torna mais íngreme, indicando que a mesma mudança em  $\theta$  produz uma mudança maior na probabilidade de resposta correta.

No modelo GEV-G TRI, a inclinação da CCI no ponto  $b$  é obtida derivando a função correspondente da CCI uma vez em relação a  $\theta$  e, em seguida, obtendo  $p'(b)$ .

$$\begin{aligned} \frac{\partial F_{GEV-RG}}{\partial \theta} \Big|_{\theta=b} &= a(1 - a\lambda(b-b))_+^{-\frac{1}{\lambda}-1} \exp \left\{ -(1 - a\lambda(b-b))_+^{-1/\lambda} \right\} \\ &= a(1)_+^{-\frac{1}{\lambda}-1} \exp \left\{ -(1)_+^{-\frac{1}{\lambda}} \right\} = a \exp\{-1\} = 0.3678794a \end{aligned} \quad (4.7)$$

o que é o mesmo para o GEV-G.

Note que nos modelos GEV-G e GEV-RG, a inclinação em  $b$  depende apenas do parâmetro  $a$ , assim como no modelo 2L e, então, a denominação de  $a$  como parâmetro de discriminação é justificada. Pode não ser o caso no modelo proposto por Bazán *et al.* (2023), onde a inclinação em  $b$  é  $\frac{d}{2^{d-1}4}$  e também depende do parâmetro  $d$ .

## 4.2 Inferência

**Função de Verossimilhança:** para as Expressões 4.5 e 4.6, utiliza-se  $F_{\lambda_j}(m_{ij})$  como notação geral. Para os modelos TRI GEV-G e GEV-RG indexados por  $\lambda_j$ , a função de verossimilhança é dada por:



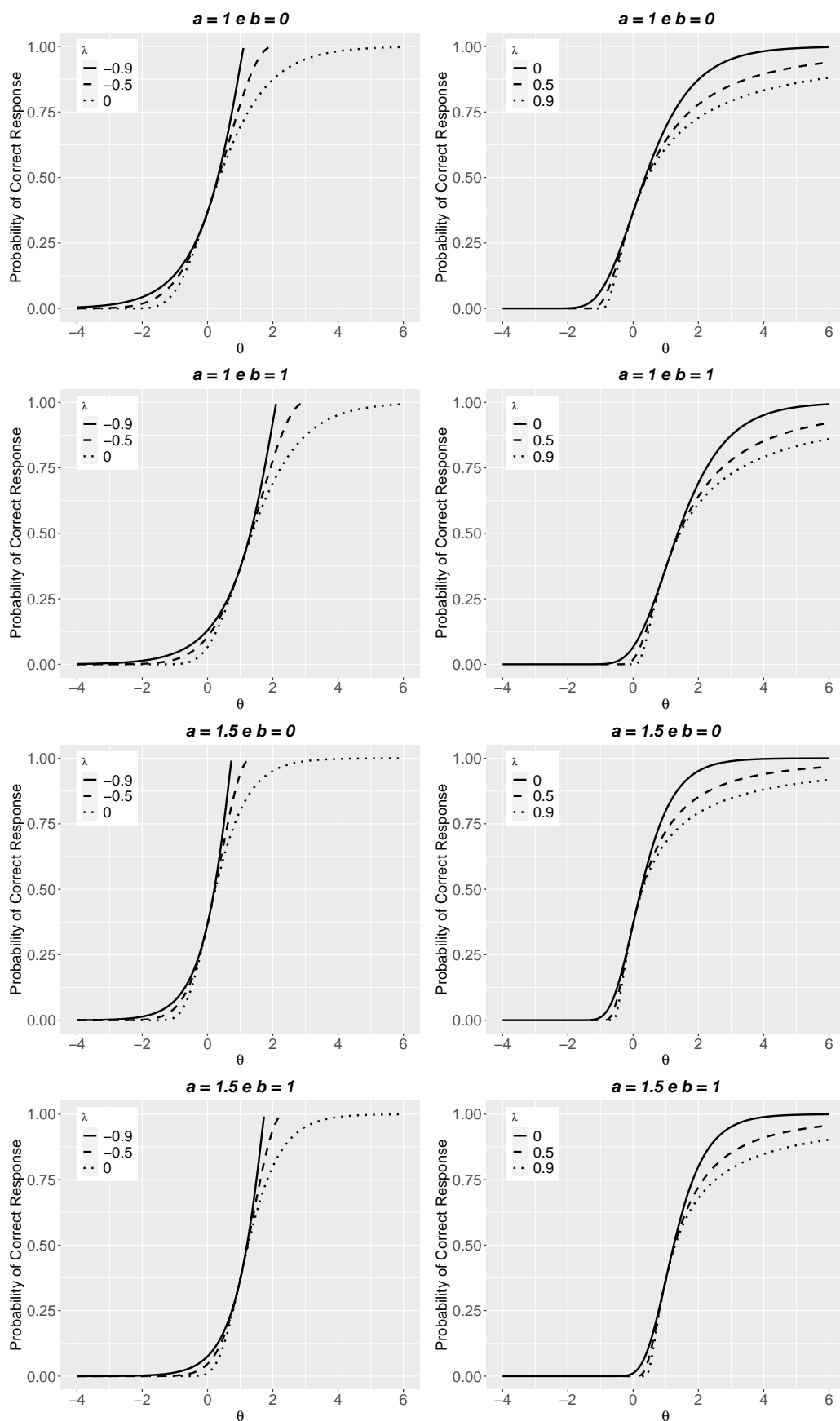


Figura 8 – Curvas característica do item para diferentes  $\lambda$ s no modelo GEV-G TRI considerando diferentes intervalos para  $\theta$ ,  $a$  e  $b$

Fonte: Elaborada pelo autor.

$$L(\beta, \theta | y, X) = \prod_{i=1}^n \prod_{j=1}^k \left\{ F_{\lambda_j}(m_{ij}) \right\}^{y_{ij}} \left\{ 1 - F_{\lambda_j}(m_{ij}) \right\}^{1-y_{ij}}, \quad (4.8)$$

na qual  $\beta = (a^T, b^T)$ ,  $a = (a_1, \dots, a_n)^T$ ,  $b = (b_1, \dots, b_n)^T$ ,  $m_{ij}$  é o preditor linear latente, com  $i = 1, \dots, n$  e  $j = 1, \dots, k$ , e  $F_{\lambda_j}$  é a função de distribuição acumulada (fda) da GEV-G e GEV-RG nas Equações (4.5 e 4.6), indexada com  $\lambda_j$  e avaliada em  $m_{ij}$ .

**Abordagem Bayesiana:** neste trabalho, considera-se especificamente a estimação Bayesiana com métodos de Monte Carlo via Cadeias de Markov (MCMC) que facilitam a amostragem eficiente da distribuição marginal posterior dos parâmetros de interesse. Essa escolha pela estimação Bayesiana foi feita considerando os trabalhos de (SWAMINATHAN *et al.*, 2003), que demonstraram que a estimação precisa dos parâmetros de itens em amostras pequenas é obtida usando a abordagem Bayesiana. Para isso, utiliza-se o pacote `rstan` no software (TEAM, 2009), que emprega o algoritmo No-U-Turn Sampler (NUTS) (HOFFMAN; GELMAN, 2014). (LUO; JIAO, 2018) demonstram, em um estudo comparativo com outros softwares BUGS, que o STAN é consideravelmente mais rápido na estimativa de modelos TRI.

**Especificação da distribuição à *Priori*:** a definição da priori desempenha um papel crucial na análise Bayesiana, especialmente quando se lida com amostras de tamanho reduzido, como discutido por (BOLFARINE; BAZAN, 2010). A seleção cuidadosa de uma distribuição a priori apropriada para o traço latente resolve problemas específicos de identificação (ALBERT; GHOSH, 2000). Na literatura sobre Teoria de Resposta ao Item (TRI), há um acordo geral sobre a definição da priori para o traço latente, tipicamente expressa como  $\theta_i \sim N(0, 1)$ , onde  $i = 1, \dots, n$ .

Opta-se por adotar priores independentes e compartilhados para  $a$ ,  $b$  e  $\lambda$ , permitindo que as correlações sejam determinadas pelos dados, conforme discutido em Bolfarine e Bazan (2010). Em outras palavras, a priori que se emprega pode ser expressa da seguinte maneira:

$$\pi(\theta, a, b, \lambda) = \prod_{i=1}^n \Phi(\theta_i) \prod_{j=1}^k \pi_1(a_j) \pi_2(b_j) \pi_3(\lambda_j) \quad (4.9)$$

em que  $\Phi(\cdot)$  é a função de densidade de probabilidade da distribuição normal padrão e  $\pi_1(\cdot)$ ,  $\pi_2(\cdot)$ ,  $\pi_3(\cdot)$  são as funções de densidade de probabilidade priori para os parâmetros  $a_j$ ,  $b_j$  e  $\lambda_j$ , respectivamente. Dessa forma,  $a_j \sim LN(0, 1)$  e  $b_j \sim N(0, 1)$  (ver, por exemplo, Bolfarine e Bazan (2010)).

Para  $\lambda_j$  realiza-se um estudo de sensibilidade na Seção 4.3.1.

**Crítérios de Comparação de Modelos:** dentro do contexto Bayesiano, existem várias abordagens para comparar modelos alternativos. Neste estudo, focaliza-se no Critério de Informação de Watanabe-Akaike (WAIC), proposto por Watanabe e Opper (2010). Este critério, baseado no

parâmetro de complexidade ( $p_{WAIC}$ ), é definido por  $WAIC = -2 \cdot lppd + 2 \cdot p_D$ , onde  $lppd$  é a densidade logarítmica de predição pontual, e  $p_D$  é o número efetivo de parâmetros do modelo. Além disso, utiliza-se outra métrica denominada 'validação cruzada leave-one-out' (LOO), proposta por Geisser e Eddy (1979). Por fim, adotamos o Critério de Informação de Desvio (DIC) definido por:  $DIC = \overline{D(\theta)} + 0,5 \cdot \text{var}(D(\theta))$ , em que  $\overline{D(\theta)}$  é o desvio médio posterior (BROOKS, 2002). Em todos esses critérios, valores menores indicam um ajuste mais adequado do modelo aos dados observados.

### 4.3 Estudos de Simulação

Nesta seção, são descritos três estudos de simulação. Um deles destina-se a analisar a sensibilidade da escolha da priori para o parâmetro  $\lambda$ , o outro tem como objetivo verificar a recuperação dos modelos propostos, e por fim têm-se um estudo dos critérios de comparação escolhidos para avaliar os modelos.

#### 4.3.1 Análise de Sensibilidade da distribuição à priori para o parâmetro $\lambda$

Para investigar a eficácia das diferentes priors nos modelos TRI propostos, conduziu-se uma análise de sensibilidade. Este estudo considerou quatro priores distintas encontradas na literatura de modelos de regressão linear aplicados aos modelos GEV para o parâmetro  $\lambda$ . As priores examinadas foram:

- a) Priori 1:  $\lambda \sim U(-2, 2)$ , proposta por Alves, Bazán e Arellano-Valle (2022);
- b) Priori 2:  $\lambda \sim U(-1, 1)$ , proposta por Roy e Dey (2014);
- c) Priori 3:  $\lambda \sim N(0, 10)$ , proposta por Bakhshi e Ahmed (2022).
- d) Priori 4:  $\lambda \sim N(0, 100)$ , proposta por Yoon *et al.* (2010).

Para este estudo, utiliza-se os dados de Bolfarine e Bazan (2010), os quais correspondem a um teste de matemática para alunos do quarto ano em escolas primárias rurais no Peru, com respostas de 974 alunos a 18 itens que se qualificam como respostas binárias (corretas ou incorretas).

