

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Modelos de regressão binária espacial bayesiana para dados desbalanceados

Alan da Silva Assunção

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

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Alan da Silva Assunção

Modelos de regressão binária espacial bayesiana para dados desbalanceados

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 Doutor em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Ricardo Sandes Ehlers

USP – São Carlos
Setembro de 2025

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

ASSUNCAO, ALAN
A111m Modelos de regressão binária espacial bayesiana
para dados desbalanceados / ALAN ASSUNCAO;
orientador Ricardo Sandes Ehlers. -- São Carlos,
2025.
168 p.

Tese (Doutorado - Programa Interinstitucional de
Pós-graduação em Estatística) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2025.

1. Regressão binária. 2. Espacial. 3. Ligação
assimétrica. 4. MCMC. 5. Monte Carlo Hamiltoniano.
I. Sandes Ehlers, Ricardo , orient. II. Título.

Alan da Silva Assunção

Bayesian spatial binary regression models for imbalanced
data

Doctoral dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Doctorate, Interagency Graduate Program in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Ricardo Sandes Ehlers

USP – São Carlos
September 2025



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Tese de Doutorado do candidato Alan da Silva Assunção, realizada em 28/08/2025.

Comissão Julgadora:

Prof. Dr. Ricardo Sandes Ehlers (USP)

Prof. Dr. Marcos Oliveira Prates (UFMG)

Profa. Dra. Marina Silva Paez (UFRJ)

Prof. Dr. João Batista de Moraes Pereira (UFRJ)

Prof. Dr. Leonardo Soares Bastos (FIOCRUZ)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

*Dedico este trabalho à minha família que,
com seu apoio e suporte, me fez chegar até aqui. Vocês são presentes de Deus na minha vida!.*

AGRADECIMENTOS

Ao meu amado Deus, soberano Criador, por Seu amor inigualável e por me sustentar todos os dias da minha vida. Ao Senhor Jesus Cristo, meu Salvador, por Sua graça infindável, e ao doce Espírito Santo, por Sua companhia perene. A Eles sejam a honra, a glória e o louvor para todo o sempre.

Aos meus amados pais, Luzia e Valdeci, e à minha amada irmã, Alaíde, por sempre me apoiarem em todos os momentos da minha vida; e aos demais parentes por sempre me ajudarem em tudo que precisei. Amo vocês!

À minha amada Kleunice, minha compromissada, por sempre me apoiar e me fortalecer nos momentos de luta. Te amo, meu bem!

Aos meus amigos, em especial João Marcos, por ser um amigo para todas as horas, amigo mais chegado que um irmão! Louvo a Deus por ter colocado vocês em minha vida e por serem pessoas tão especiais. Que Deus os abençoe sempre!

Aos meus pastores, Wendell de Carvalho e Fernanda de Carvalho, da Igreja Batista Nova Aliança em Timon-MA — suas palavras motivadoras e exortações foram essenciais para que eu chegasse até este momento ímpar. Ao meu amigo e líder de geração, pastor Fabrício Gomes, pela amizade e pelos conselhos dados ao longo desses quase dez anos em que caminhamos juntos. Agradeço também aos meus irmãos de fé por cada palavra de bênção liberada sobre minha vida.

Ao meu orientador, professor Ricardo Ehlers — nunca imaginei que um dia teria a oportunidade de tê-lo como orientador. Que privilégio! Só tenho a agradecer por todo o conhecimento e aprendizado que obtive durante esse tempo de pesquisa. Obrigado pela confiança em mim depositada.

Aos professores e funcionários do Departamento de Estatística da UFPI, por me darem todo o suporte de que precisei em diversos momentos da minha pesquisa.

Aos professores do PIPGEs, por todo o conhecimento adquirido durante o curso. Aos funcionários do PIPGEs, do ICMC-USP e do Des-UFSCar, pela gentileza com que sempre me trataram.

Aos colegas de curso, por toda a ajuda que recebi, especialmente durante o período da pandemia. Agradeço ao Flávio Barros que, gentil e amavelmente, me ajudou a usar o cluster Euler do Centro de Ciências Matemáticas Aplicadas à Indústria (CeMEAI). Sua ajuda foi muito importante — que Deus o abençoe!

Ao CeMEAI, pelo suporte computacional que possibilitou grande parte do desenvolvimento das análises e resultados apresentados nesta tese. Sem esse suporte, tenho certeza de que os resultados que obtive não seriam possíveis.

E a todos que colaboraram direta ou indiretamente para a conclusão deste trabalho, os meus mais sinceros agradecimentos.

“Cada experiência que Deus nos dá... é a preparação perfeita para um futuro que só Ele pode ver.”

(Corrie Ten Boom)

RESUMO

ASSUNÇÃO, A. S. **Modelos de regressão binária espacial bayesiana para dados desbalanceados**. 2025. 168 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, 2025.

Modelos de regressão binária são excelentes propostas de modelagem para dados dicotômicos, que nos permitem relacionar a probabilidade do evento de interesse com as covariáveis disponíveis. Nesse tipo de cenário, é comum a ocorrência de desbalanceamento dos dados, isto é, uma proporção de zeros (ou uns) significativamente diferente de uns (ou zeros), fazendo com que funções de ligações simétricas não sejam boas alternativas no momento de ajustar o modelo. Neste trabalho, propomos uma classe de modelos de regressão binária ajustada com funções de ligação assimétricas, a saber: as funções de ligação potência e reversa de potência. Além disso, incrementamos efeitos aleatórios espaciais em nossa regressão, assumindo assim, que os dados binários podem ser referenciados no espaço. A regressão binária resultante torna-se um tipo especial de modelo hierárquico bayesiano cuja estrutura espacial modelamos por meio de uma distribuição a priori mais flexível que o modelo CAR (*Conditional autoregressive*) padrão: a distribuição a priori G-Wishart. Através de um conjunto de dados motivacionais, apresentamos uma extensão do modelo de regressão binária espacial proposto. Para ambos os modelos, fornecemos estudos de simulação e aplicação em dados reais, além de análise bayesiana de detecção de pontos influentes. A estimação dos parâmetros é feita de forma completamente bayesiana, no qual buscamos a maximização da eficiência computacional no processo de estimação. O desempenho dos modelos aqui propostos será avaliado e comparados através de métricas de diagnóstico bayesiana e qualidade preditiva. O desempenho do nosso algoritmo será avaliado através de simulações computacionais, como também, em conjunto de dados reais. Para concluir esta tese, apresentamos de forma preliminar e motivacional um estudo de simulação com o modelo espacial DAGAR.

Palavras-chave: Regressão Binária, Espacial, Ligação Assimétrica, MCMC, Monte Carlo Hamiltoniano.

ABSTRACT

ASSUNÇÃO, A. S. **Bayesian spatial binary regression models for imbalanced data**. 2025. 168 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, 2025.

Binary regression models are excellent modeling approaches for dichotomous data, allowing us to relate the probability of the event of interest to the available covariates. In this type of scenario, it is common to encounter data imbalance, that is, a proportion of zeros (or ones) significantly different from ones (or zeros), which makes symmetric link functions poor alternatives when fitting the model. In this work, we propose a class of binary regression models fitted with asymmetric link functions, namely: the power and reverse power link functions. Additionally, we incorporate spatial random effects into our regression, thereby assuming that the binary data can be spatially referenced. The resulting binary regression becomes a special type of Bayesian hierarchical model whose spatial structure is modeled through a more flexible prior distribution than the standard CAR (Conditional Autoregressive) model: the G-Wishart prior distribution. Through a motivational dataset, we present an extension of the proposed spatial binary regression model. For both models, we provide simulation studies and applications to real data, as well as a Bayesian analysis for the detection of influential points. Parameter estimation is fully Bayesian, with a focus on maximizing computational efficiency during the estimation process. The performance of the models proposed herein will be assessed and compared using Bayesian diagnostic metrics and predictive quality measures. The performance of our algorithm will be evaluated through computational simulations and applications to real-world datasets. Finally, as a preliminary and motivational perspective, we present a simulation study using the DAGAR spatial model.

Keywords: Binary Regression, Spatial, Asymmetric Link, MCMC, Monte Carlo Hamiltonian.

LISTA DE ILUSTRAÇÕES

Figura 1 – Trimestre menos chuvoso ('AGOSTO-SETEMBRO-OUTUBRO'=1) para cada município do estado do Piauí ao longo de 30 anos.	34
Figura 2 – Funções de distribuição acumulada e funções densidade de probabilidade para os modelos: C-P, C-RP, LA-P e LA-RP usando diferentes valores de λ	51
Figura 3 – Função de distribuição acumulada para as distribuições Logística, Normal padrão, Cauchy, Gumbel Reversa e Gumbel. Simetria em torno de $F(0) = 0,5$ para $x \in (-4,4)$	54
Figura 4 – Ajuste de modelos com funções de ligação Logito (linha sólida), Probit (linha tracejada), e Cloglog (linha pontilhada) ao subconjunto de dados ortopédicos com inclinação pélvica como uma covariável. Figura (a): ajuste dos modelos aos dados originais. Figura (b): dados agrupados em categorias.	56
Figura 5 – Curva ROC associado a um modelo ingênuo.	65
Figura 6 – Mapa de estações (16 estações) de monitoramento de umidade relativa do ar para o estado brasileiro do Maranhão. A medida de plotagem refere-se à média da umidade relativa do ar para o ano de 2010.	69
Figura 7 – Mapa do PIB das unidades federativas da Região Nordeste (Brasil) para o ano de 2019.	70
Figura 8 – Grafo circular para 8 sub-regiões.	76
Figura 9 – Distribuição da variável resposta para cada município do estado da Bahia.	81
Figura 10 – Estimativas pontuais e intervalos de credibilidade de 95% a posteriori, usando todas as covariáveis. Modelos C-P e Logito.	82
Figura 11 – Gráficos de traço e autocorrelação dos modelos C-P e Logito.	83
Figura 12 – Probabilidades a posteriori estimadas de serem <i>outliers</i> com função convexa ψ_1 e divergência de Bregman normalizada com função convexa ψ_1 , contra observações deletadas para os dados de 'trimestre menos chuvoso' para cada município baiano.	85
Figura 13 – Divergência de Bregman normalizada com função convexa ψ_1 de possíveis municípios influentes para pelo menos 8 modelos.	88
Figura 14 – Probabilidades preditivas do evento de interesse para cada município da Bahia sob os modelos GR-P, Probit, Logito e Cloglog.	91

Figura 15 – Tipos de dentes (M, molar; PM, pré-molar; C, canino; IC, incisivo), localizações (bucal, ou lado da bochecha, versus lingual, ou lado da língua) e numeração das medições (como 1–6 para o segundo molar no quadrante superior esquerdo), para um sujeito hipotético sem dentes ausentes.	94
Figura 16 – Proporção empírica de CAL em relação às mandíbulas superior e inferior para os 288 indivíduos.	95
Figura 17 – <i>Trace-plots</i> e gráficos de autocorrelação para o modelos GR-P.	102
Figura 18 – <i>Trace-plots</i> e gráficos de autocorrelação para os modelos GR-RP.	103
Figura 19 – Probabilidades a posteriori estimadas de serem <i>outliers</i> com função convexa ψ_1 , e divergência de Bregman normalizada com função convexa ψ_1 , contra observações deletadas para os dados periodontais.	106
Figura 20 – Gráfico (a): probabilidades preditivas a posteriori de doença periodontal para um sujeito aleatório com idade de 25 anos, IMC 25, e combinando todos os valores de gênero, hábitos de fumar e níveis de HbA1c. Gráfico (b): probabilidades preditivas posteriores de doença periodontal para um sujeito aleatório com idade de 50 anos, IMC 40, e combinando todos os valores de gênero, hábitos de fumar e níveis de HbA1c. HNFC: Homem - Não fuma - Diabetes controlado; HNFA: Homem - Não fuma - Diabetes alto; HFC: Homem - Fumo - Diabetes controlado; MFA: Homem - Fumo - Diabetes alto; MNFC: Mulher - Não fuma - Diabetes controlado; MNFA: Mulher - Não fuma - Diabetes alto; MFC: Mulher - Fumo - Diabetes controlado; MFA: Mulher - Fumo - Diabetes alto;	107
Figura 21 – Diferença entre o grafo sob o modelo CAR e o grafo acíclico direcionado sob o modelo DAGAR, considerando a direção Oeste \rightarrow Leste, para os estados que compõem o norte brasileiro.	114
Figura 22 – Ilustração de como um objeto no Rcpp serve de transição entre objetos do R e do C++.	132
Figura 23 – Funções de distribuição acumulada e funções densidade de probabilidade para os modelos: L-P e L-RP, usando diferentes valores de λ	149
Figura 24 – Funções de distribuição acumulada e funções densidade de probabilidade para os modelos: N-P, N-RP, GR-P e GR-RP, usando diferentes valores de λ	150
Figura 25 – Funções de distribuições acumuladas e funções densidade de probabilidade para os modelos G-AT e GR-AT, usando diferentes valores de λ	154
Figura 26 – <i>Box-plot</i> de cada métrica, para cada modelo, baseado em 100 conjuntos de dados gerados de um modelo Logito. pc - ponto de corte estimado	157
Figura 27 – <i>Box-plot</i> de cada métrica para cada modelo. Gráfico superior : $m = 200$ e $\lambda = 0, 5$; gráfico inferior: $m = 200$ e $\lambda = 4$. pc - ponto de corte estimado	158
Figura 28 – <i>Box-plot</i> de cada métrica para cada modelo. Gráfico superior: $m = 500$ e $\lambda = 0, 5$; gráfico inferior: $m = 500$ e $\lambda = 4$. pc - ponto de corte estimado	159

Figura 29 – Gráficos de traço e autocorrelação dos modelos ajustados.	160
Figura 30 – Probabilidades preditivas para cada município da Bahia, sob os modelos C-P, N-P, N-RP, L-P, L-RP, LA-P, LA-RP e GR-RP. Modelos ajustados sob priori G-Wishart (GW-E).	161

LISTA DE ALGORITMOS

Algoritmo 1 – Metropolis-Hastings	44
Algoritmo 2 – Amostrador de Gibbs	45
Algoritmo 3 – Monte Carlo Hamiltoniano	47

LISTA DE TABELAS

Tabela 1 – Distribuições potência e reversa de potência.	49
Tabela 2 – fda e fdp das distribuições usadas neste trabalho.	50
Tabela 3 – Número estimado de “Normal=1” sob diferentes funções de ligação.	56
Tabela 4 – Matriz de confusão.	63
Tabela 5 – Métricas de desempenho para cenário de desbalanceamento. t_0 e t_1 são valores de ponto de corte em $(0, 1)$, sendo neste caso, $t_0 = 0$ e $t_1 = 1$	65
Tabela 6 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $m = 200$ e $m = 500$. 5000 iterações, descartando 2500 como <i>burn-in</i> . Modelos: C-P, C-RP, GR-P, GR-RP e LA-P	78
Tabela 7 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $m = 200$ e $m = 500$. 5000 iterações, descartando 2500 como <i>burn-in</i> . Modelos: L-P, L-RP, N-P, N-RP e LA-RP	79
Tabela 8 – Comparação de modelos para diferentes conjuntos de dados de tamanho 100, gerados a partir de um modelo Logito.	80
Tabela 9 – Características do conjunto de dados ‘trimestre menos chuvoso’ para os municípios do estado da Bahia.	81
Tabela 10 – Estimativas dos parâmetros usando diferentes funções de ligação: C-P, C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, Probit, Logito e Cloglog. As estimativas são a média a posteriori, desvio padrão (em parênteses) e intervalo de credibilidade de 95%.	84
Tabela 11 – Comparação de modelos usando diferentes funções de ligação: C-P, N-P, N-RP, L-P, L-RP, LA-P, LA-RP, Logito, Probit, Cloglog, GR-P e GR-RP.	86
Tabela 12 – Comparação de modelos usando diferentes funções de ligação: C-P, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato de Mohammadi e Wit (2015) (GW-E); ii) atribuindo priori CAR para a estrutura espacial (CAR).	87

Tabela 13 – Comparação de modelos segundo métricas de qualidade da predição, usando diferentes funções de ligação: C-P, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P, LA-RP, Logito, Probitto e Cloglog. P_{cmax} - estimativa do ponto de corte.	89
Tabela 14 – Comparação de modelos segundo métricas de qualidade da predição, usando diferentes funções de ligação: C-P, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato de Mohammadi e Wit (2015) (GW-E); ii) atribuindo priori CAR para a estrutura espacial (CAR)	90
Tabela 15 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostra $r = 50$. 5000 iterações, descartando 2500 como <i>burn-in</i>	99
Tabela 16 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostra $r = 200$. 5000 iterações, descartando 2500 como <i>burn-in</i>	100
Tabela 17 – Comparação de modelos para diferentes conjuntos de dados de tamanho 100, gerados a partir de um modelo Logito.	101
Tabela 18 – Estimativas dos parâmetros usando diferentes funções de ligação: C-P, C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, Probitto, Logito e Cloglog. As estimativas são a média a posteriori, desvio padrão (em parênteses) e intervalo de credibilidade de 95%.	104
Tabela 19 – Comparação de modelos usando diferentes funções de ligação: C-P,C-RP, N-P, N-RP, L-P, L-RP, LA-P, LA-RP, Logito, Probitto, Cloglog, GR-P e GR-RP.	105
Tabela 20 – Comparação de modelos usando diferentes funções de ligação: C-P,C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato de Mohammadi e Wit (2015) (GW-E); ii) e atribuindo priori CAR para a estrutura espacial (CAR).	109
Tabela 21 – Comparação de modelos, segundo métricas de qualidade da predição, usando diferentes funções de ligação C-P,C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, Logito, Probitto e Cloglog. P_{cmax} - estimativa do ponto de corte.	110

Tabela 22 – Comparação de modelos, segundo métricas de qualidade da predição, usando diferentes funções de ligação: C-P, C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato de Mohammadi e Wit (2015) (GW-E); ii) e atribuindo priori CAR para a estrutura espacial (CAR).	111
Tabela 23 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para o modelo C-P, baseado em 100 replicações para o tamanho de amostras $r = 100$. 5000 iterações, descartando 2500 como <i>burn-in</i>	117
Tabela 24 – Comparação do modelo C-P sob os modelos espaciais G-Wishart, CAR e DAGAR para diferentes conjuntos de dados simulados de tamanho $r = 100$	119
Tabela 25 – Comparação do modelo C-P sob os modelos espaciais G-Wishart, CAR e DAGAR para diferentes conjuntos de dados simulados de tamanho $r = 200$	120
Tabela 26 – fda e fdp das distribuições assimétricas.	153
Tabela 27 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $n = 200$ e $n = 500$. 5000 iterações, descartando 2500 como <i>burn-in</i> . Modelos: G-AT e GR-AT.	155
Tabela 28 – Comparação de modelos para diferentes conjuntos de dados de tamanho 100, gerados a partir de um modelo Logito. 5000 iterações, descartando as primeiras 2500 como <i>burn-in</i>	155
Tabela 29 – Comparação de modelos baseada em métricas de predição para diferentes conjuntos de dados de tamanho 100, gerados a partir de um modelo Logito. pc - ponto de corte estimado.	157
Tabela 30 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $m = 200$ e $m = 500$. 5000 iterações, descartando 2500 como <i>burn-in</i> . Modelos: C-P, C-RP, GR-P, GR-RP e LA-P.	162
Tabela 31 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $m = 200$ e $m = 500$. 5000 iterações, descartando 2500 como <i>burn-in</i> . Modelos: L-P, L-RP, N-P, N-RP e LA-RP	163

Tabela 32 – Comparação de modelos segundo métricas de qualidade da predição, usando diferentes funções de ligação: Logito, Probit e Cloglog, sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato (GW-E); ii) e atribuindo priori CAR para a estrutura espacial (CAR). P _{max} - ponto de corte estimado.	165
Tabela 33 – Comparação de modelos usando as ligações G-AT e GR-AT	166

LISTA DE SÍMBOLOS

Ω — matriz de precisão

G-Wishart $_{\mathbf{W}}(\kappa, \mathbf{S})$ — distribuição G-Wishart com graus de liberdade κ e matriz escala \mathbf{S}

$p(\theta)$ — distribuição a priori de θ

$l(\theta; \mathbf{y})$ — função de verossimilhança

$p(\theta|\mathbf{y})$ — distribuição a posteriori de θ

$p(\mathbf{y})$ — distribuição marginal de \mathbf{y}

\propto — proporcional a...