Assim, ajusta-se as quatro priors mencionadas anteriormente e para a estimativa, utiliza-se o procedimento descrito na Seção 4.2 e o pacote `rstan` para encontrar estimativas com as seguintes especificações: duas cadeias MCMC, 20.000 iterações executadas, 10.000 iterações descartadas, e utilizamos um espaçamento de 10. Para comparar as priors, utilizam-se os critérios de comparação de ajuste de cada modelo descrito na Seção 4.2, o tempo computacional e o erro padrão das estimativas.

A tabela 12 apresenta as estimativas dos parâmetros para os modelos TRI GEV-G 1P, GEV-G 2P, GEV-RG 1P e GEV-RG 2P, considerando os valores do erro padrão das estimativas e a estatística  $\hat{R}$ , com os critérios de comparação de modelos (WAIC, LOO) e o tempo de execução dos algoritmos MCMC (Time).

Tabela 12 – Resultados da análise de sensibilidade dos modelos GEV-G 1P, GEV-G 2P, GEV-RG 1P, GEV-RG 2P para as quatro distribuições de prioris

Models	Parameter	Priori 1			Priori 2			Priori 3			Priori 4		
		mean	se_mean	$\hat{R}$	mean	se_mean	$\hat{R}$	mean	se_mean	$\hat{R}$	mean	se_mean	$\hat{R}$
GEV-G 1P	$b$	-0,231	0,004	1,002	-0,240	0,004	1,004	-0,244	0,010	1,023	-0,239	0,014	1,031
	$\lambda$	-0,069	0,008	1,002	-0,085	0,007	1,003	0,282	0,058	1,026	0,265	0,050	1,039
	$\theta$	0,003	0,012	1,001	0,002	0,012	1,001	0,005	0,032	1,014	0,009	0,033	1,015
	WAIC	17677,048			17731,180			<b>17641,282</b>			17642,504		
	LOO	17713,902			17766,883			<b>17671,440</b>			17671,640		
	Time	10554,236			<b>5940,355</b>			91813,393			128733,389		
GEV-RG 1P	$b$	0,536	0,006	1,007	0,555	0,003	1,000	0,552	0,020	1,096	0,516	0,025	1,185
	$\lambda$	-0,39	0,008	1,004	-0,343	0,004	1,000	-0,716	0,041	1,137	-0,816	0,054	1,133
	$\theta$	-0,016	0,016	1,004	-0,011	0,009	1,000	0,014	0,06	1,062	-0,025	0,061	1,061
	WAIC	17696,680			17743,483			17644,467			<b>17636,543</b>		
	LOO	17727,846			17775,321			17672,135			<b>17664,552</b>		
	Time	22719,445			<b>14001,349</b>			142719,633			92267,699		
GEV-G 2P	$a$	1,008	0,007	1,002	0,944	0,007	1,004	1,048	0,011	1,001	1,058	0,023	1,007
	$b$	-0,224	0,004	1,002	-0,211	0,006	1,004	-0,223	0,004	1,001	-0,225	0,006	1,009
	$\lambda$	-0,004	0,011	1,002	-0,009	0,013	1,006	-0,025	0,013	1,001	-0,025	0,022	1,008
	$\theta$	0,002	0,011	1,001	0,005	0,016	1,002	0,004	0,011	1,001	0,000	0,017	1,003
	WAIC	17614,784			17638,591			<b>17606,932</b>			17608,932		
	LOO	17650,996			17676,565			<b>17643,466</b>			17644,490		
Time	22514,442			<b>6333,258</b>			58592,932			31319,548			
GEV-RG 2P	$a$	0,927	0,013	1,004	0,912	0,012	1,006	0,920	0,014	1,009	0,917	0,012	1,007
	$b$	0,609	0,006	1,004	0,619	0,007	1,004	0,601	0,008	1,006	0,609	0,008	1,005
	$\lambda$	-0,501	0,012	1,006	-0,415	0,011	1,006	-0,532	0,022	1,01	-0,54	0,018	1,008
	$\theta$	-0,009	0,013	1,001	-0,013	0,015	1,002	-0,013	0,019	1,004	-0,011	0,015	1,003
	WAIC	17584,810			17618,922			17584,419			<b>17571,771</b>		
	LOO	17616,469			17651,503			17614,996			<b>17601,920</b>		
Time	32439,363			<b>15938,446</b>			70078,914			87515,436			

Fonte: Elaborada pelo autor.

Observa-se que nos modelos 2P não há sensibilidade na escolha das prioris estudadas, pode-se observar, na Tabela 12, que as estimativas dos parâmetros para as quatro prioris coincidem nos modelos GEV-G 2P e GEV-RG 2P. Já no caso dos modelos 1P pode-se ver sensibilidade na estimação do parâmetro  $\lambda$ , em geral, as prioris 1( $U(-2, 2)$ ) e 2( $U(-1, 1)$ ) as estimativas coincidem, e nas prioris 3( $N(0, 10)$ ) e priori 4( $N(0, 100)$ ) as estimativas também coincidem, porém há uma diferença nas estimativas quando se comparam as prioris 1 e 2 com as prioris 3 e 4.

Para analisar qual seria o ganho dos critérios WAIC e LOO quando se opta pela distribuição à priori 3( $N(0, 10)$ ) com relação as outras distribuições à priori, é necessário analisar o percentual de perda relativo. Para calcular essa percentagem, subtraímos o resultado associado à distribuição da Priori 3 dos resultados das distribuições que estamos avaliando, e então dividimos esse resultado pelo resultado da distribuição à priori 3. Para o modelo GEV-G 1P tem-se que: a distribuição à priori 1 tem um percentual de perda relativo no WAIC de 0,203% e no LOO de 0,240%; a distribuição à priori 2 tem um percentual de perda relativo no WAIC de 0,510% e no LOO de 0,54%; e a distribuição à priori 4 tem um percentual de perda relativo no WAIC de 0,007% e no LOO de 0,001% com relação a distribuição à priori 3. E para o modelo GEV-RG

1P temos: a distribuição à *priori* 1 tem um percentual de perda relativo no WAIC de 0,296% e no LOO de 0,315%; e a distribuição à *priori* 2 tem um percentual de perda relativo no WAIC de 0,561% e no LOO de 0,584%. A diferença de perda relativo entre os critérios WAIC e LOO são menores de 1% para todas as prioris estudadas nos modelos 1P.

No caso dos modelos 2P, observando os critérios WAIC e LOO as distribuições à *priori* para o parâmetro  $\lambda$  que tiveram os menores valores foram a distribuição à *priori* 3( $N(0,10)$ ) para o modelo GEV-G 2P e a distribuição à *priori* 4( $N(0,100)$ ) para o modelo GEV-RG 2P, como analisado anteriormente com os modelos 1P. Porém, comparando o percentual de perda relativo pode-se ver que esse ganho não foi significativo. Analisando o WAIC e LOO para o modelo GEV-G 2P: a distribuição à *priori* 1 tem um percentual de perda relativo no WAIC de 0,045% e no LOO de 0,043% com relação a distribuição à *priori* 3; a distribuição à *priori* 2 tem um percentual de perda relativo no WAIC e no LOO de 0,180% com relação a distribuição à *priori* 3; e a distribuição à *priori* 4 tem um percentual de perda relativo no WAIC de 0,011% e no LOO de 0,006% com relação a distribuição à *priori* 3; E para o modelo GEV-RG 2P temos: a *priori* 1 tem um percentual de perda relativo no WAIC de 0,074% e no LOO de 0,083% com relação a distribuição à *priori* 4; a distribuição à *priori* 2 tem um percentual de perda relativo no WAIC de 0,268% e no LOO de 0,282% com relação a distribuição à *priori* 4, e a distribuição à *priori* 3 tem um percentual de perda relativo no WAIC de 0,072% e no LOO de 0,074% com relação a distribuição à *priori* 4. A diferença de perda relativa entre os critérios WAIC e LOO são menos de 1% para todas as prioris estudadas nos modelos 2P.

Quando se analisa o tempo computacional para todos os modelos de 1P e 2P, observa-se que a distribuição à *priori* 2( $U(-1,1)$ ) possui o menor tempo em todos os modelos. Para o modelo GEV-G 1P o tempo computacional para a distribuição à *priori* 2 é 15 vezes menor que a distribuição à *priori* 3( $N(0,10)$ ), para o modelo GEV-RG 1P a distribuição à *priori* 2 é 10 vezes menor que a distribuição à *priori* 3( $N(0,10)$ ), para o modelo GEV-G 2P o tempo computacional para a distribuição à *priori* 2 é 9 vezes menor que a distribuição à *priori* 3( $N(0,10)$ ) e para o modelo GEV-RG 2P a distribuição à *priori* 2 é 4 vezes menor que a distribuição à *priori* 3( $N(0,10)$ ).

Como análise geral, quando são observados os critérios WAIC e LOO a escolha seria a distribuição à *priori* 3( $N(0,10)$ ), mas o tempo computacional é inviável quando comparada com a distribuição à *priori* 1( $U(-2,2)$ ) e a distribuição à *priori* 2( $U(-1,1)$ ). Dessa forma, analisando as distribuições à *priori* 1 e 2, observa-se que o percentual de perda relativo da distribuição à *priori* 1( $U(-2,2)$ ) é menor que o percentual de perda da distribuição à *priori* 2( $U(-1,1)$ ) para todos os modelos de 1P e 2P e, apesar de o tempo computacional dessa distribuição à *priori* não ser o melhor é comparável com o tempo computacional da distribuição à *priori* 2( $U(-1,1)$ ). Diante disso sugere-se o uso da distribuição à *priori* 1( $U(-2,2)$ ).

### 4.3.2 Estudo de recuperação dos parâmetros

Foi desenvolvido um estudo de simulação com o propósito de avaliar a precisão na recuperação dos parâmetros dos modelos TRI propostos. Foram considerados três diferentes tamanhos de amostra:  $n = 500, 1.000$  e  $2.000$ . Para a geração dos verdadeiros valores dos parâmetros no modelo TRI, adotou-se a seguinte abordagem: os parâmetros de discriminação  $a_j$  com  $j = 1, \dots, 20$  foram gerados a partir de uma distribuição  $U(0, 4, 2)$ , enquanto o parâmetro de dificuldade  $b_j$  e o de forma  $\lambda_j$ , com  $j = 1, \dots, 20$  foram extraídos de uma distribuição  $N(0, 1)$ . Além disso, a habilidade latente  $\theta_i$  foi gerada também com base em uma distribuição  $N(0, 1)$  para  $i = 1, \dots, n$ . Dessa forma geramos os dados para o modelo TRI GEV-G 2P e o modelo TRI GEV-RG 2P.

As distribuições *à priori* consideradas para os parâmetros foram aquelas que demonstraram os melhores resultados no estudo de simulação 4.3.1: para o parâmetro de discriminação  $a_j$ , foi adotada a  $LN(0, 1)$ ; para o parâmetro de dificuldade  $b_j$ , a distribuição  $N(0, 1)$ ; para o parâmetro  $\lambda_j$ , considerou-se a distribuição  $U(-2, 2)$ ; e, por fim, para o parâmetro de habilidade  $\theta_i$ , foi utilizada a distribuição *à priori*  $N(0, 1)$ .