$E_{p(\theta|\mathbf{y})}[h(\theta)] = \int h(\theta)p(\theta|\mathbf{y})d\theta$ — esperança da função $h(\theta)$ para alguma função $h(\cdot)$, sob distribuição $p(\theta|\mathbf{y})$

$q(\theta^*|\theta^t)$ — distribuição proposta no algoritmo de Metropolis Hastings

min — refere-se à função que calcula o mínimo de uma sequência de dados

$p(\theta_j|\boldsymbol{\theta}_{-j})$ — distribuição condicional completa de θ_j , dado $\boldsymbol{\theta}_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d)^\top$

\mathbb{R}^d — Conjunto dos números reais d -dimensionais

\mathbf{M} — matriz de massas do vetor de posição \mathbf{p}

$H(\boldsymbol{\theta}, \mathbf{p})$ — função hamiltoniana

$-\mathcal{L}(\boldsymbol{\theta})$ — $\log p(\boldsymbol{\theta}|\mathbf{y})$

\mathbf{p} — vetor de posição do HMC

N_d — Normal multivariada de dimensão d

\mathbf{M} — matriz de massa no HMC

$\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})$ — gradiente da log posteriori negativa de $\boldsymbol{\theta}$

\mathbf{I}_d — matriz identidade de dimensão d

$\mathbf{G}(\boldsymbol{\theta})$ — matriz de massas do RMHMC

$H_P(y|\lambda)$ — distribuição potência H

λ — parâmetro de forma das distribuições potência e reversa de potência

$H_{RP}(y|\lambda)$ — distribuição reversa de potência H

$h_P(y|\lambda)$ — função densidade de probabilidade potência H

$h_P(y|\lambda)$ — função densidade de probabilidade reversa de potência H

μ e σ — parâmetros de localização e escala, respectivamente

$F_\lambda(y)$ — função de distribuição acumulada parametrizada em função de λ

$\exp(\cdot)$ — função exponencial

$\text{sign}(\cdot)$ — função sinal

$\arctan(\cdot)$ — função tangente inversa

\top — transposta de uma matriz ou vetor

p_i — $P(Y_i = 1)$

\mathbf{X} — matriz de delineamento

β — vetor de coeficientes de regressão

Y — variável dependente ou variável resposta em um modelo estatístico

$\eta(\cdot)$ — Preditor linear do modelo de regressão

$F^{-1}(\cdot)$ — Função de ligação

$L(\beta)$ — função de verossimilhança do vetor de coeficientes de regressão do modelo de regressão binária bayesiano padrão

\equiv — equivalente

$\Phi(\cdot)$ — fda da distribuição de probabilidade Normal padrão

$p_t(\tilde{y}_i)$ — distribuição de probabilidade (desconhecida) que representa o verdadeiro processo gerador de dados para \tilde{y}_i

\tilde{y}_i — observação futura

$p(\boldsymbol{\theta}|\mathbf{y})$ — distribuição a posteriori do parâmetro $\boldsymbol{\theta}$

$V_{l=1}^N$ — variância amostral

$L(a, \tilde{y}; \mathbf{y})$ — medida de perda em que a seria a “ação” tomada para prever \tilde{y} , este último seria a observação a ser predita e \mathbf{y} seria os dados observados

$L(\tilde{\mathbf{y}}, \mathbf{a}; \mathbf{y})$ — medida de perda

$p_v(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})$ — distribuição a posteriori de $\boldsymbol{\theta}$ com uma perturbação v

ψ — uma função estritamente convexa e diferenciável

D_ψ — divergência de Bregman funcional sob função convexa ψ

$D_{\|\psi\|}(\cdot)$ — divergência Bregman normalizada

\mathcal{B} — operador definido em Danilevicz e Ehlers (2022)

\mathbb{R}^d — conjunto dos números reais d - dimensionais

\mathbf{s}_i — uma localização espacial

D — um subconjunto d -dimensional

∂_i — conjunto de sub-regiões vizinhas à sub-região i

\mathbf{W} — matriz de proximidade; quando $w_{ij} = 1$ para $i \sim j$ e 0 caso contrário, \mathbf{W} recebe o nome de matriz de adjacência

$i \sim j$ — sub-região i é vizinha da sub-região j

$\phi_i, i = 1 \dots, m$ — efeitos aleatórios espaciais

\mathbf{S} — matriz escala da distribuição a priori G-Wishart

$M_{\mathbf{W}}$ — representa o conjunto de matrizes simétricas positivas definidas restritas a terem entradas nulas fora da diagonal referente aos zeros em \mathbf{W}

$\delta = \log \lambda$ — reparametrização do parâmetro de forma das funções de ligação potência e reversa de potência

μ_i — média Preditiva a posteriori da observação \tilde{y}_i

\mathbf{B} — matriz triangular estritamente inferior

$\mathbf{F} = \text{diag}(\tau_1, \dots, \tau_m)$ — matriz diagonal com τ_1, \dots, τ_m

$\mathbf{L} = \mathbf{I}_m - \mathbf{B}$

$N(i)$ — o conjunto de vizinhos tal que $b_{ij} = 0$ para todo $j \notin N(i)$

$n_{<i} = |N(i)|$ — denota a cardinalidade do conjunto de vizinhos da unidade de área i

ρ — parâmetro presente na matriz de precisão do modelo DAGAR e CAR

π — um direcionamento arbitrário das observações

ϵ — vetor de termos de erro no modelo CAR e no modelo SAR

H_{AT} — fda de uma distribuição assimétrica transmutada como modelo *baseline H*

H_{AT} — fdp de uma distribuição assimétrica transmutada como modelo *baseline H*

τ_ϕ — precisão marginal do modelo DAGAR

$|\cdot|$ — determinante de uma matriz

SUMÁRIO

1	INTRODUÇÃO	33
1.1	Regressão Binária com Função de Ligação Assimétrica	35
1.2	Modelagem para Dados Espaciais	37
1.3	Proposta do Trabalho e Objetivos	38
1.4	Estrutura do Trabalho	39
2	CONCEITOS PRELIMINARES	41
2.1	Inferência Bayesiana	41
2.2	Monte Carlo via Cadeias de Markov	42
2.2.1	<i>Metropolis-Hastings</i>	43
2.2.2	<i>Amostrador de Gibbs</i>	44
2.2.3	<i>Monte Carlo Hamiltoniano</i>	45
2.2.3.1	<i>Riemann Manifold Monte Carlo Hamiltoniano - RMHMC</i>	47
2.3	Distribuições Potência e Reversa de Potência	48
2.4	Modelo de Regressão Binária Bayesiano	52
2.4.1	<i>Desbalanceamento de Classes em Classificação Binária</i>	55
2.5	Comparação de Modelos	57
2.5.1	<i>Crítério de informação amplamente aplicável WAIC (widely applicable information criterion)</i>	57
2.5.2	<i>Avaliação de Modelos via Medida L</i>	58
2.6	Diagnóstico de Modelo e Detecção de Outliers	60
2.6.1	<i>Divergência de Bregman Funcional Normalizada</i>	62
2.7	Comparação de Modelos Baseada na Predição	62
3	MODELOS DE REGRESSÃO BINÁRIA ESPACIAL	67
3.1	Tipos de Dados Espaciais	67
3.2	Modelos Hierárquicos Espaciais para Dados de Área	69
3.3	Modelo de Regressão Binária Espacial	73
3.4	Estudo de Simulação	75
3.5	Aplicação	80
3.6	Discussões Finais	92
4	MODELOS DE REGRESSÃO BINÁRIA ESPACIAL: UMA GENERALIZAÇÃO	93
4.1	Dados Motivacionais	93

4.2	O Modelo Hierárquico Espacial	95
4.3	Estudo de Simulação	97
4.4	Aplicação	101
4.5	Discussões Finais	112
5	REGRESSÃO BINÁRIA ESPACIAL COM MODELOS DE GRAFOS ACÍCLICOS DIRECIONADOS AUTOREGRESSIVOS (DAGAR)	113
5.1	O modelo DAGAR	113
5.2	Estudo de Simulação	116
5.3	Discussões Finais	118
6	CONCLUSÕES E TRABALHOS FUTUROS	121
6.1	Produções	122
6.1.1	<i>Trabalhos Apresentados em Eventos</i>	122
6.2	Propostas de Trabalhos Futuros	123
	REFERÊNCIAS	125
	APÊNDICE A RCPP: UMA BREVE INTRODUÇÃO	131
	APÊNDICE B EXPRESSÕES MATEMÁTICAS ÚTEIS	139
B.1	Funções de Ligação Potência	139
B.1.1	<i>Amostrando os Efeitos Aleatórios Espaciais</i>	141
B.2	Funções de Ligação Reversa de Potência	142
B.2.1	<i>Amostrando os Efeitos Aleatórios Espaciais</i>	144
B.3	Funções de Ligação Assimétrica Transmutada	144
B.3.1	<i>Amostrando os Efeitos Aleatórios Espaciais</i>	146
	APÊNDICE C CAPÍTULO 2	149
C.1	Curvas das Fda's e Fdp's dos Modelos L-P, L-RP, N-P, N-RP, GR-P e GR-RP	149
	APÊNDICE D CAPÍTULO 3	151
D.1	Modelos Simultaneamente Autoregressivos (SAR)	151
D.2	Distribuições Assimétricas Transmutadas	152
D.3	Estudo de Simulação das Métricas de Predição	156
D.4	<i>Trace-plot</i> e gráficos de Autocorrelação dos Modelos Ajustados	160
D.5	Probabilidades Preditivas	161
D.6	Estudo de Simulação - Modelo CAR	161
	APÊNDICE E CAPÍTULO 4	165
E.1	Comparação de Modelos segundo Métricas de Qualidade da Predição	165