Ajusta-se dessa maneira, os modelos TRI GEV-G 2P e GEV-RG 2P, dados pelas equações 4.5 e 4.6, respectivamente. Para isso, utiliza-se o procedimento descrito na Seção 4.2 e o pacote `rstan` para obter estimativas com as seguintes especificações: 2 cadeias MCMC, 20,000 iterações executadas com descarte de 10,000 iterações, e um espaçamento de 10. Para este estudo foram utilizadas  $M = 100$  replicações.

Para avaliar a precisão das estatísticas, utiliza-se 5 critérios: desvio padrão (SD), viés, erro quadrático médio (RMSE), a Probabilidade de Cobertura (PC) dos intervalos de confiança de 95%, e  $\hat{R}$ . A média do desvio padrão posterior é definida como:  $SD(\hat{\eta}) = \frac{1}{M} \sum_{m=1}^M SD^{(m)}(\eta)$  e o viés foi definido como:  $Vis(\hat{\eta}) = \frac{1}{M} \sum_{m=1}^M (\hat{\eta}^{(m)} - \eta)$  onde  $\hat{\eta}^{(m)}$  denota a média posterior e  $SD^{(m)}(\eta)$  representa o desvio padrão posterior de  $\eta$  obtido no  $m$ -ésimo conjunto de dados para  $m = 1, \dots, M$ .

O RMSE para o parâmetro  $\eta$  é raiz quadrada da variância amostral das estimativas posteriores ao longo diferentes conjuntos de dados simulados fornecidos por

$RMSE(\hat{\eta}) = \sqrt{\frac{1}{M} \sum_{m=1}^M (\hat{\eta}^{(m)} - \frac{1}{M} \sum_{i=1}^M \hat{\eta}^{(i)})^2}$ . Finalmente, a probabilidade de cobertura é definida como:  $PC(\hat{\eta}) = \frac{\text{\#de intervalos HPD de 95\% contendo } \eta, \text{ em } M \text{ simulações}}{M}$ .

A Tabela 13 apresenta os 5 critérios utilizados na análise da recuperação dos parâmetros -  $\theta$ ,  $b$ ,  $a$  e  $\lambda$ - para o modelo TRI GEV-G 2P. Foi considerada a média para cada conjunto de parâmetros. Conforme o número de indivíduos aumenta, observa-se uma tendência decrescente no Desvio Padrão (SD), Viés e o Erro Quadrático Médio (RMSE) para esses parâmetros.

Ao comparar o aumento no número de indivíduos de 500 para 2.000, observam-se reduções nos valores dos parâmetros: o desvio padrão (SD) diminui de 0,306 para 0,290 e o

erro quadrático médio (RMSE) de 0,260 para 0,248 para o parâmetro  $\theta$ . Para o parâmetro  $b$ , o SD é reduzido pela metade, de 0,119 para 0,066, e o RMSE diminui de 0,129 para 0,111. Um padrão semelhante é observado na redução do SD e RMSE para o parâmetro  $a$ , diminuindo pela metade de 0,286 para 0,151 e de 0,782 para 0,347, respectivamente. Em relação ao parâmetro de  $\lambda$ , o SD e o RMSE diminuem de 0,338 para 0,225 e de 0,923 para 0,726, respectivamente.

Um ponto a ser destacado é a probabilidade de cobertura, variando entre 90% e 95% para todos os parâmetros. Isso sugere que, em pelo menos 90% das análises, o valor verdadeiro está contido no intervalo de credibilidade. Além disso, independentemente do tamanho da amostra, foi observada a convergência das cadeias com  $\hat{R}$  próximo de 1,00. Esse resultado sugere que, apesar da amostra relativamente pequena, considerando a quantidade de parâmetros a serem estimados, o modelo TRI GEV-G 2P apresenta uma boa recuperação dos parâmetros.

Tabela 13 – Estimativas de parâmetros para o modelo TRI GEV-G 2P com dados simulados desbalanceados

	n = 500					n = 1000					n = 2000				
	Sd	Viés	RMSE	PC	$\hat{R}$	Sd	Viés	RMSE	PC	$\hat{R}$	Sd	Viés	RMSE	PC	$\hat{R}$
$\theta$	0,306	-0,032	0,260	0,941	1,011	0,297	-0,022	0,302	0,939	1,009	0,290	-0,028	0,248	0,941	1,008
$b$	0,119	-0,044	0,129	0,962	1,027	0,089	-0,022	0,145	0,958	1,027	0,066	-0,026	0,111	0,950	1,021
$a$	0,286	0,088	0,782	0,953	1,020	0,218	0,076	0,752	0,956	1,018	0,151	0,029	0,347	0,942	1,016
$\lambda$	0,338	0,021	0,923	0,907	1,018	0,287	0,001	0,839	0,904	1,020	0,225	0,016	0,726	0,895	1,019

Fonte: Elaborada pelo autor.

A Tabela 14 apresenta as métricas usadas para avaliar a recuperação do modelo TRI GEV-RG 2P para três tamanhos diferentes de amostras. Da mesma forma que na tabela anterior, observamos que à medida que o número de indivíduos na amostra aumenta, ocorre uma redução nos valores de Desvio Padrão (SD), Viés e Raiz do Erro Quadrático Médio (RMSE) para os parâmetros do modelo TRI GEV-RG 2P.

Ao comparar os resultados entre amostras de 500 e 2,000 indivíduos, nota-se reduções nos valores dos parâmetros. Para o parâmetro de  $\theta$ , houve uma diminuição no SD de 0,318 para 0,300, enquanto o RMSE manteve-se constante, mantendo o valor de 0,294 para ambos os tamanhos de amostra. Já o parâmetro de  $b$  mostra uma redução significativa, com o SD e o RMSE diminuindo pela metade, respectivamente, de 0,121 para 0,063 e de 0,097 para 0,057. Essa mesma tendência foi evidenciada na redução do SD e RMSE para o parâmetro de  $a$ , diminuindo pela metade de 0,279 para 0,133 e de 0,865 para 0,217, respectivamente. No caso do parâmetro  $\lambda$ , observou-se uma queda do SD de 0,339 para 0,220 e do RMSE de 0,238 para 0,083, aproximando-se de zero, o que sugere uma boa recuperação desse parâmetro.

Assim como no caso do modelo TRI GEV-G 2P, a variação da probabilidade de cobertura permaneceu entre 89% e 95% para todos os parâmetros. Além disso, independentemente do tamanho da amostra, foi observada a convergência das cadeias com  $\hat{R}$  próximo de 1,00.

Conclui-se, portanto, que tanto o modelo TRI GEV-G 2P quanto o GEV-RG 2P demonstraram uma boa recuperação dos parâmetros, apresentando resultados consistentes independente-



mente do tamanho da amostra considerado.

Tabela 14 – Estimativas de parâmetros para o modelo TRI GEV-RG 2P com dados simulados desbalanceados

	n = 500					n = 1000					n = 2000				
	Sd	Viés	RMSE	PC	$\hat{R}$	Sd	Viés	RMSE	PC	$\hat{R}$	Sd	Viés	RMSE	PC	$\hat{R}$
$\theta$	0,318	-0,032	0,294	0,950	1,010	0,306	-0,023	0,299	0,944	1,005	0,300	-0,028	0,299	0,943	1,003
$b$	0,121	-0,032	0,097	0,962	1,021	0,089	-0,016	0,076	0,959	1,013	0,063	-0,022	0,057	0,944	1,009
$a$	0,279	0,064	0,865	0,965	1,017	0,203	0,055	0,495	0,961	1,009	0,133	0,022	0,217	0,953	1,009
$\lambda$	0,339	0,092	0,238	0,917	1,014	0,278	0,080	0,139	0,907	1,010	0,220	0,050	0,083	0,896	1,008

Fonte: Elaborada pelo autor.

### 4.3.3 Estudos dos Critérios de Comparação de Modelos

Realiza-se um estudo comparativo entre os modelos TRI GEV-G 2P e GEV-RG 2P, com o objetivo de avaliar esses modelos em relação ao modelo logístico de dois parâmetros (2L). Utiliza-se a mesma estratégia de geração de dados empregada no estudo de recuperação dos parâmetros. Para três diferentes tamanhos de amostra ( $N = 500, 1.000, 2.000$ ), gera-se os dados para o modelo TRI GEV-G 2P e ajustam-se tanto o modelo TRI GEV-G 2P quanto o modelo 2L. Para este estudo consideram-se 100 replicações. Os parâmetros considerados para essa análise foram os seguintes:  $a_j \sim U(0, 4, 2)$ ,  $b_j \sim N(0, 1)$  e  $\lambda_j \sim N(0, 1)$ , com  $j = 1, \dots, 20$ ; e  $\theta_i \sim N(0, 1)$ , para  $i = 1, \dots, N$ , como realizado no estudo anterior. As distribuições a priori consideradas para os parâmetros foram as consideradas no estudo de recuperação. Uma abordagem análoga foi utilizada para a geração dos valores do modelo TRI GEV-RG 2P.

Utiliza-se os critérios DIC, WAIC e LOO conforme o procedimento descrito na Seção 4.2 e o pacote `rstan` para obter estimativas com as seguintes especificações: 2 cadeias MCMC, 20.000 iterações executadas com descarte de 10.000 iterações, espaçamento de 10, foram realizadas 100 réplicas. A convergência das cadeias foi verificada pela estatística  $\hat{R}$  de (GELMAN; RUBIN *et al.*, 1992).

A Tabela 15 mostra os resultados desse estudo para os três tamanhos de amostra. É perceptível que o modelo verdadeiro é selecionado em todas as réplicas e para todos os parâmetros, totalizando uma taxa de 100%. Além disso, a média dos critérios foi calculada a partir de 100 amostras, revelando que o modelo verdadeiro sempre oferece um ajuste superior em comparação com o modelo 2L.

Os resultados da análise demonstram que os modelos propostos se mostraram mais apropriados em relação ao modelo 2L para a representação do verdadeiro modelo TRI. Isso reforça a robustez e a eficácia dos modelos proposto ao oferecer um ajuste mais consistente e preciso, evidenciando sua capacidade de descrever, com maior fidelidade, os dados observados, independentemente do tamanho da amostra considerado.

Além disso, os critérios de comparação DIC, WAIC e LOO demonstraram ser eficazes na avaliação e seleção dos modelos propostos, fornecendo uma análise detalhada e abrangente



Tabela 15 – Desempenho dos critérios de ajuste para comparação de modelos

n	Modelo Verdadeiro							
	GEV-G 2P				GEV-RG 2P			
	Modelo Ajustado	Crítérios	%	Média	Modelo Ajustado	Crítérios	%	Média
500	2L	DIC	0	11012,1	2L	DIC	0	8178,8
	2L	WAIC	0	12274,1	2L	WAIC	0	8968,9
	2L	LOO	0	12315,2	2L	LOO	0	8984,6
	GEV-G 2P	DIC	100	10576,7	GEV-RG 2P	DIC	100	7872,9
	GEV-G 2P	WAIC	100	11793,0	GEV-RG 2P	WAIC	100	8683,5
	GEV-G 2P	LOO	100	11841,5	GEV-RG 2P	LOO	100	8715,1
1000	2L	DIC	0	14587,8	2L	DIC	0	16339,6
	2L	WAIC	0	16213,8	2L	WAIC	0	17843,8
	2L	LOO	0	16263,5	2L	LOO	0	17869,6
	GEV-G 2P	DIC	100	14024,3	GEV-RG 2P	DIC	100	15686,6
	GEV-G 2P	WAIC	100	15601,9	GEV-RG 2P	WAIC	100	17268,0
	GEV-G 2P	LOO	100	15660,8	GEV-RG 2P	LOO	100	17327,2
2000	2L	DIC	0	35754,2	2L	DIC	0	35533,6
	2L	WAIC	0	35850,2	2L	WAIC	0	35580,3
	2L	LOO	0	31088,3	2L	LOO	0	31289,6
	GEV-G 2P	DIC	100	34476,1	GEV-RG 2P	DIC	100	34421,4
	GEV-G 2P	WAIC	100	34598,2	GEV-RG 2P	WAIC	100	34537,6
	GEV-G 2P	LOO	100	35754,2	GEV-RG 2P	LOO	100	35533,6

Fonte: Elaborada pelo autor.

da adequação dos mesmos em relação aos dados gerados.