E.2	Modelos G-AT e GR-AT Aplicados aos Dados de Periodontite	166
APÊNDICE F	CAPÍTULO 5	167

INTRODUÇÃO

Em muitos problemas científicos, é comum analisar um experimento ou fenômeno natural cujo resultado é dicotômico. Por exemplo, em mercado de crédito, bancos estão interessados em avaliar as chances de um cliente ser adimplente (ou inadimplente) no momento da concessão de um dado empréstimo; uma semente deve germinar ou não germinar sob certas condições experimentais; um inseto submetido a um teste de inseticida deve sobreviver ou morrer quando exposto a uma certa dose; na área da saúde, pacientes, quando submetidos a um teste clínico para comparar formas alternativas de tratamento podem ou não apresentar os sintomas da doença após um período de testagens (COLLETT, 2002). Como é conhecido na literatura, dados com essa característica são nomeados de dados de resposta binária¹ e é comum denominarmos genericamente, os dois resultados possíveis da variável resposta como sucesso (evento de interesse) e fracasso.

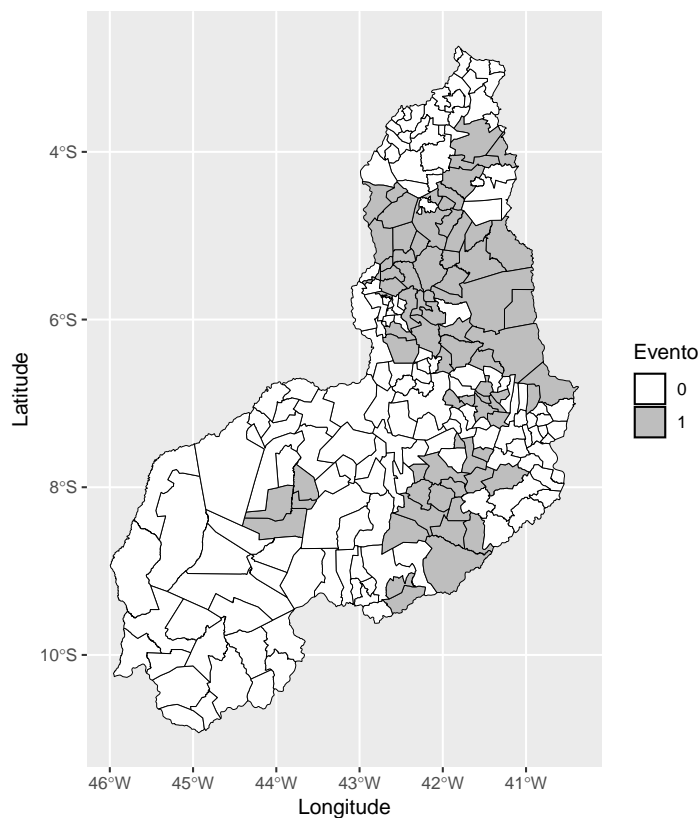
Dentre as formas de analisar esse tipo de dado, modelos de regressão binária são ferramentas muito úteis, nos permitindo relacionar a probabilidade de sucesso com as variáveis de interesse (covariáveis) disponíveis (LI *et al.*, 2016). No trabalho aplicado de Jaber, Juhász e Csonka (2021), por exemplo, os autores investigaram como as lesões de ciclistas, classificadas em “maior gravidade” e “menor gravidade”, estavam associadas à infraestrutura viária e as condições de superfície através de regressão binária. Entre as conclusões apontadas, os autores destacaram que para as covariáveis relacionadas à superfície da via, como estradas boas, pavimentadas e sinalizadas estão associadas a uma maior probabilidade de lesões graves devido à expectativa de maior frequência de ciclistas em estradas com boas condições. Além desses exemplos, aplicações em áreas como Ciências Sociais, Medicina, Física, Epidemiologia, Sistemas de Informação e Biomedicina também podem ser citadas. Para análise mais profunda e detalhada de modelos de regressão binária, indicamos ao leitor o trabalho de Collett (2002). Nesta tese, trabalharemos especificamente nesta classe de modelos de regressão cuja variável resposta é do tipo binária.

¹ apesar de o termo mais antigo ser *quantal*, conforme cita (COLLETT, 2002)

Dados binários podem ser observados em muitos contextos, apresentando características intrínsecas da própria natureza do problema em questão, sendo estas informações importantes para a modelagem dos dados e para compreender o fenômeno em estudo. Dados binários espaciais seriam um exemplo no qual a natureza espacial das observações entra como fonte de informação no processo de modelagem dos dados. Segundo [Banerjee, Carlin e Gelfand \(2003\)](#), existem alguns tipos de dados espaciais. Neste trabalho, abordaremos os chamados dados de área.

Na [Figura 1](#) temos a ilustração de um conjunto de dados binários de natureza espacial. Estes dados referem-se a um compilado de 30 anos de observação para cada município do estado brasileiro da Piauí sobre a informação do ‘trimestre menos chuvoso’. Os dados originais foram categorizados atribuindo valor 1 aos municípios cujo ‘trimestre menos chuvoso’ foi ‘AGOSTO-SETEMBRO-OUTUBRO’ e zero caso contrário. Os dados foram extraídos de [Alvares *et al.* \(2013\)](#), e disponibilizados na base de dados espaciais da Embrapa (<https://geoinfo.dados.embrapa.br/>). Perceba que os municípios que apresentam o evento de interesse formam certos padrões de agrupamento, revelando que a natureza espacial dos dados pode ser uma informação importante no processo de modelagem.

Figura 1 – Trimestre menos chuvoso (‘AGOSTO-SETEMBRO-OUTUBRO’=1) para cada município do estado do Piauí ao longo de 30 anos.



Fonte: Elaborada pelo autor

1.1 Regressão Binária com Função de Ligação Assimétrica

Em modelos de regressão binária, a especificação da função de ligação desempenha papel importante no desenvolvimento das análises. Normalmente, as funções de ligação surgem como a função inversa de uma função de distribuição acumulada (fda) (BAZÁN; BOLFARINE; BRANCO, 2010).

Quando a fda é simétrica, a curva de resposta de probabilidade fornecida possui forma simétrica em torno de 0,5. Atualmente, as funções Logito e Probit são dois exemplos de funções de ligação que podem ser consideradas simétricas, e dão origem aos modelos de regressão binária mais conhecidos na literatura. Contudo, assumir função de ligação simétrica em algumas situações nem sempre agrega qualidade de ajuste do modelo aos dados, e esta deficiência tende a aumentar quando o número de observações nas categorias de respostas são significativamente diferentes umas das outras, dando origem aos chamados dados desbalanceados (*imbalanced data*).

Autores como Chen, Dey e Shao (1999), Bazán, Bolfarine e Branco (2010) já sinalizaram sobre os problemas inerentes à inferências dos dados quando há má especificação da função de ligação. Czado e Santner (1992), por exemplo, mostram que há um aumento substancial no viés e no erro quadrático médio das estimativas dos coeficientes de regressão, quando se assume falsamente uma função de ligação Logística para o modelo. Wang e Dey (2011) vão além, e argumentam que esses efeitos indesejáveis são de maior magnitude quando a má especificação envolvida está mais relacionada à assimetria dos dados do que à curtose.²

Bazán, Bolfarine e Branco (2010) e Huayanay *et al.* (2019) argumentam que, de uma forma geral, quando a probabilidade de uma dada resposta binária aproxima-se de 0 em uma taxa diferente da qual ela se aproxima de 1, funções de ligação assimétricas são mais úteis para análise de dados binários que as funções de ligação simétricas. Por isso, visando fornecer alternativas satisfatórias, muitos esforços são desenvolvidos na literatura para se propor funções de ligação mais flexíveis no que tange à modelagem de dados categóricos (WANG; DEY, 2011). Prentice (1976), por exemplo, propôs uma função de ligação baseada na função de distribuição F-Snedecor, na qual as funções de ligação Logito, Clog-log, Log-log³, Probit, Laplace e Exponencial entram como casos particulares; Aranda-Ordaz (1981) introduziu dois modelos uniparamétricos para tratar de casos com a presença e a ausência de assimetria nos desvios (*departures*) em um modelo de regressão binária logístico; Guerrero e Johnson (1982) e Morgan (1983) também apresentaram modelos uniparamétricos, em que os primeiros propõem uma transformação de Box-Cox na *odds ratio* com o intuito de apresentar uma generalização do modelo Logístico, enquanto o segundo apresentou um modelo Logístico cúbico para modelar os desvios simétricos da curva Logística.

Mais adiante, Stukel (1988) propôs uma classe de modelos com dois parâmetros para

² curtose está relacionada ao peso da(s) cauda(s) da distribuição.

³ As funções de ligação Clog-log e Log-log são assimétricas de valor fixo, sendo a primeira assimétrica à direita, e a segunda, assimétrica à esquerda.

estender o uso do modelo de regressão Logístico padrão para curvas de probabilidade assimétricas. Contudo, na presença de covariáveis, esse modelo produz distribuições a posteriori impróprias para muitos tipos de distribuições a priori não-informativas. [Chen, Dey e Shao \(1999\)](#) também propuseram uma classe de modelos de ligação assimétrica para dados de resposta binária, permitindo, porém, lidar com uma ampla variedade de distribuições a posteriori próprias para os coeficientes de regressão, utilizando prioris impróprias padrão.

[Kim, Chen e Dey \(2008\)](#) introduziram uma nova classe de modelos de função de ligação assimétrica (*GT-link*) baseada na distribuição *t* generalizada. [Wang e Dey \(2011\)](#) argumentam que apesar dessa classe de funções de ligação ser generalizada para muitos problemas nos quais há a presença de assimetria nos dados, a restrição no parâmetro de forma δ , com $0 < \delta < 1$, presente no modelo, reduz grandemente uma variedade de assimetrias possíveis. Por isso, os autores propõem uma nova classe de funções de ligação assimétricas baseada na distribuição de valores extremos generalizada (*generalized extreme value - GEV*) com a incorporação de um parâmetro de forma ξ irrestrito, permitindo, assim, uma ampla gama de assimetrias.

Trabalhos posteriores como o de [Bazán, Bolfarine e Branco \(2010\)](#) revisaram funções de ligação assimétricas e apresentaram uma abordagem unificada para duas ligações Probit assimétricas propostas na literatura; [Li et al. \(2016\)](#) propuseram um modelo de regressão binária não-paramétrica flexível baseada na função de ligação GEV adicionada de uma priori de processo Gaussiano para a função de regressão latente. [Bazán et al. \(2017\)](#) e [Lemonte e Bazán \(2018\)](#) apresentaram modelos de regressão binária com funções de ligação potência e reversa de potência, introduzindo um parâmetro de assimetria. [Huayanay et al. \(2019\)](#) fizeram um estudo de comparação de performance para várias funções de ligação assimétricas. [Alves, Bazán e Arellano-Valle \(2023\)](#) introduziram funções de ligação Clog-log flexíveis para modelos de regressão binomial que incluem um parâmetro extra para explicar algum nível de desbalanceamento.

Neste trabalho, investigaremos o desempenho da nossa classe de modelos de regressão binária sob as funções de ligação potência e reversa de potência apresentadas por [Bazán et al. \(2017\)](#) e [Lemonte e Bazán \(2018\)](#).

No que tange aos procedimentos de inferência para os modelos de regressão binária, algumas abordagens são apresentadas na literatura. [Aranda-Ordaz \(1981\)](#), [Guerrero e Johnson \(1982\)](#), [Lemonte e Bazán \(2018\)](#)⁴ apresentaram métodos inferenciais baseados em abordagem frequentista utilizando máxima verossimilhança. Trabalhos como os de [Chen, Dey e Shao \(1999\)](#), [Kim, Chen e Dey \(2008\)](#) e [Wang e Dey \(2011\)](#) foram desenvolvidos por uma metodologia bayesiana. [Huayanay et al. \(2019\)](#), buscando uma melhor eficiência na amostragem dos parâmetros do modelo, optaram por uma abordagem bayesiana baseada no método de Monte Carlo Hamiltoniano - HMC (*Hamiltonian Monte Carlo*) ([DUANE et al., 1987](#); [BETANCOURT, 2017](#))- utilizando o amostrador No-U-Turn ([HOFFMAN; GELMAN et al., 2014](#)). Para a classe de modelos proposta neste trabalho, utilizaremos uma metodologia completamente bayesiana.

⁴ o método IRLS (*Iteratively Reweighted Least Squares*) é um desses exemplos ([COLLETT, 2002](#))

Definida a classe de funções de ligação a ser utilizada, descrevemos, na próxima Seção, a forma como abordaremos a modelagem da parte espacial dos nossos modelos.

1.2 Modelagem para Dados Espaciais

Dados de área são um tipo especial de dados espaciais e consistem de um conjunto de sub-regiões finitas com fronteiras bem definidas, em que cada uma delas possui uma medida agregada de sua população (MORRIS *et al.*, 2019). Não é difícil perceber, em uma investigação epidemiológica, a importância de descrever o comportamento de taxas de determinadas doenças, como câncer de pulmão, por município e ano em um estado específico. Além da epidemiologia, trabalhos produzidos em campos como Geologia, Processamento de Imagens e Saúde, são outros exemplos que podem ser citados (CRESSIE, 1993).

A modelagem para dados agregados por região normalmente assume a forma de um modelo hierárquico, no qual especificamos efeitos aleatórios latentes para cada sub-região. Dessa forma, os efeitos espaciais aleatórios entram em um segundo estágio de uma especificação hierárquica bayesiana, e estes efeitos aleatórios são modelados conjuntamente como tendo distribuição Normal multivariada, com vetor de médias nulo e matriz de precisão Ω , que incorpora a dependência espacial entre as sub-regiões. Com isso, os conhecidos modelos autoregressivos CAR (*Conditional autoregressive*) e ICAR (*intrinsic conditional autoregressive*) são usados como especificações de distribuição a priori para a matriz de precisão. Exemplos desse uso podem ser encontrados na literatura em modelagem de risco de doença, por exemplo, em que se utiliza um modelo Poisson log-normal incrementado de um componente ICAR para capturar o efeito aleatório espacial e um componente para capturar a heterogeneidade não-espacial (MORRIS *et al.*, 2019).

Sob as especificações dos modelos CAR ou ICAR, a matriz de precisão condiciona a estrutura de vizinhança, intrínseca ao problema espacial, a uma forma paramétrica. No entanto, um tratamento totalmente bayesiano pode ser oferecido ao problema quando em vez de assumirmos uma forma paramétrica conhecida para a matriz de precisão, atribuímos uma distribuição a priori para Ω para representar a incerteza sobre a relação espacial entre os efeitos aleatórios, e assim refletir a intuição de que a correlação espacial não segue uma simples forma paramétrica.

Nesse enfoque, a distribuição G-Wishart é uma das alternativas disponíveis como distribuição a priori para modelos de grafos Gaussianos (ROVERATO, 2002). Essa distribuição é uma priori conjugada e possui aplicação em problemas como os da área da Física, Biomedicina e Tecnologia. Em nossa pesquisa, incrementaremos, aos modelos de regressão binários propostos neste projeto, componentes de efeitos aleatórios espaciais. Porém, em vez de assumir uma forma paramétrica padrão CAR (ou ICAR) para Ω , utilizaremos uma abordagem mais flexível ao fazer $\Omega \sim \text{G-Wishart}_{\mathcal{W}}(\kappa, \mathcal{S})$, em que G-Wishart significa uma distribuição G-Wishart, κ é o parâmetro referente aos graus de liberdade e \mathcal{S} é uma matriz escala restrita a ter entradas nulas

para cada zero na matriz de adjacência W .

Dentre os modelos paramétricos para a matriz de precisão Ω , além dos modelos CAR e ICAR citados acima, o modelo DAGAR (*directed acyclic graph auto-regressive*) (DATTA *et al.*, 2019) é mais recente na literatura e possui maior escalabilidade quando o número de sub-regiões é grande. Além disso, apresenta desempenho tão bom quanto os modelos paramétricos tradicionais como o modelo CAR. Este modelo também agrega interpretabilidade aprimorada, produzindo boa recuperação dos efeitos aleatórios espaciais (DATTA *et al.*, 2019). Ofereceremos um estudo de simulação inicial com um dos modelos trabalhados nesta tese, sob este modelo espacial, no último Capítulo (Capítulo 5).

Antes de finalizar a Seção, convém destacar que dados binários em cenários comuns normalmente apresentam problema de desbalanceamento, não sendo diferente no contexto espacial (Veja a Figura 1). Os dados trabalhados nesta tese, por exemplo, apresentados nas Seções 3.5 e 4.1, apresentam baixíssima proporção do evento de interesse. Poderíamos ainda conjecturar dados espaciais em um cenário de doença rara, no qual naturalmente o número de casos de interesse por unidade de área seria pequeno. Enfim, esses são exemplos de um conjunto de muitas possibilidades.

1.3 Proposta do Trabalho e Objetivos

A proposta desta tese consiste em desenvolver uma metodologia de inferência bayesiana para modelos de regressão binária espacial sob esparsidade, utilizando funções de ligação assimétricas. Dois tipos de modelos são apresentados: i) o primeiro é aplicável a uma única região composta por m sub-regiões; ii) o segundo estende o primeiro para r regiões assumidas idênticas, cada uma contendo m sub-regiões. Nesse contexto, o modelo do primeiro caso é recuperado como uma situação especial quando $r = 1$. As funções de ligação utilizadas são as que foram apresentadas nos trabalhos de Bazán *et al.* (2017) e Lemonte e Bazán (2018), ou seja, a classe de funções de ligação assimétricas potência e reversa de potência.

A estrutura da regressão binária dos modelos propostos assume que os dados podem ser referenciados no espaço, dando origem a um modelo de natureza espacial. A codificação do relacionamento espacial será estabelecida por meio dos modelos mais recentes no segmento. Primeiro, modelamos a matriz de precisão Ω da distribuição dos efeitos aleatórios espaciais por meio de uma abordagem bastante flexível, especificando uma distribuição a priori G-Wishart, isto é, $\Omega \sim G\text{-Wishart}_W(\kappa, \mathcal{S})$ de acordo com Roverato (2002). Em um segundo momento, como forma de análise preliminar, atribuímos à matriz de precisão o modelo paramétrico DAGAR como distribuição a priori (DATTA *et al.*, 2019).

Explanadas as características pertinentes aos modelos propostos, as principais contribuições desta tese, de forma sucinta, são:

1. apresentar uma classe de modelos de regressão binária espacial para dados desbalanceados, utilizando funções de ligação assimétricas;
2. apresentar uma metodologia de inferência completamente bayesiana para o trato desta nova classe de modelos, levando em consideração a eficiência computacional na amostragem dos parâmetros e a natureza esparsa da matriz de precisão;
3. apresentar uma análise de desempenho dos modelos propostos, considerando as funções de ligação selecionadas e os modelos espaciais especificados para Ω para fins de comparação; nesta análise de desempenho, serão avaliadas a robustez, a qualidade do ajuste e performance preditiva dos modelos estudados;
4. apresentar, dentro desse contexto de análise de desempenho de modelos, ferramentas diagnósticos que também serão utilizadas para detecção de observações influentes no conjunto de dados. Essas ferramentas são baseadas em trabalhos recentes na literatura, tais como [Goh e Dey \(2014\)](#) e [Danilevicz e Ehlers \(2022\)](#).