## 4.4 Aplicações

Nesta seção, foram empregadas duas aplicações práticas de dados reais para demonstrar a eficácia dos modelos propostos. Uma delas foi realizada utilizando dados do teste de matemática previamente utilizados no capítulo anterior, enquanto a segunda aplicação foi baseada em dados relacionados à infecção pulmonar. Esses conjuntos de dados reais foram escolhidos intencionalmente para avaliar e ilustrar o desempenho dos modelos propostos em contextos distintos e relevantes.

### 4.4.1 Teste de Matemática

Neste estudo, utiliza-se o conjunto de dados apresentado por [Bolfarine e Bazan \(2010\)](#) para ilustrar as abordagens bayesianas propostas na análise de modelos TRI com respostas binárias. Esse conjunto de dados corresponde a um teste de matemática administrado a 974 alunos do quarto ano em escolas primárias rurais no Peru. O teste consiste em 18 itens classificados como binários, os quais abordavam distintas habilidades e respondidos independentemente pelos alunos. Considerando a habilidade latente  $\theta$  dos alunos, suas respostas aos itens também foram tratadas como independentes. Esse conjunto de dados foi o mesmo usado no estudo de simulação a priori da Seção 4.3.1

Mostra-se, nessa aplicação, o ajuste dos modelos paramétricos da TRI discutidos, anteriormente, na seção 4.1 usando os dados do teste matemático. Para comparar os modelos propostos, são considerados os modelos logísticos de TRI 1L, 2L, RLPE 1P, RLPE 2P propostos

por [Bolfarine e Bazan \(2010\)](#) e os modelos 1PGLogit e 2PGLogit propostos por [Zhang et al. \(2022\)](#) e também os modelos propostos no capítulo 3. Em todos os casos, as distribuições anteriores utilizadas são  $\theta \sim N(0, 1)$ ,  $b \sim N(0, 1)$ ,  $a \sim LN(0, 1)$ . Para  $\lambda$ , considera-se a priori  $U(-2, 2)$  para os modelos propostos. O código *Stan* está disponível no Apêndice C.

Para a análise utiliza-se o software R, e o algoritmo NUTS que está implementado no pacote *rstan*. Os parâmetros MCMC utilizados foram 5.000 iterações, espaçamento de 5, e descartamos as 2.000 primeiras iterações para alcançar a convergência usando 2 cadeias MCMC. Consideramos a verificação de convergência baseada na estatística ( $\hat{R}$ ) ([GELMAN; RUBIN et al., 1992](#)). Para comparar os modelos, usam-se os critérios de seleção DIC, WAIC e LOO discutidos na Seção 4.2. Esses valores são mostrados na Tabela 16.

Tabela 16 – Critérios de comparação de modelos para os dados do teste de matemática

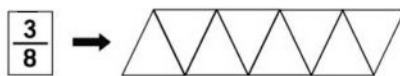
	WAIC	LOO	Tempo Relativo
L 1P	17968,508	17977,448	1,0
L 2P	17750,767	17767,340	1,5
RLPE 1P	17835,716	17851,014	5,0
RLPE 2P	17708,040	17733,863	9,8
RP-RG 1P	17721,592	17750,574	2,4
RP-RG 2P	17694,963	17727,336	6,0
GLogit 1P	17612,545	17672,829	1,9
GLogit 2P	17588,262	17648,337	2,6
GEV-G 1P	17677,006	17712,047	2,1
GEV-G 2P	17609,283	17643,411	4,0
GEV-RG 1P	17700,392	17729,633	4,8
<b>GEV-RG 2P</b>	<b>17583,697</b>	<b>17612,398</b>	2,9

Fonte: Elaborada pelo autor.

Os resultados apresentados na Tabela 16 indicam que, com base nos critérios de avaliação WAIC e LOO, o modelo GEV-RG 2P se destaca como o mais adequado entre os modelos propostos. Além disso, ao classificar os valores dos critérios de comparação, os modelos Glogit 2P e GEV-G 2P também demonstram desempenho superior em relação aos demais modelos. Consequentemente, espera-se que as estimativas das CCIs sejam mais precisas ao utilizar o modelo assimétrico proposto na TRI. Com relação ao tempo relativo, o modelo TRI GEV-RG 2P é 2,9 vezes maior que o modelo TRI 1L, o que pode ser justificado pela quantidade de parâmetros maior a ser estimado, e relativamente inferior aos outros modelos propostos por [Bolfarine e Bazan \(2010\)](#).

Neste estágio, opta-se por realizar uma comparação entre os modelos TRI GEV-G 2P, GEV-RG 2P propostos e RLPE que foi previamente examinados por [Bolfarine e Bazan \(2010\)](#) durante a análise dos dados do Teste de Matemática. Dessa forma, segue-se a mesma ideia empregada no capítulo 3. O foco de análise é direcionado para três itens específicos: Item 14, Item 2 e Item 11, os mesmos itens previamente estudados por [Bolfarine e Bazan \(2010\)](#). De acordo com o modelo RLPE 2P, o Item 14 é categorizado como assimétrico, enquanto o Item 2 exibe algum grau de assimetria e o Item 11 não apresenta qualquer assimetria (para mais detalhes, consulte [Bolfarine e Bazan \(2010\)](#)). A Figura 9 ilustra os itens selecionados.

Item 14. You should depict the fraction in the figure (**Asymmetric Item**)



Item 2. Look closely at these number (**Medium asymmetric Item**)

**14**   **89**   **700**   **302**

**25**   **106**   **41**   **99**

Mark with an ("X") all numbers **LESS THAN 100**

Item 11. Complete the sum (**Symmetric Item**)

$$300 + \boxed{\phantom{000}} = 350$$

Figura 9 – Itens 14, 2 e 11 para o conjunto de dados do teste de matemática

Fonte: [Bolfarine e Bazan \(2010\)](#).

A Tabela 17 mostra os parâmetros dos itens para os modelos de TRI alternativos que apresentaram melhor desempenho quando comparados ao RLPE 2P.

Comparando os modelos GEV-G 2P e GEV-RG 2P pode-se observar, na Tabela 17, que houve uma mudança de sinal no parâmetro  $\lambda$  no item 14 e item 2, isso é esperado pelo fato de serem modelos reversos. Com relação ao parâmetro  $a$  a magnitude são próximas para o caso simétrico e com alguma assimetria, e são mais divergentes no caso em que a assimetria não é evidenciada. Já no parâmetro  $b$ , o sinal está contrário nos itens com alguma assimetria e simétrico para esses modelos, sendo que o modelo TRI GEV-RG 2P possui o mesmo sinal que o modelo TRI RLPE 2P. Portanto, pode-se concluir que este modelo TRI GEV-RG 2P permite uma interpretação mais parecida com o modelo TRI RLPE 2P e, mediante a Tabela 16, mostrou-se o modelo mais apropriado.

Agora, comparando os modelos TRI GEV-RG 2P e RLPE 2P em relação ao parâmetro  $\lambda$ , observa-se magnitude e sinais diferentes, no caso assimétrico, e no caso com alguma assimetria temos  $\lambda = 0,08$ , ou seja, próximo de zero, isso significa que esse item possui assimetria e essa assimetria é próxima ao modelo TRI Gumbel (que é um caso especial do GEV-RG quando tem-se o parâmetro lambda igual a zero). Agora observa-se que, no caso simétrico  $\lambda = -0,38$  para o modelo TRI GEV-RG 2P, isto é esperado para este modelo, pois se sabe que ele não detecta itens simétricos. Em relação aos parâmetros de dificuldade e discriminação observa-se sinais iguais, porém magnitudes diferentes em relação a esses dois modelos TRI GEV-RG 2P e

RLPE 2P, os quais serão explorados futuramente.

Tabela 17 – Parâmetros dos itens 14, 2 e 11 para modelos alternativos de TRI nos dados de matemática

Itens	Modelos	Parâmetros					
		$a$	$\hat{R}_a$	$b$	$\hat{R}_b$	$\lambda$	$\hat{R}_\lambda$
Item 14 (Assimétrico)	RPLE	4,71	1,00	-1,62	1,00	0,20	1,00
	GEV-G 2P	1,36	1,01	-1,00	1,00	1,08	1,01
	GEV-RG 2P	0,47	1,01	-0,03	1,02	-1,46	1,02
Item 02 (Com algum grau de assimetria)	RPLE 2P	1,06	1,00	0,38	1,00	2,12	1,00
	GEV-G	0,63	1,01	-0,94	1,01	-0,54	1,04
	GEV-RG 2P	0,79	1,00	0,08	1,01	0,08	1,01
Item 11 (Assimetria não evidente)	RPLE	2,01	1,00	0,12	1,00	1,25	1,00
	GEV-G 2P	1,13	1,00	-0,25	1,01	-0,22	1,00
	GEV-RG 2P	1,06	1,01	0,37	1,01	-0,38	1,01

Fonte: Elaborada pelo autor.

Pode-se observar, na Figura 10 que, para o item 14, o ICC do modelo GEV-RG mostra que se o indivíduo não possui habilidade ou esta habilidade é pequena ( $\theta < -1,6$ ), a probabilidade de acertar o item é praticamente zero, mas se ele tiver algum nível de habilidade ( $-1,6 < \theta < 1,6$ ) a probabilidade de acertar o item aumenta significativamente. E se essa habilidade for maior que 2, a probabilidade de acertar o item é praticamente 1. Observe que o GEV-RG 2P e o RLPE 2P mostram que esse item é assimétrico.

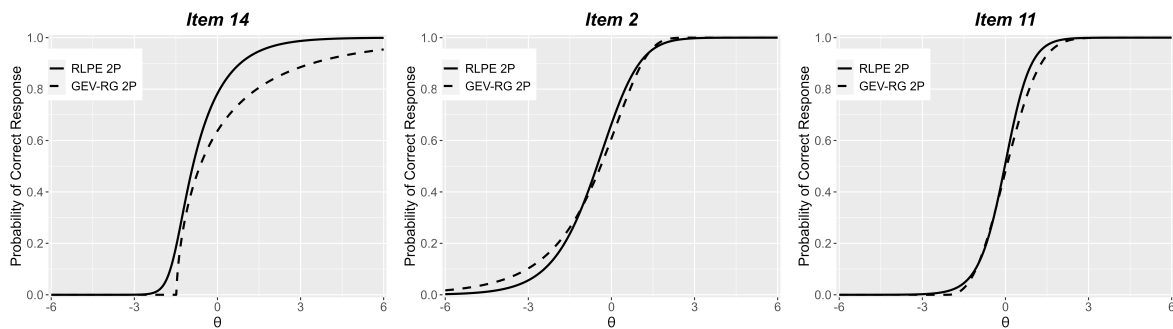


Figura 10 – Curvas características do item (ICCs) para os itens 14, 2 e 11 sob os modelos de TRI GEV-RG 2P e RLPE 2P nos dados de matemática

Fonte: Elaborada pelo autor.

No que se refere ao item 2, é importante notar que a possibilidade de acertá-lo, mesmo que alguém tenha uma habilidade muito baixa ( $\theta < -3$ ), não é nula. Por outro lado, se a habilidade for maior do que 1, a probabilidade de acertar o item é praticamente 1. Portanto, pode-se concluir que este item não é complexo.