1.4 Estrutura do Trabalho

No Capítulo 2, apresentamos inicialmente o fundamento da abordagem bayesiana, os modelos de regressão binária bayesiana com funções de ligação assimétricas, e as métricas de avaliação de desempenho dos modelos baseadas na qualidade do ajuste e na qualidade da predição. Estas últimas são as métricas popularmente conhecidas no contexto de regressão binária.

No Capítulo 3 apresentamos uma revisão sucinta dos modelos para dados de área, o primeiro dos modelos de regressão binária espacial abordados neste trabalho, bem como um estudo de simulação que avaliará o desempenho do modelo proposto sob as diferentes funções de ligação assimétricas adotadas, e sob diferentes cenários; apresentamos também uma aplicação a dados reais.

Chegando ao Capítulo 4, descrevemos uma versão generalizada do modelo de regressão apresentado no Capítulo 3, bem como um estudo de simulação do modelo proposto e uma aplicação a um conjunto de dados reais.

Apresentamos um estudo de simulação com o modelo espacial DAGAR no Capítulo 5 e finalizamos esta tese no Capítulo 6, discorrendo sobre as produções e conclusões decorrentes deste trabalho, bem como propostas de projetos de pesquisa futuros.

CONCEITOS PRELIMINARES

Neste Capítulo, apresentamos os conceitos fundamentais para o desenvolvimento das metodologias de análises e obtenção de resultados ao longo desta tese. Na Seção 2.1, apresentamos a noção básica de inferência bayesiana; na Seção 2.2, alguns dos famosos métodos de amostragem MCMC (*Markov Chain Monte Carlo*); os modelos potência e reversa de potência que dão origem às funções de ligação flexíveis adotadas neste trabalho são apresentados na Seção 2.3; o modelo de regressão binária bayesiano, na Seção 2.4. Apresentamos as métricas utilizadas para diagnóstico e seleção de modelos na Seção 2.5; a metodologia de detecção de *outliers* bayesiana é abordada na Seção 2.6; na Seção 2.7, temos as métricas de avaliação de desempenho de modelos baseada na predição.

2.1 Inferência Bayesiana

A classe de modelos trabalhada nesta tese será analisada sob enfoque bayesiano. A inferência bayesiana possui uma vasta literatura e apresenta muitas atratividades, como a interpretabilidade intuitiva e adequação a muitos problemas complexos. A abordagem bayesiana permite uma estrutura coesa para combinar modelos de dados complexos e conhecimento externo aos dados do experimento em estudo.

Apresentemos tal abordagem iniciando pelo caso mais simples, o univariado. Assuma que haja uma quantidade θ de interesse, e que essa quantidade, tipicamente não observável e desconhecida, pode representar alguma característica da população em estudo. A abordagem bayesiana postula que o conhecimento a priori que se tem sobre θ , incorporado através de uma distribuição de probabilidade $p(\theta)$, chamada distribuição a priori (*prior distribution*), possa ser aumentado após observar uma amostra de tamanho n da variável aleatória Y relacionada a θ , ou seja, $\mathbf{y} = (y_1, \dots, y_n)^\top$.

A distribuição a priori $p(\theta)$ reflete o grau de incerteza que se tem sobre θ e, consequen-

temente, esta informação variará de analista para analista (produzindo modelos distintos). Sob o enfoque bayesiano, tanto as quantidades observáveis quanto os parâmetros de um modelo estatístico são considerados quantidades aleatórias. A regra de atualização do conhecimento a priori é dada pelo teorema de Bayes, e basicamente une a informação advinda da função de verossimilhança $l(\theta; \mathbf{y}) = p(\mathbf{y}|\theta)$ e da informação a priori $p(\theta)$. Logo, se θ for contínuo, temos

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}, \theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)p(\theta)}{\int p(\theta, \mathbf{y})d\theta}, \quad (2.1)$$

em que $p(\theta|\mathbf{y})$ é chamada de distribuição a posteriori (*posterior distribution*) e $p(\mathbf{y})$ é a distribuição marginal de \mathbf{y} , e funciona como uma espécie de constante de normalização para a distribuição em (2.1). Se θ for discreto, a fórmula em (2.1) pode ser expressa em

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}, \theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)p(\theta)}{\sum_{\theta} p(\theta, \mathbf{y})}, \quad (2.2)$$

em que o somatório se estende sobre o intervalo de todos os possíveis valores de θ .

Em (2.1), a constante $p(\mathbf{y})$ não depende de θ , e a expressão da distribuição a posteriori resultante fica especificada a menos de uma constante de normalização, que é a chamada distribuição a posteriori não normalizada. Assim, a expressão em (2.1) pode ser resumida na seguinte fórmula

$$p(\theta|\mathbf{y}) \propto l(\theta; \mathbf{y})p(\theta), \quad (2.3)$$

em que $p(\theta|\mathbf{y})$ é proporcional à verossimilhança vezes a distribuição a priori. A expressão em (2.3) é bastante útil, por exemplo, em processos de estimação, em que necessitamos apenas da expressão proporcional da posteriori para simular valores de sua distribuição. Esta será a abordagem utilizada quando tratarmos sobre o processo de estimação dos parâmetros relacionados aos modelos trabalhados nesta tese. Para obtenção de mais detalhes sobre estatística bayesiana, os leitores podem consultar [Gelman *et al.* \(2004\)](#) e [Gamerman e Lopes \(2006\)](#).

2.2 Monte Carlo via Cadeias de Markov

Métodos de Monte Carlo via cadeias de Markov, conhecidos abreviadamente por métodos MCMC (*Markov Chain Monte Carlo*), são métodos que utilizam as propriedades das cadeias de Markov¹ para simular o comportamento de variáveis aleatórias (ou vetores aleatórios). Na literatura, há uma vasta aplicação de métodos MCMC, em que diferentes estratégias têm sido propostas para poder simular cadeias de Markov (CM) cuja distribuição de equilíbrio é a distribuição pretendida, buscando-se, por exemplo, dentre alguns objetivos, eficiência computacional. Referências como [Gamerman e Lopes \(2006\)](#) podem ser consultadas para obtenção de mais informações.

¹ Cadeias de Markov é um típico processo estocástico em que a distribuição de X_{t+1} , condicionada no histórico passado X_t, X_{t-1}, \dots, X_1 , depende apenas do estado imediatamente anterior, ou seja, X_t .

Uma situação bastante comum em modelos bayesianos é que a distribuição a posteriori não pode ser trabalhada analiticamente, seja pelo fato de $p(\mathbf{y})$ em (2.1) não possuir uma forma analítica explícita, ou porque a posteriori não possui forma padrão (GAMERMAN; LOPES, 2006). Nesta situação, métodos MCMC se tornam ferramentas poderosas, pois nos permitem trabalhar com a distribuição a posteriori na sua forma não normalizada, tal como em (2.3).

O fundamento dos métodos MCMC consiste basicamente em simular valores da(s) variável (eis) aleatória (s) de interesse por meio das CM's, e utilizar as propriedades estatísticas das amostras geradas para fazer aproximações das propriedades da distribuição a posteriori. Basicamente, utilizamos o resultado da lei forte dos grandes números cujo resultado nos diz que, para amostras independentes e identicamente distribuídas (iid) da distribuição $p(\theta|\mathbf{y})$, a média amostral converge quase certamente (*almost surely* - a.s) para a esperança verdadeira, ou seja, de posse de uma amostra particular $\theta_1, \dots, \theta_T$ gerada da distribuição $p(\theta|\mathbf{y})$, e sabendo que $E_{p(\theta|\mathbf{y})}[h(\theta)] = \int h(\theta)p(\theta|\mathbf{y})d\theta$, temos

$$\frac{1}{T} \sum_{i=1}^T h(\theta_i) \xrightarrow{a.s} E_{p(\theta|\mathbf{y})}[h(\theta)], \text{ quando } T \rightarrow \infty, \quad (2.4)$$

em que h é qualquer função dada. A partir de 2.4, chegamos ao seguinte resultado:

$$\int h(\theta)p(\theta|\mathbf{y})d\theta \approx \frac{1}{T} \sum_{i=1}^T h(\theta_i). \quad (2.5)$$

Após um dado valor inicial θ_0 , a CM iniciará deste valor e varrerá todo o suporte da distribuição a posteriori até alcançar a distribuição de equilíbrio. O período inicial de simulação antes da cadeia alcançar a distribuição de equilíbrio é chamado de *burn-in*. Uma vez que a cadeia alcançou a distribuição de equilíbrio, os valores gerados podem ser retirados da cadeia a um certo valor espaçado, no qual os valores tomados podem ser considerados amostras aproximadamente independentes da distribuição a posteriori de θ (GAMERMAN; LOPES, 2006).