É importante observar que para o item 11, os ICCs dos modelos RLPE 2P e GEV-RG 2P são muito próximos, uma vez que  $\lambda = 1,22$  para o modelo RLPE, portanto, a curva é simétrica. Isso porque, quando o parâmetro  $\lambda$  se aproxima de zero o modelo RLPE aproxima-se do modelo 2L. Quanto ao modelo GEV-RG 2P, tem-se  $\lambda = -0,36$ , o que também se aproxima de zero. Pode-se concluir que o item 11 é um item simétrico, ou seja, o indivíduo não precisa ter sub-processos sequenciais para acertar, portanto, a resposta é considerada simétrica.

A Figura 11 mostra a relação entre os parâmetros  $a$ ,  $b$  e  $\lambda$  para o modelo TRI GEV-RG 2P. Observa-se que existe alguma relação entre os parâmetros  $b$  e o parâmetro  $\lambda$ , o que indica que esses dois parâmetros podem estar relacionados. Os outros parâmetros aparentemente não apresentam relação entre si.

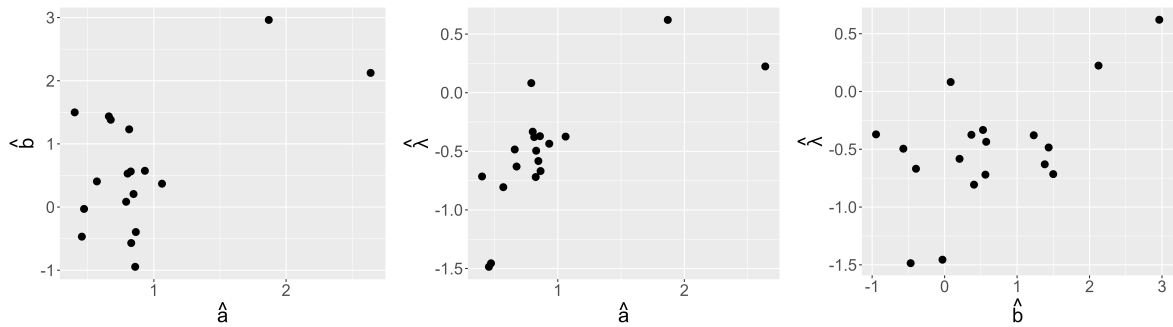


Figura 11 – Relação entre os parâmetros  $\hat{a}$ ,  $\hat{b}$  e  $\hat{\lambda}$  para o modelo TRI GEV-RG 2P.

Fonte: Elaborada pelo autor.

Para comparar os modelos TRI GEV-RG 2P e RLPE 2P, pode-se observar a distribuição das estimativas do parâmetro de habilidade ( $\hat{\theta}$ ) de ambos os modelos dada pela Figura 12. Embora as distribuições das estimativas dos parâmetros de habilidade sejam muito próximas, pode-se observar, no modelo TRI GEV-RG 2P, uma pequena assimetria na cauda direita, já no modelo RLPE 2P essa assimetria é menos evidente. Em resumo, o valor médio posterior do parâmetro habilidade para o modelo GEV-RG TRI é igual a  $-0,012173$ , onde o valor mínimo é igual a  $-2,145247$  e o valor máximo é igual a  $2,530677$ . Para o modelo TRI RLPE, o valor médio posterior do parâmetro habilidade igual a  $0,00534$ , valor mínimo igual a  $-2,07524$  e valor máximo igual a  $2,29141$ .

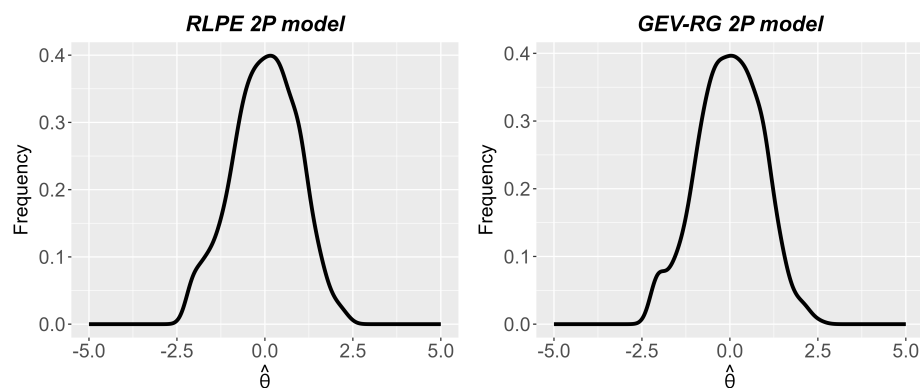


Figura 12 – Distribuições das estimativas dos parâmetros ( $\hat{\theta}$ ) para os modelos TRI GEV-RG 2P e RLPE 2P

Fonte: Elaborada pelo autor.

### 4.4.2 LSAT

Para essa aplicação usa-se o banco de dados LSAT que faz parte do pacote ltm (Latent Trait Models) em R, para ilustrar a performance dos modelos que estamos propondo discutidos na seção 4.1 versus o que já estão na literatura (1L, 2L, RLPE 1P, RLPE 2P, RP-RG 1P, RP-RG 2P, 1PGLogit e 2PGLogit)(Bolfarine e Bazan (2010); Alves e Bazán (2022); Zhang et al. (2022)). Esse banco de dados contém 1000 observações (candidatos a escolas de direito nos Estados Unidos) e 5 variáveis (itens do teste LSAT), todas medidas em escala nominal (0 ou 1 para cada resposta correta ou incorreta, respectivamente). Assim como na aplicação do teste de matemática, as distribuições anteriores utilizadas para as variáveis  $\theta \sim N(0, 1)$ ,  $b \sim N(0, 1)$  e  $a \sim LN(0, 1)$ . Para a variável  $\lambda$ , consideramos uma distribuição à priori uniforme  $U(-2, 2)$  para os modelos propostos.

Como parâmetros das cadeias MCMC, considera-se: 20.000 iterações com 10.000 iterações de descarte, aliadas a um intervalo de espaçamento de 10 iterações para alcançar a convergência usando duas cadeias MCMC. Como estatística para medir a convergência das cadeias MCMC foi utilizado o  $(\hat{R})$  proposto por Gelman, Rubin et al. (1992). Para comparar os modelos, usam-se os critérios de seleção DIC, WAIC e LOO discutidos na Seção 4.2. Esses valores são mostrados na Tabela 18. Todas as análises foram feitas no software R, utilizando o algoritmo NUTS que está implementado no pacote rstan.

Tabela 18 – Critérios de comparação de modelos para os dados do LSAT

MODELOS	WAIC	LOO	Tempo(minutos)
L 1P	4908,369	4919,161	<b>9,694</b>
L 2P	4904,152	4912,184	17,694
RLPE 1P	4908,203	4919,676	36,621
RLPE 2P	4901,887	4913,504	65,115
RP-RG 1P	4905,460	4914,827	33,044
RP-RG 2P	4898,637	4909,118	34,911
GLogit 1P	4838,893	4848,601	12,818
GLogit 2P	4841,758	4852,919	21,215
GEV-G 1P	4873,481	4884,064	18,4778
GEV-G 2P	4845,166	4854,457	41,569
<b>GEV-RG 1P</b>	<b>4560,170</b>	<b>4568,005</b>	23,180
GEV-RG 2P	4822,128	4832,897	24,406

Fonte: Elaborada pelo autor.

Mediante o exame da Tabela 18, é possível perceber que o modelo GEV-RG 1P se destaca como o mais adequado para esses dados, levando em consideração os critérios de comparação WAIC e LOO. Observando o *ranking* dos modelos com melhor desempenho, constata-se que o modelo GEV-G 2P e o modelo GEV-RG 2P ocupam as posições de segundo e terceiro lugares, respectivamente, para esses dados referentes ao LSAT. É importante ressaltar que a escolha do modelo mais apropriado depende dos critérios de comparação adotados e que a análise de dados

deve ser realizada com cautela, a fim de assegurar a confiabilidade dos resultados obtidos.

## 4.5 Comentários Finais

Com base na análise realizada, podemos inferir que as generalizações de funções fundamentadas na ICC *cloglog* surgem como uma alternativa promissora à tradicional Teoria da Resposta ao Item (TRI). Além disso, o procedimento de estimação bayesiana elaborado para os modelos TRI demonstrou eficácia na obtenção de ajustes adequados aos dados.

Os estudos de sensibilidade a priori (ver Seção 4.3.1) evidenciaram que a distribuição a priori  $U(-2, 2)$  se mostra a mais apropriada para o parâmetro adicional  $\lambda$ , associado à assimetria, em todos os modelos propostos, tanto os de 1P quanto os de 2P. Adicionalmente, ressaltou-se a relevância de empregar uma ICC flexível para capturar a assimetria da variável latente, o que pode aprimorar tanto o ajuste dos modelos quanto a interpretação dos resultados.

Com o estudo da recuperação dos parâmetros (ver Seção 4.3.2), foi possível verificar que os modelos propostos demonstraram uma robusta capacidade de recuperação dos parâmetros. Os resultados obtidos foram consistentes e estáveis, independentemente do tamanho da amostra considerado. Isso sugere que esses modelos são confiáveis e adequados para a análise de dados, mesmo em cenários com variações no número de observações.

Os resultados do estudo dos critérios de comparação (ver Seção 4.3.3), evidenciam consistentemente a seleção do modelo verdadeiro em todas as réplicas e para todos os parâmetros, alcançando uma taxa de 100%.

Ao aplicar os modelos aos dados do Teste de Matemática (conforme detalhado na Seção 4.4.1), observou-se que o modelo TRI GEV-RG 2P apresentou um ajuste superior em comparação com outras abordagens TRI encontradas na literatura. Da mesma forma, na análise dos dados do pacote *1tm* (ver Seção 4.4.2), os modelos propostos destacaram-se pelo desempenho superior em relação aos modelos existentes.

Este estudo abre caminho para diversas extensões em pesquisas futuras. Por exemplo, uma investigação mais aprofundada sobre a identificabilidade dos modelos propostos e uma exploração mais detalhada da interpretação do parâmetro adicional associado à assimetria.





---

# COMENTÁRIOS FINAIS E DESENVOLVIMENTOS FUTUROS

---

Neste capítulo, são apresentados comentários finais referentes aos resultados principais e às contribuições deste trabalho. Também são descritas as produções resultantes desta pesquisa. Por fim, são feitos comentários gerais sobre possíveis desenvolvimentos futuros.

## 5.1 Conclusões

Nesta pesquisa, exploramos abordagens alternativas para a classificação de dados desbalanceados, focalizando modelos de regressão binomial e teoria de resposta ao item em cenários distintos. Em ambas as situações, examinamos o uso de uma nova classe de funções de ligação assimétrica, baseada nas distribuições Gumbel e sua versão reversa. Essas funções surgem como alternativas eficazes às convencionais, como logística e normal, nos contextos estudados, demonstrando um desempenho notável em diversas situações e instigando debates sobre sua aplicabilidade em diferentes áreas do conhecimento.

A relevância das funções de ligação assimétricas para dados desbalanceados não é uma descoberta recente, conforme evidenciado no capítulo 2. Em face do desbalanceamento, as ligações assimétricas propostas mostraram-se apropriadas, constituindo uma valiosa contribuição para a literatura e estimulando reflexões para estudos futuros.

Há potenciais extensões dos métodos desenvolvidos em pesquisas futuras, como a exploração de modelos mistos, conforme investigado por (HUAYANAY *et al.*, 2019) ao estudar a família de ligação potencia para esse contexto. Além disso, sugere-se a realização de estudos adicionais, incluindo análises de simetria e curtose das distribuições propostas neste trabalho.

No terceiro e quarto capítulo, propusemos adaptações das CCIs, baseando-nos na CCI *cloglog*, como alternativa aos modelos tradicionais na Teoria de Resposta ao Item. Desenvolve-

mos um método de estimação Bayesiana para esses modelos TRI, ressaltando a importância do uso da CCI flexível para abordar a assimetria da variável latente.

Os modelos propostos não só são uma alternativa viável aos métodos tradicionais, mas também oferecem uma ferramenta valiosa e eficiente para análise de dados, capturando a assimetria presente. Esta capacidade é fundamental em diversos campos, como ciências sociais, economia e medicina, onde dados frequentemente apresentam distribuições assimétricas que podem afetar a interpretação e análise dos resultados.

Ao proporcionar uma abordagem flexível para modelar a assimetria da variável latente, os modelos propostos têm potencial para melhorar a precisão e a interpretação das análises estatísticas. Isso beneficia não apenas a avaliação educacional, mas também uma variedade de outras áreas.

Este trabalho abre um vasto leque de possibilidades para futuras extensões, destacando-se a aplicação dessas curvas características a modelos politômicos, o que representaria um avanço significativo na modelagem estatística. Em todas essas situações, a estimação por máxima verossimilhança pode desempenhar um papel crucial como uma abordagem complementar, oferecendo uma validação robusta e independente dos resultados obtidos através da estimação bayesiana.

Além disso, aprimorar a estimação bayesiana para torná-la mais rápida utilizando outros métodos e software é uma direção promissora. Essa integração de métodos reforça a confiabilidade e a precisão dos modelos propostos, ampliando ainda mais seu potencial de aplicação em diversos campos de estudo.

## 5.2 Produções

Alves, J. S., Bazán, J. L., & Arellano-Valle, R. B. (2022). Flexible cloglog links for binomial regression models as an alternative for imbalanced medical data. *Biometrical Journal*, 65(3), 2100325.

Alves, J. S. B., & Bazán, J. L. (2022). New Flexible Item Response Models for Dichotomous Responses with Applications. In *The Annual Meeting of the Psychometric Society* (pp. 311-323). Cham: Springer Nature Switzerland.

## 5.3 Desenvolvimentos Futuros

Até o momento, o desenvolvimento do capítulo 4 está concluído e está pronto para ser submetido para publicação. Tanto o capítulo 2 (ALVES; BAZÁN; ARELLANO-VALLE, 2022), quanto o capítulo 3 (ALVES; BAZÁN, 2022) consistem em artigos já publicados. Diante desse cenário, algumas sugestões para trabalhos futuros são:

- a) Para o caso de regressão binomial, a inclusão de modelos mistos nas análises, como realizado por (HUAYANAY *et al.*, 2019), representa uma proposta de grande relevância. Essa integração oferece uma flexibilidade adicional e aumenta a precisão na modelagem de conjuntos de dados complexos.
- b) Ampliar os novos modelos para abranger estruturas com traços latentes multidimensionais, conforme desenvolvido por Fragoso e Cúri (2013). Isso pode permitir explorar a adaptação desses modelos para situações com múltiplos traços latentes, oferecendo uma compreensão mais abrangente e complexa dos dados;
- c) Explorar investigações mais profundas sobre versões simplificadas desses modelos, mantendo a característica de discriminação constante. Com base nos insights do Capítulo 3, expandir esses modelos para cenários mistos, potencialmente lidando com situações onde múltiplas condições coexistem ou interagem simultaneamente;
- d) Algumas ampliações dos métodos elaborados nesta pesquisa podem ser consideradas em futuros estudos. Por exemplo, é viável conduzir uma investigação acerca da identificabilidade dos modelos e uma exploração mais aprofundada na interpretação do parâmetro extra relacionado à assimetria.
- e) Explorar a estimação por máxima verossimilhança como uma alternativa para aprimorar a análise estatística. Além disso, investir em melhorias na estimação bayesiana para torná-la mais eficiente e ágil, por meio da adoção de métodos e software alternativos, representaria uma estratégia valiosa para otimizar o processo de modelagem e análise de dados. Essas abordagens complementares têm o potencial de aprimorar a precisão e a velocidade da inferência estatística, ampliando assim as possibilidades de aplicação em uma variedade de contextos de pesquisa.



## REFERÊNCIAS

---

- ALBERT, J.; CHIB, S. Bayesian analysis of binary and polychotomous response data. **Journal of the American Statistical Association**, v. 88, p. 669–679, 1993. Citado na página 24.
- ALBERT, J.; GHOSH, M. Item response modeling. In **D. Dey, S. G. and Mallick(eds.), Generalized Linear Models: A Bayesian Perspective**, New York: Addison-Wesley, p. 174–193, 2000. Citado nas páginas 51 e 64.
- ALVES, J. S.; BAZÁN, J. L.; ARELLANO-VALLE, R. B. Flexible cloglog links for binomial regression models as an alternative for imbalanced medical data. **Biometrical Journal**, Wiley Online Library, v. 65, n. 3, p. 2100325, 2022. Citado nas páginas 48, 52, 53, 65 e 80.
- ALVES, J. S. B.; BAZÁN, J. L. New flexible item response models for dichotomous responses with applications. In: SPRINGER. **The Annual Meeting of the Psychometric Society**. [S.l.], 2022. p. 311–323. Citado nas páginas 76 e 80.
- ANDERSON, D.; BURNHAM, K. Model selection and multi-model inference. **Second. NY: Springer-Verlag**, v. 63, n. 2020, p. 10, 2004. Citado na página 34.
- ANYOSA, S. A. C. **Binary regression using power and power reverse links**. Dissertação (Mestrado) — Universidade de São Paulo, 2017. Citado na página 33.
- AYALA, R. J. D. **The theory and practice of item response theory**. [S.l.]: Guilford Publications, 2013. Citado na página 45.
- BAKHSHI, A. K.; AHMED, M. M. Bayesian extreme value analysis of kinematic-based surrogate measure of safety to detect crash-prone conditions in connected vehicles environment: A driving simulator experiment. **Transportation research part C: emerging technologies**, Elsevier, v. 136, p. 103539, 2022. Citado na página 65.
- BASU, S.; MUKHOPADHYAY, S. Binary response regression with normal scale mixture links. In: DEY, D. K.; GHOSH, S. K.; MALLICK, B. K. (Ed.). **Generalized Linear Models: A Bayesian perspective**. New York: Marcell Decker, 2000. p. 231–239. Citado na página 24.
- BAZÁN, J.; TORRES-AVILÉS, F.; SUZUKI, A. K.; LOUZADA, F. Power and reversal power links for binary regressions: An application for motor insurance policyholders. **Applied Stochastic Models in Business and Industry**, Wiley Online Library, v. 33, n. 1, p. 22–34, 2017. Citado nas páginas 24, 26, 27, 29, 48, 58 e 59.
- BAZÁN, J. L.; BOLFARINE, H.; BRANCO, M. D. A skew item response model. **Bayesian analysis**, International Society for Bayesian Analysis, v. 1, n. 4, p. 861–892, 2006. Citado nas páginas 21, 46 e 47.
- BAZÁN, J. L.; BOLFARINE, H.; BRANCO, M. D. A framework for skew-probit links in binary regression. **Communications in Statistics - Theory and Methods**, v. 39, p. 678–697, 2010. Citado na página 24.

- BAZÁN, J. L.; FLORES, S. E. A.; AZEVEDO, C. L.; DEY, D. K. Revisiting the samejima–bolfarine–bazán irt models: New features and extensions. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 37, n. 1, p. 1–25, 2023. Citado na página 62.
- BAZÁN, J. L.; ROMEO, J. S.; RODRIGUES, J. *et al.* Bayesian skew-probit regression for binary response data. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 28, n. 4, p. 467–482, 2014. Citado nas páginas 33, 34 e 35.
- BAZÁN, J. L.; E., F. S.; N., A. C. L.; DEY, D. K. Revisiting the samejima–bolfarine–bazán irt models: New features and extensions. **Brazilian Journal of Probability and Statistics**, v. 37(1), p. 1–25, 2023. Citado na página 58.
- BERKSON, J. Application of the logistic function to bio-assay. **Journal of the American Statistical Association**, American Statistical Association, Taylor & Francis, Ltd., v. 39, n. 227, p. 357–365, 1944. Citado na página 25.
- BIRNBAUM, A. L. Some latent trait models and their use in inferring an examinee’s ability. **Statistical theories of mental test scores**, Addison-Wesley, 1968. Citado nas páginas 21 e 47.
- BLISS, C. I. The calculation of the dosage-mortality curve. **Annals of Applied Biology**, v. 22, n. 1, p. 134–167, 1935. Citado na página 25.
- BOLFARINE, H.; BAZAN, J. L. Bayesian estimation of the logistic positive exponent irt model. **Journal of Educational and Behavioral Statistics**, SAGE Publications Sage CA: Los Angeles, CA, v. 35, n. 6, p. 693–713, 2010. Citado nas páginas 21, 46, 47, 48, 49, 51, 52, 53, 55, 59, 64, 65, 71, 72, 73 e 76.
- BOLT, D. M.; LIAO, X. Item complexity: A neglected psychometric feature of test items? **Psychometrika**, Springer, v. 87, n. 4, p. 1195–1213, 2022. Citado na página 49.
- BROOKS, S. Discussion of the paper bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society**, v. 64, p. 616–618, 2002. Citado nas páginas 34, 52 e 65.
- BURA, E.; GASTWIRTH, J. L. The binary regression quantile plot: assessing the importance of predictors in binary regression visually. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 43, n. 1, p. 5–21, 2001. Citado na página 40.
- CHEN, M.-H.; DEY, D. K.; SHAO, Q.-M. A new skewed link model for dichotomous quantal response data. **Journal of the American Statistical Association**, Taylor & Francis, v. 94, n. 448, p. 1172–1186, 1999. Citado nas páginas 21 e 24.
- \_\_\_\_\_. Bayesian analysis of binary data using skewed logit models. **Calcutta Statistical Association Bulletin**, SAGE Publications Sage India: New Delhi, India, v. 51, n. 1-2, p. 11–30, 2001. Citado nas páginas 24 e 48.
- CONGDON, P. **Bayesian statistical modelling**. [S.l.]: John Wiley & Sons, 2007. v. 704. Citado na página 51.
- CZADO, C. Bayesian inference of binary regression models with parametric link. **Journal of Statistical Planning and Inference**, v. 41, n. 2, p. 121–140, 1994. Citado na página 24.
- CZADO, C.; SANTNER, T. The effect of link misspecification on binary regression inference. **Journal of Statistical Planning and Inference**, v. 33, p. 213–231, 1992. Citado nas páginas 21, 24 e 48.

DEVIDAS, M.; GEORGE, E. O.; ZELTERMAN, D. Generalized logistic models for low—dose response data. **Statistics in medicine**, Wiley Online Library, v. 12, n. 9, p. 881–892, 1993. Citado na página 35.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Citado nas páginas 34 e 41.

FINNEY, D. Probit analysis, 3rd edn, 333 pp. **Cambridge University Press, London**, 1971. Citado na página 40.

FISHER, R. A. On the mathematical foundations of theoretical statistics. **Philosophical Transactions of the Royal Society**, v. 222, p. 309–368, 1922. Citado na página 25.