2.2.1 Metropolis-Hastings

Um procedimento geral, conhecido em métodos MCMC para simular uma CM cuja distribuição de equilíbrio é a distribuição a posteriori de interesse, é o método de Metropolis-Hastings - MH. O método de MH consiste basicamente em descrever como o próximo estado da CM pode ser obtido. Para fins de ilustração, assuma que θ^t é o estado atual da cadeia. Então, o estado candidato θ^* é gerado de uma distribuição proposta $q(\theta^*|\theta^t)$, em que a proposta é aceita $\theta^{t+1} \leftarrow \theta^*$ como o próximo estado consecutivo da cadeia de Markov se

$$\min \left\{ 1, \frac{p(\theta^*|\mathbf{y}) q(\theta^t|\theta^*)}{p(\theta^t|\mathbf{y}) q(\theta^*|\theta^t)} \right\} \geq u, \quad (2.6)$$

em que u é uma amostra da variável aleatória uniforme no intervalo $(0, 1)$ (GAMERMAN; LOPES, 2006). Caso contrário, a proposta é rejeitada e atribui-se θ^t ao próximo estado da

Algoritmo 1 – Metropolis-Hastings

- 1: Especifique um valor inicial $\theta^{(1)}$ e faça $t=1$.
- 2: Dada a posição atual $\theta^{(t)}$, gere um novo valor θ^* de $q(\cdot|\theta^{(t)})$.
- 3: Calcule a probabilidade de aceitação $\alpha(\theta^{(t)}, \theta^*) = \min \left\{ 1, \frac{p(\theta^*|\mathbf{y})q(\theta^{(t)}|\theta^*)}{p(\theta^{(t)}|\mathbf{y})q(\theta^*|\theta^{(t)})} \right\}$ e gere $u \sim U(0, 1)$
- 4: Se $u \leq \alpha$, então aceite o novo valor e faça $\theta^{(t+1)} = \theta^*$. Caso contrário rejeite e faça $\theta^{(t+1)} = \theta^{(t)}$
- 5: Faça $t = t + 1$ e volte ao passo 2.

cadeia, ou seja, $\theta^{t+1} \leftarrow \theta^t$. No algoritmo 1, tem-se um pseudo-código que resume o método de Metropolis-Hastings em alguns passos.

Aplicando o procedimento acima um número suficientemente grande de vezes, digamos T ; retirada a amostra de aquecimento de tamanho k , $\theta^{(k+1)}, \theta^{(k+2)}, \dots, \theta^{(T)}$ é uma amostra da distribuição a posteriori de θ .

2.2.2 Amostrador de Gibbs

Em algumas situações em que se tem um vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^\top$ de dimensão d para amostrar valores, pode ser conveniente atualizá-lo de forma sucessiva, fazendo com que cada elemento do vetor de parâmetros sejam atualizados um de cada vez. Esse é justamente o caso do amostrador de Gibbs (*Gibbs sampling*).

No amostrador de Gibbs, cada elemento de $\boldsymbol{\theta}$ é atualizado um a um utilizando valores gerados das distribuições condicionais completas $p(\theta_j|\boldsymbol{\theta}_{-j}, \mathbf{y})$, em que $\boldsymbol{\theta}_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d)^\top$, para $j = 1, \dots, d$. Em um caso geral, θ_j pode ser um subvetor de $\boldsymbol{\theta}$.

A distribuição condicional completa $p(\theta_j|\boldsymbol{\theta}_{-j}, \mathbf{y})$ é a distribuição do j -ésimo componente de $\boldsymbol{\theta}$ condicionada nos demais componentes. O amostrador de Gibbs possui a vantagem de não termos de trabalhar com a distribuição a posteriori conjunta de $\boldsymbol{\theta}$, em vez disso, permite-nos trabalhar com distribuições univariadas (ou pelo menos com distribuições que possuem dimensões menores, caso θ_j seja um subvetor de $\boldsymbol{\theta}$) e evita a necessidade do passo de aceitação-rejeição apresentado na Seção anterior pelo método de Metropolis-Hastings, pois neste caso todos os valores propostos são sempre aceitos. É fácil verificar isso ao substituir a distribuição condicional completa para a j -ésima componente na regra de Metropolis-Hastings em (2.6). Assim, seja θ_j^* o valor proposto pela distribuição condicional completa de θ_j , então

$$\frac{p(\theta_j^*, \boldsymbol{\theta}_{-j}^t | \mathbf{y}) p(\boldsymbol{\theta}^t | \theta_j^*, \boldsymbol{\theta}_{-j}^t)}{p(\boldsymbol{\theta}^t | \mathbf{y}) p(\theta_j^*, \boldsymbol{\theta}_{-j}^t | \boldsymbol{\theta}^t)} = 1. \quad (2.7)$$

Desta forma, podemos concluir que o amostrador de Gibbs é um caso especial do método de Metropolis-Hastings. O algoritmo 2 resume o método apresentado pelo amostrador de Gibbs.

Após iterar o algoritmo 2 T vezes, retirada a amostra de aquecimento de tamanho k , temos $(\theta_1, \dots, \theta_d)^{k+1}, \dots, (\theta_1, \dots, \theta_d)^T$ como amostra da distribuição a posteriori de $\boldsymbol{\theta}$. Na literatura,

Algoritmo 2 – Amostrador de Gibbs

- 1: Inicialize $\boldsymbol{\theta}^{(1)} = (\theta_1, \dots, \theta_d)^\top$ e faça $t = 1$
- 2: Obtenha um novo valor $\boldsymbol{\theta}^{(t)}$ a partir de $\boldsymbol{\theta}^{(t-1)}$ através da geração sucessiva dos valores

$$\begin{aligned}\theta_1^{(t)} &\sim p(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_d^{(t-1)}) \\ \theta_2^{(t)} &\sim p(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_d^{(t-1)}) \\ &\vdots \\ \theta_d^{(t)} &\sim p(\theta_d | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{d-1}^{(t)})\end{aligned}$$

- 3: Faça $t = t + 1$ e volte ao passo 2

alguns algoritmos MCMC são propostos utilizando variações e combinações dos métodos de Metropolis-Hastings e do amostrador de Gibbs. Algumas possibilidades, por exemplo, seriam utilizar o amostrador de Gibbs por passos de Metropolis, em que a distribuição condicional completa de cada parâmetro seria a distribuição de interesse no algoritmo de MH; também seria possível utilizar o Metropolis-Hastings em blocos, assim $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^\top$ poderia ser dividido em B blocos, e cada um deles amostrado por vez pelo algoritmo de Metropolis-Hastings (GAMERMAN; LOPES, 2006).

2.2.3 Monte Carlo Hamiltoniano

Em busca de maior eficiência computacional, alguns métodos MCMC têm sido propostos com o intuito de otimizar o processo de convergência para a distribuição de equilíbrio, sem a necessidade de ter que se gerar um tamanho excessivamente grande da cadeia de Markov. Métodos como o de Metropolis-Hastings que possuem mecanismo de geração dos valores propostos do tipo caminho aleatório (*random walk*), podem ser bastante ineficientes em explorar o espaço paramétrico da distribuição a posteriori dos parâmetros.

O método de Monte Carlo Hamiltoniano (HMC), proposto originalmente com o nome de Monte Carlo Híbrido por Duane *et al.* (1987) para simulação de dinâmica molecular, possui um mecanismo diferente para gerar os valores candidatos da cadeia de Markov ao utilizar a informação das derivadas das distribuições a posteriori dos parâmetros. Através da informação das derivadas, a cadeia pode ser direcionada para as regiões de maior densidade a posteriori. Dessa forma a convergência para as distribuições de equilíbrio se torna mais rápida.

Basicamente, o método HMC se fundamenta em fazer simulações em um sistema físico fictício para gerar estados propostos sob um esquema de Metropolis-Hastings. O estado proposto é simulado usando simulação discreta da dinâmica hamiltoniana. Para uma revisão mais detalhada do método HMC, os leitores podem consultar Betancourt (2017) e Neal (2010).

Seja $\boldsymbol{\theta} \in \mathbb{R}^d$ a expressão que denota um vetor de parâmetros d -dimensional; $p(\boldsymbol{\theta} | \mathbf{y})$, a distribuição a posteriori de $\boldsymbol{\theta}$ e $\mathbf{p} \in \mathbb{R}^d$, um vetor de parâmetros auxiliares que são independentes

de $\boldsymbol{\theta}$. Sob a dinâmica hamiltoniana, o vetor de parâmetros $\boldsymbol{\theta}$ é interpretado como o vetor de posição da “partícula” e $-\log p(\boldsymbol{\theta}|\mathbf{y})$ como sua energia potencial; \mathbf{p} é interpretado como o momento da partícula e $\mathbf{p}^\top \mathbf{M}^{-1} \mathbf{p}$, como sua energia cinética. A função hamiltoniana que descreve o sistema fictício é dada por

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\mathcal{L}(\boldsymbol{\theta}) + \mathbf{p}^\top \mathbf{M} \mathbf{p} / 2, \quad (2.8)$$

em que $-\mathcal{L}(\boldsymbol{\theta}) = \log p(\boldsymbol{\theta}|\mathbf{y})$. Ao invés de gerarmos amostras da distribuição a posteriori de $\boldsymbol{\theta}$, geramos amostras da distribuição conjunta de $(\boldsymbol{\theta}, \mathbf{p})$, que possui a seguinte proporcional:

$$f(\boldsymbol{\theta}, \mathbf{p}) \propto p(\boldsymbol{\theta}|\mathbf{y}) \exp(-\mathbf{p}^\top \mathbf{M}^{-1} \mathbf{p} / 2) \propto \exp[-H(\boldsymbol{\theta}, \mathbf{p})]. \quad (2.9)$$