FRAGOSO, T. M.; CÚRI, M. Improving psychometric assessment of the beck depression inventory using multidimensional item response theory. **Biometrical Journal**, Wiley Online Library, v. 55, n. 4, p. 527–540, 2013. Citado na página 81.

GEISSER, S.; EDDY, W. F. A predictive approach to model selection. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 74, n. 365, p. 153–160, 1979. Citado nas páginas 34, 52 e 65.

GELMAN, A.; GILKS, W. R.; ROBERTS, G. O. Weak convergence and optimal scaling of random walk Metropolis algorithms. **The Annals of Applied Probability**, Institute of Mathematical Statistics, v. 7, n. 1, p. 110 – 120, 1997. Disponível em: <<https://doi.org/10.1214/aoap/1034625254>>. Citado na página 34.

GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for bayesian models. **Statistics and computing**, Springer, v. 24, n. 6, p. 997–1016, 2014. Citado na página 34.

GELMAN, A.; JAKULIN, A.; PITTAU, M. G.; SU, Y.-S. A weakly informative default prior distribution for logistic and other regression models. 2008. Citado na página 35.

GELMAN, A.; RUBIN, D. B. *et al.* Inference from iterative simulation using multiple sequences. **Statistical science**, Institute of Mathematical Statistics, v. 7, n. 4, p. 457–472, 1992. Citado nas páginas 40, 52, 70, 72 e 76.

GONÇALVES, F. B.; SILVA, J. V.; LOSCHI, R. H. Flexible bayesian modelling in dichotomous item response theory using mixtures of skewed item curves. **British Journal of Mathematical and Statistical Psychology**, Wiley Online Library, v. 76, n. 1, p. 69–86, 2023. Citado na página 46.

GUERRERO, V. M.; JOHNSON, R. A. Use of the box-cox transformation with binary response models. **Biometrika**, Oxford University Press, Biometrika Trust, v. 69, n. 2, p. 309–314, 1982. Citado nas páginas 21 e 24.

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. **Fundamentals of item response theory**. [S.l.]: Sage, 1991. v. 2. Citado na página 45.

HAND, D. Handbook of modern item response theory. **Biometrics**, International Biometric Society, v. 54, n. 4, p. 1680, 1998. Citado na página 58.

HARO-LÓPEZ, R. A.; MALLICK, B.; SMITH, A. F. M. Binary regression using data adaptive robust link functions. In: DEY, D.; GHOSH, S.; MALLICK, B. (Ed.). **Generalized Linear Models: A Bayesian Perspective**. New York: Marcel Dekker, 2000. p. 243–253. Citado na página 24.

HOFFMAN, M. D.; GELMAN, A. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. **J. Mach. Learn. Res.**, v. 15, n. 1, p. 1593–1623, 2014. Citado nas páginas 33, 46, 51, 59 e 64.

HUAYANAY, A. de la C.; BAZAN, J. L.; CANCHO, V. G.; DEY, D. K. Performance of asymmetric links and correction methods for imbalanced data in binary regression. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 89, n. 9, p. 1694–1714, 2019. Citado nas páginas 28, 79 e 81.

JIANG, X.; DEY, D. K.; PRUNIER, R.; WILSON, A. M.; HOLSINGER, K. E. A new class of flexible link functions with application to species co-occurrence in cape floristic region. **The Annals of Applied Statistics**, Institute of Mathematical Statistics, v. 7, n. 4, p. 2180–2204, 2013. ISSN 19326157. Citado na página 25.

JOHNSON, V. E.; ALBERT, J. H. Regression models for ordinal data. **Ordinal Data Modeling**, Springer, p. 126–157, 1999. Citado na página 51.

KIM, S.; CHEN, M.-H.; DEY, D. K. Flexible generalized t-link models for binary response data. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 95, n. 1, p. 93–106, 2008. Citado na página 25.

KOTZ, S.; NADARAJAH, S. **Extreme Value Distributions: Theory and Applications**. [S.l.]: Imperial College Press, London, 2000. v. 2. Citado na página 60.

LEE, D.; SINHA, S. Identifiability and bias reduction in the skew-probit model for a binary response. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 89, n. 9, p. 1621–1648, 2019. Citado na página 37.

LEE, S.; BOLT, D. M. Asymmetric item characteristic curves and item complexity: Insights from simulation and real data analyses. **psychometrika**, Springer, v. 83, n. 2, p. 453–475, 2018. Citado nas páginas 21 e 58.

LEMONTE, A. J.; BAZÁN, J. L. New links for binary regression: an application to coca cultivation in peru. **Test**, Springer, v. 27, n. 3, p. 597–617, 2018. Citado nas páginas 24, 26, 48 e 59.

LI, D.; WANG, X.; LIN, L.; DEY, D. K. Flexible link functions in nonparametric binary regression with gaussian process priors. **Biometrics**, v. 72, n. 3, p. 707–719, 2016. Citado nas páginas 24 e 59.

LORD, F. A theory of test scores. **Psychometric monographs**, 1952. Citado nas páginas 21 e 47.

LUO, Y.; JIAO, H. Using the stan program for bayesian item response theory. **Educational and psychological measurement**, SAGE Publications Sage CA: Los Angeles, CA, v. 78, n. 3, p. 384–408, 2018. Citado nas páginas 51 e 64.



- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **The journal of chemical physics**, American Institute of Physics, v. 21, n. 6, p. 1087–1092, 1953. Citado na página 33.
- MICCERI, T. The unicorn, the normal curve, and other improbable creatures. **Psychological bulletin**, American Psychological Association, v. 105, n. 1, p. 156, 1989. Citado na página 21.
- MORGAN, B. Estimating the dose for a given mortality. **Analysis of Quantal Response Data**, (Chapman and Hall, London), p. 59–67, 1992. Citado na página 40.
- NARANJO, L.; MARTÍN, J.; PÉREZ, C. Bayesian binary regression with exponential power link. **Computational Statistics & Data Analysis**, v. 71, p. 464 – 476, 2014. Citado na página 24.
- NEAL, R. M. *et al.* Mcmc using hamiltonian dynamics. **Handbook of Markov Chain Monte Carlo**, Chapman & Hall/CRC., v. 2, n. 11, p. 116–162, 2011. Citado nas páginas 33 e 46.
- PATZ, R. J.; JUNKER, B. W. A straightforward approach to markov chain monte carlo methods for item response models. **Journal of educational and behavioral Statistics**, Sage Publications, v. 24, n. 2, p. 146–178, 1999. Citado na página 51.
- PAULA, G.; TUDER, R. Utilization of logistic regression to improve diagnosis of pulmonary infection. **Ciência e Cultura**, v. 38, p. 1046–1050, 1986. Citado na página 41.
- PAULA, G. A. Influence and residuals in restricte generalized linear models. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 51, n. 2-4, p. 315–331, 1995. Citado nas páginas 42, 43 e 44.
- PRENTICE, R. L. A generalization of the probit and logit methods for dose-response curves. **Biometrika**, v. 32, p. 761–768, 1976. Citado nas páginas 21 e 24.
- RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 54, n. 3, p. 507–554, 2005. Citado na página 34.
- ROBITZSCH, A. On the choice of the item response model for scaling pisa data: Model selection based on information criteria and quantifying model uncertainty. **Entropy**, Multidisciplinary Digital Publishing Institute, v. 24, n. 6, p. 760, 2022. Citado nas páginas 47 e 55.
- ROY, V.; DEY, D. K. Propriety of posterior distributions arising in categorical and survival models under generalized extreme value distribution. **Statistica Sinica**, JSTOR, p. 699–722, 2014. Citado na página 65.
- RUBIN, M. L.; CHAN, W.; YAMAL, J.-M.; ROBERTSON, C. S. A joint logistic regression and covariate-adjusted continuous-time markov chain model. **Statistics in Medicine**, Wiley Online Library, v. 36, n. 28, p. 4570–4582, 2017. Citado na página 37.
- SAHU, S. K. Bayesian estimation and model choice in item response models. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 72, n. 3, p. 217–232, 2002. Citado nas páginas 47 e 51.
- SAMEJIMA, F. Departure from normal assumptions: A promise for future psychometrics with substantive mathematical modeling. **Psychometrika**, Springer, v. 62, n. 4, p. 471–493, 1997. Citado na página 21.

- \_\_\_\_\_. Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika*, Springer, v. 65, n. 3, p. 319–335, 2000. Citado nas páginas 21, 45, 46, 47, 48 e 49.
- SHAW, W. T.; BUCKLEY, I. R. C. The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *ArXiv e-prints*, jan. 2009. Citado nas páginas 27, 29 e 48.
- SILVA, A. N. da; ANYOSA, S.; BAZAN, J. L. Modelagem bayesiano de regressão binária para dados desbalanceados usando novas ligações. *Revista Brasileira de Biometria*, v. 38, n. 4, p. 385–417, 2020. Citado na página 34.
- SKRONDAL, A.; RABE-HESKETH, S. **Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models**. [S.l.]: Crc Press, 2004. Citado na página 58.
- SPIEGELHALTER, D.; THOMAS, A.; BEST, N.; GILKS, W. Bugs 0.5 examples (vol. 1 version i). **Cambridge, UK: University of Cambridge**, 1996. Citado na página 51.
- SPRENT, P. **Data driven statistical methods**. [S.l.]: Routledge, 2019. Citado na página 40.
- Stan Development Team. **RStan: the R interface to Stan**. 2020. R package version 2.21.2. Disponível em: <<http://mc-stan.org/>>. Citado nas páginas 34, 46 e 59.
- STUKEL, T. A. Generalized logistic models. *Journal of the American Statistical Association*, v. 83, p. 426–431, 1988. Citado nas páginas 21 e 24.
- SWAMINATHAN, H.; HAMBLETON, R. K.; SIRECI, S. G.; XING, D.; RIZAVI, S. M. Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates. *Applied psychological measurement*, Sage Publications, v. 27, n. 1, p. 27–51, 2003. Citado nas páginas 51 e 64.
- TEAM, R. D. C. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2009. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>. Citado nas páginas 51 e 64.
- TRIKALINOS, T. A.; HOAGLIN, D. C.; SCHMID, C. H. **Empirical and Simulation-Based Comparison of Univariate and Multivariate Meta-Analysis for Binary Outcomes**. Agency for Healthcare Research and Quality (US), Rockville (MD), 2013. (AHRQ Methods for Effective Health Care). Disponível em: <<http://europepmc.org/books/NBK132562>>. Citado na página 37.
- WANG, X.; DEY, D. K. *et al.* Generalized extreme value regression for binary response data: An application to b2b electronic payments system adoption. *The Annals of Applied Statistics*, Institute of Mathematical Statistics, v. 4, n. 4, p. 2000–2023, 2010. Citado nas páginas 24, 26, 29, 35 e 59.
- WATANABE, S.; OPPER, M. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, v. 11, n. 12, p. 3571–3594, 2010. Citado nas páginas 34, 52 e 64.
- YOON, S.; CHO, W.; HEO, J.-H.; KIM, C. E. A full bayesian approach to generalized maximum likelihood estimation of generalized extreme value distribution. *Stochastic environmental research and risk assessment*, Springer, v. 24, p. 761–770, 2010. Citado na página 65.

---

ZHANG, J.; ZHANG, Y.-Y.; TAO, J.; CHEN, M.-H. Bayesian item response theory models with flexible generalized logit links. **Applied Psychological Measurement**, SAGE Publications Sage CA: Los Angeles, CA, v. 46, n. 5, p. 382–405, 2022. Citado nas páginas [46](#), [72](#) e [76](#).



---

## CÓDIGOS - CAPÍTULO 2

---

O Capítulo 2, cuja publicação já ocorreu, representa um marco significativo neste trabalho. Sua total replicabilidade é um aspecto fundamental, garantindo a transparência e a confiabilidade dos resultados apresentados. Todos os códigos utilizados na análise podem ser acessados e replicados facilmente através desse link: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202100325>, o que possibilita uma verificação detalhada dos procedimentos metodológicos empregados e uma validação independente dos resultados obtidos.



---

## CÓDIGOS - CAPÍTULO 3

---

---

### Código-fonte 1 – Aplicação teste de Matemática

---

```
1: P_RG<-"  
2: data {  
3:   int<lower=0> n;  
4:   int<lower=0> p;  
5:   int<lower=0,upper=1> Y[n,p];  
6: }  
7: parameters {  
8:   vector[n] theta;  
9:   vector[p] b;  
10:  vector<lower=0>[p] a;  
11:  vector[p] lambda;  
12: }  
13: transformed parameters{  
14:  vector[p] m[n];  
15:  vector[p] pp[n];  
16:  vector<lower=0,upper=1>[p] prob[n];  
17:  vector <lower=0>[p] delta;  
18:  
19:  
20:  for(i in 1:n){  
21:    for (j in 1:p){  
22:      delta[j] = exp(lambda[j]);  
23:      m[i,j] = a[j]*(theta[i]-b[j]);  
24:      pp[i,j] = exp(-exp(m[i,j]));  
25:      prob[i,j] = pow(1 - pp[i,j], delta[j]);  
26:    }
```

```

27:   }
28: }
29: model {
30:   theta ~ normal(0,1);
31:   b ~ normal(0,1);
32:   a ~ lognormal(0,1);
33:   //lambda ~ gamma(0.25,0.25);
34:   lambda ~ uniform(-2,2);
35:
36:
37:   for(i in 1:n){
38:     for (j in 1:p){
39:       Y[i,j] ~ bernoulli(prob[i,j]);
40:     }
41:   }
42: }
43: generated quantities {
44:   vector[p] loglik_y[n];
45:   vector[p] Y_rep[n];
46:
47:
48:   for (i in 1: n){
49:     for (j in 1: p){
50:       loglik_y[i,j] = bernoulli_lpmf(Y[i,j] | prob[i,j]);
51:       Y_rep[i,j] = bernoulli_rng(prob[i,j]);
52:     }
53:   }
54: }
55: "

```

---

```

1:   RP_RG<-"
2:   data {
3:     int<lower=0> n;
4:     int<lower=0> p;
5:     int<lower=0,upper=1> Y[n,p];
6:   }
7:   parameters {
8:     vector[n] theta;
9:     vector[p] b;
10:    vector<lower=0>[p] a;
11:    vector[p] lambda;

```



```
12: }
13: transformed parameters{
14:   vector[p] m[n];
15:   vector[p] pp[n];
16:   vector<lower=0,upper=1>[p] prob[n];
17:   vector <lower=0>[p] delta;
18:
19:
20:   for(i in 1:n){
21:     for (j in 1:p){
22:       delta[j] = exp(lambda[j]);
23:       m[i,j] = a[j]*(theta[i]-b[j]);
24:       pp[i,j] = exp(-exp(-m[i,j]));
25:       prob[i,j] = 1 - pow((1-pp[i,j]), delta[j]);
26:     }
27:   }
28: }
29: model {
30:   theta ~ normal(0,1);
31:   b ~ normal(0,1);
32:   a ~ lognormal(0,1);
33:   //lambda ~ gamma(0.25,0.25);
34:   lambda ~ uniform(-2,2);
35:
36:   for(i in 1:n){
37:     for (j in 1:p){
38:       Y[i,j] ~ bernoulli(prob[i,j]);
39:     }
40:   }
41: }
42: generated quantities {
43:   vector[p] loglik_y[n];
44:   vector[p] Y_rep[n];
45:
46:
47:   for (i in 1: n){
48:     for (j in 1: p){
49:       loglik_y[i,j] = bernoulli_lpmf(Y[i,j] | prob[i,j]);
50:       Y_rep[i,j] = bernoulli_rng(prob[i,j]);
51:     }
52:   }
53: }
```

54: "

---

```

1: TS_G<-"
2: data {
3:   int<lower=0> n;
4:   int<lower=0> p;
5:   int<lower=0,upper=1> Y[n,p];
6: }
7: parameters {
8:   vector[n] theta;
9:   vector[p] b;
10:  vector<lower=0>[p] a;
11:  vector[p] lambda;
12: }
13: transformed parameters{
14:   vector[p] m[n];
15:   vector[p] pp[n];
16:   vector[p] ppp[n];
17:   vector<lower=0,upper=1>[p] prob[n];
18:   vector<lower=-1,upper=1>[p] delta;
19:
20:   for(i in 1:n){
21:     for (j in 1:p){
22:       delta[j] = (exp(lambda[j])-1)/(exp(lambda[j])+1);
23:       m[i,j] = a[j]*(theta[i]-b[j]);
24:       pp[i,j] = exp(-exp(-m[i,j]));
25:       ppp[i,j] = delta[j] * pp[i,j];
26:       prob[i,j] = (pp[i,j]) * (1 + delta[j] - ppp[i,j]);
27:     }
28:   }
29: }
30: model {
31:   theta ~ normal(0,1);
32:   b ~ normal(0,1);
33:   a ~ lognormal(0,1);
34:   //lambda ~ gamma(0.25,0.25);
35:   lambda ~ uniform(-2,2);
36:
37:   for(i in 1:n){
38:     for (j in 1:p){
39:       Y[i,j] ~ bernoulli(prob[i,j]);

```

```
40:     }
41:   }
42: }
43: generated quantities {
44:   vector[p] loglik_y[n];
45:   vector[p] Y_rep[n];
46:
47:   for (i in 1: n){
48:     for (j in 1: p){
49:       loglik_y[i,j] = bernoulli_lpmf(Y[i,j] | prob[i,j]);
50:       Y_rep[i,j] = bernoulli_rng(prob[i,j]);
51:     }
52:   }
53: }
54: "
```

---

```
1: TS_RG<-"
2: data {
3:   int<lower=0> n;
4:   int<lower=0> p;
5:   int<lower=0,upper=1> Y[n,p];
6: }
7: parameters {
8:   vector[n] theta;
9:   vector[p] b;
10:  vector<lower=0>[p] a;
11:  vector[p] lambda;
12: }
13: transformed parameters{
14:  vector[p] m[n];
15:  vector[p] pp[n];
16:  vector[p] ppp[n];
17:  vector<lower=-1,upper=1>[p] delta;
18:  vector<lower=0,upper=1>[p] prob[n];
19:
20:
21:  for(i in 1:n){
22:    for (j in 1:p){
23:      delta[j] = (exp(lambda[j])-1)/(exp(lambda[j])+1);
24:      m[i,j] = a[j]*(theta[i]-b[j]);
25:      pp[i,j] = exp(-exp(m[i,j]));
```

```
26:         ppp[i,j] = delta[j] * pp[i,j];
27:         prob[i,j] = (1 - pp[i,j])*(1 + ppp[i,j]);
28:     }
29: }
30: }
31: model {
32:     theta ~ normal(0,1);
33:     b ~ normal(0,1);
34:     a ~ lognormal(0,1);
35:     //lambda ~ gamma(0.25,0.25);
36:     lambda ~ uniform(-2,2);
37:
38:     for(i in 1:n){
39:         for (j in 1:p){
40:             Y[i,j] ~ bernoulli(prob[i,j]);
41:         }
42:     }
43: }
44: generated quantities {
45:     vector[p] loglik_y[n];
46:     vector[p] Y_rep[n];
47:
48:     for (i in 1: n){
49:         for (j in 1: p){
50:             loglik_y[i,j] = bernoulli_lpmf(Y[i,j] | prob[i,j]);
51:             Y_rep[i,j] = bernoulli_rng(prob[i,j]);
52:         }
53:     }
54: }
55: "
```

---

---

## CÓDIGOS - CAPÍTULO 4

---

---

### Código-fonte 2 – Aplicação teste de Matemática

---

```
1: GEV_G2P <- "  
2: data {  
3:   int<lower=0> n;  
4:   int<lower=0> p;  
5:   int<lower=0,upper=1> Y[n,p];  
6: }  
7: parameters {  
8:   vector[n] theta;  
9:   vector[p] b;  
10:  vector<lower=0>[p] a;  
11:  vector[p] lambda;  
12: }  
13: transformed parameters{  
14:   vector[p] m[n];  
15:   vector[p] w[n];  
16:   vector[p] aux[n];  
17:   vector<lower=0,upper=1>[p] prob[n];  
18:  
19:   for(i in 1:n){  
20:     for (j in 1:p){  
21:       m[i,j] = a[j]*(theta[i] - b[j]);  
22:       w[i,j] = 1 + lambda[j]*m[i,j];  
23:       aux[i,j] = if_else(w[i,j]>0, w[i,j], 0);  
24:       prob[i,j] = exp(-pow(aux[i,j], (-1/lambda[j])));  
25:     }  
26:   }
```

```

27: }
28: model {
29:   theta ~ normal(0,1);
30:   b ~ normal(0,1);
31:   a ~ lognormal(0,1);
32:   lambda ~ uniform(-2,2);
33:
34:   for(i in 1:n){
35:     for (j in 1:p){
36:       Y[i,j] ~ bernoulli(prob[i,j]);
37:     }
38:   }
39: }
40: generated quantities {
41:   vector[p] loglik_y[n];
42:   vector[p] Y_rep[n];
43:   for (i in 1: n){
44:     for (j in 1: p){
45:       loglik_y[i,j] = bernoulli_lpmf(Y[i,j] | prob[i,j]);
46:       Y_rep[i,j] = bernoulli_rng(prob[i,j]);
47:     }
48:   }
49: }
50: "

```

---

```

1:   GEV_RG2P<- "
2:   data {
3:     int<lower=0> n;
4:     int<lower=0> p;
5:     int<lower=0,upper=1> Y[n,p];
6:   }
7:   parameters {
8:     vector[n] theta;
9:     vector[p] b;
10:    vector<lower=0>[p] a;
11:    vector[p] lambda;
12:  }
13:  transformed parameters{
14:    vector[p] m[n];
15:    vector[p] w[n];
16:    vector[p] aux[n];

```

---

```
17:     vector<lower=0,upper=1>[p] prob[n];
18:
19:     for(i in 1:n){
20:         for (j in 1:p){
21:             m[i,j] = a[j]*(theta[i]-b[j]);
22:             w[i,j] = 1 - lambda[j]*m[i,j];
23:             aux[i,j] = if_else(w[i,j]>0, w[i,j], 0);
24:             prob[i,j] = 1 - exp(-pow(aux[i,j], (-1/lambda[j])));
25:         }
26:     }
27: }
28: model {
29:     theta ~ normal(0,1);
30:     b ~ normal(0,1);
31:     a ~ lognormal(0,1);
32:     lambda ~ uniform(-2,2);
33:
34:     for(i in 1:n){
35:         for (j in 1:p){
36:             Y[i,j] ~ bernoulli(prob[i,j]);
37:         }
38:     }
39: }
40: generated quantities {
41:     vector[p] loglik_y[n];
42:     vector[p] Y_rep[n];
43:
44:
45:     for (i in 1: n){
46:         for (j in 1: p){
47:             loglik_y[i,j] = bernoulli_lpmf(Y[i,j] | prob[i,j]);
48:             Y_rep[i,j] = bernoulli_rng(prob[i,j]);
49:
50:         }
51:     }
52: }
53: "
```

---