Para tempo contínuo t , amostrando \mathbf{p} de uma distribuição Normal multivariada $N_d(\mathbf{0}, \mathbf{M})$, em que \mathbf{M}^2 é a chamada matriz de massa, $(\boldsymbol{\theta}, \mathbf{p})$ resulta no estado atual da cadeia. Então, a evolução do sistema é dada pelas equações diferenciais de Hamilton

$$\begin{aligned} \frac{\partial \boldsymbol{\theta}}{\partial t} &= \frac{\partial H(\boldsymbol{\theta}, \mathbf{p})}{\partial \mathbf{p}} = \mathbf{M}^{-1} \mathbf{p} \\ \frac{\partial \mathbf{p}}{\partial t} &= -\frac{\partial H(\boldsymbol{\theta}, \mathbf{p})}{\partial \boldsymbol{\theta}} = -\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}), \end{aligned} \quad (2.10)$$

em que $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$ é o gradiente de $\mathcal{L}(\boldsymbol{\theta})$ com relação a $\boldsymbol{\theta}$. Porém, devido ao fato de as equações diferenciais acima não poderem ser resolvidas analiticamente, métodos de resolução numérica necessitarão ser empregados para discretizar a dinâmica hamiltoniana. Um método comumente empregado, e que garante a reversibilidade do valor proposto (NEAL, 2010), é o *Leapfrog*. Esse método discretiza a dinâmica hamiltoniana nos seguintes passos:

$$\begin{aligned} \mathbf{p}^{(\tau+\epsilon/2)} &= \mathbf{p}^{(\tau)} + (\epsilon/2) \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^\tau) \\ \boldsymbol{\theta}^{(\tau+\epsilon)} &= \boldsymbol{\theta}^{(\tau)} + \epsilon \mathbf{p}^{(\tau+\epsilon/2)} \\ \mathbf{p}^{(\tau+\epsilon/2)} &= \mathbf{p}^{(\tau+\epsilon/2)} + (\epsilon/2) \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^{(\tau+\epsilon)}), \end{aligned} \quad (2.11)$$

para um tamanho de passo $\epsilon > 0$ especificado. Após um dado número L de passos de tempo³, isto resulta em uma proposta $(\boldsymbol{\theta}^*, \mathbf{p}^*)$. Contudo, a discretização do sistema apresentada acima introduz um erro, e a aceitação do estado proposto precisa ser decidida pela regra de Metropolis-Hastings. Como mencionado antes, a distribuição conjunta de $(\boldsymbol{\theta}, \mathbf{p})$ é a distribuição alvo, e o estado proposto $(\boldsymbol{\theta}^*, \mathbf{p}^*)$, ao final da simulação do sistema, é aceito com probabilidade

$$\begin{aligned} \alpha[(\boldsymbol{\theta}, \mathbf{p}), (\boldsymbol{\theta}^*, \mathbf{p}^*)] &= \min \left[\frac{f(\boldsymbol{\theta}^*, \mathbf{p}^*)}{f(\boldsymbol{\theta}, \mathbf{p})}, 1 \right] \\ &= \min [\exp[H(\boldsymbol{\theta}, \mathbf{p}) - H(\boldsymbol{\theta}^*, \mathbf{p}^*)], 1], \end{aligned} \quad (2.12)$$

e rejeitado, caso contrário, fazendo com que a cadeia permaneça no estado atual. Ao final, uma sequência de amostras de $\boldsymbol{\theta}$ é obtida simplesmente descartando os valores das variáveis auxiliares \mathbf{p} . O Algoritmo 3 apresenta um pseudo-código que resume os passos do HMC.

² No HMC padrão, $\mathbf{M} = \mathbf{I}_d$ em que \mathbf{I}_d é a matriz identidade de dimensão d

³ Ou seja, as equações dadas em (2.11) são repetidas L vezes

Algoritmo 3 – Monte Carlo Hamiltoniano

-
- 1: Gere um valor inicial $\boldsymbol{\theta}^{(0)}$ e faça $t=1$.
 - 2: amostre $\mathbf{p}^* \sim N_d(\mathbf{0}, \mathbf{I}_d)$,
 - 3: faça $(\boldsymbol{\theta}^{(t)}, \mathbf{p}^{(t)}) = (\boldsymbol{\theta}^{(t-1)}, \mathbf{p}^*)$ e $H_0 = H(\boldsymbol{\theta}^{(t)}, \mathbf{p}^{(t)})$,
 - 4: **para** $l = 1, \dots, L$ **faça** ▷ solução Leapfrog L vezes
 - 5: $\mathbf{p}^* = \mathbf{p}^* + (\epsilon/2)\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}^{(t-1)})$
 - 6: $\boldsymbol{\theta}^{(t-1)} = \boldsymbol{\theta}^{(t-1)} + \epsilon\mathbf{p}^*$
 - 7: $\mathbf{p}^* = \mathbf{p}^* + (\epsilon/2)\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}^{(t-1)})$
 - 8: **fim para**
 - 9: faça $(\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)}) = (\boldsymbol{\theta}^{(t-1)}, \mathbf{p}^*)$, $H_1 = H(\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)})$ e gere $u \sim U(0, 1)$,
 - 10: calcule $\alpha[(\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)}), (\boldsymbol{\theta}^{(t)}, \mathbf{p}^{(t)})] = \min[\exp(H_0 - H_1), 1]$;
 - 11: faça $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(L)}$ se $\alpha[(\boldsymbol{\theta}^{(L)}, \mathbf{p}^{(L)}), (\boldsymbol{\theta}^{(t)}, \mathbf{p}^{(t)})] > u$, e $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t)}$ caso contrário.
-

Antes de finalizar essa Seção, cabe ressaltar algumas observações sobre a abordagem HMC. No HMC, as tarefas mais custosas, computacionalmente falando, são as avaliações da proporcional da distribuição a posteriori e o seu gradiente com relação a cada parâmetro de interesse. A partir disso, ao utilizar o método *Leapfrog* para discretizar a dinâmica hamiltoniana, as tarefas acima mencionadas são avaliadas um número L de vezes, podendo este variar consideravelmente a depender da complexidade do modelo estatístico (BEAM; GHOSH; DOYLE, 2016). Com isso, apesar de ser um método mais eficiente para inspecionar o espaço paramétrico do(s) parâmetro(s) a ser(em) estimado(s), evitando o comportamento de caminho aleatório mencionado antes, o HMC apresenta um custo maior em gerar amostras da distribuição proposta, principalmente se o número de parâmetros a serem estimados for grande (problema alto-dimensional).

2.2.3.1 Riemann Manifold Monte Carlo Hamiltoniano - RMHMC

Ainda sobre o HMC, os autores Girolami e Calderhead (2011) desenvolveram uma modificação no mecanismo proposto para gerar valores candidatos do estado atual da cadeia, alterando o movimento da cadeia, que a priori é pela métrica de distância euclideana, para a métrica de Riemann. O HMC proposto com essa modificação recebe o nome de *Riemann Manifold HMC* ou RMHMC. Resumidamente, a modificação consiste em mudar a matriz de massa $\mathbf{M} = \mathbf{I}_d$ por $\mathbf{G}(\boldsymbol{\theta})$, em que \mathbf{I}_d é uma matriz identidade de dimensão d , tal que a função Hamiltoniana é redefinida para

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\mathcal{L}(\boldsymbol{\theta}) + \frac{1}{2}\log |\mathbf{G}(\boldsymbol{\theta})| + \mathbf{p}^\top \mathbf{G}(\boldsymbol{\theta}) \mathbf{p} / 2. \quad (2.13)$$

Essa alteração explora melhor as propriedades geométricas da distribuição a posteriori, pois $\mathbf{G}(\boldsymbol{\theta})$ adapta-se de forma conveniente à geometria local da distribuição a posteriori. No entanto, esse tipo de método é bastante custoso computacionalmente, uma vez que $\mathbf{G}(\boldsymbol{\theta})$ varia de acordo com $\boldsymbol{\theta}$ e assim $\boldsymbol{\theta}$ e \mathbf{p} não são mais independentes.

Uma solução simples, que também será utilizada neste trabalho⁴, é considerar a matriz

⁴ O RMHMC será usado como método de amostragem dos parâmetros $\boldsymbol{\beta}$ e δ (veja Seção 2.3) no modelo

$G(\theta)$ da mesma forma que a abordada em [Hartmann e Ehlers \(2017\)](#), porém avaliada na moda a posteriori, a saber

$$G(\theta) = -E \left(\frac{d^2 \mathcal{L}(\theta)}{d\theta^\top \theta} \right) \Big|_{\theta=\hat{\theta}} = -E \left(\frac{d^2 \log p(\mathbf{y}|\theta)}{d\theta^\top \theta} \right) \Big|_{\theta=\hat{\theta}} - \frac{d^2 \log p(\theta)}{d\theta^\top \theta} \Big|_{\theta=\hat{\theta}}, \quad (2.14)$$

ou seja, a matriz de informação do modelo mais a matriz hessiana negativa da log priori avaliadas na moda a posteriori.

Para exemplificar a eficácia do algoritmo acima descrito, [Hartmann e Ehlers \(2017\)](#) avaliam o desempenho dos métodos RMHMC e MH (Seção 2.2.1) por meio de um estudo de simulação e aplicação a dados reais. Os resultados mostram, por meio do cálculo do tamanho efetivo da amostra (*Effective sample size* -ESS) e da capacidade em recuperar os verdadeiros valores dos parâmetros do modelo, que o método RMHMC apresenta desempenho muito melhor que o método MH sob os cenários considerados.

2.3 Distribuições Potência e Reversa de Potência

Nesta Seção, abordaremos as distribuições potência e reversa de potência. A construção desta Seção baseou-se nos trabalhos de [Huayanay et al. \(2019\)](#), [Alves, Bazán e Arellano-Valle \(2023\)](#) e [Huayanay, Bazán e Russo \(2025\)](#), e portanto, o leitor interessado pode consultar esses materiais para obter mais detalhes sobre o assunto.

Basicamente, para produzir uma distribuição de potência, o caminho é simples: considere uma função de distribuição acumulada (fda) H , que será denotada como linha de base (*baseline*), então a nova função de distribuição acumulada potência H é obtida realizando a seguinte exponenciação:

$$H_P(y|\lambda) = H(y)^\lambda, \quad y \in \mathbb{R}, \quad (2.15)$$

em que $\lambda > 0$ é interpretado como o parâmetro de forma da distribuição. A função de distribuição acumulada reversa de potência H associada é obtida da seguinte forma:

$$H_{RP}(y|\lambda) = 1 - H(-y)^\lambda, \quad y \in \mathbb{R}, \quad (2.16)$$

em que novamente $\lambda > 0$. As distribuições H_P e H_{RP} satisfazem a propriedade de reversibilidade, ou seja, $H_P(y) + H_{RP}(-y) = 1$ ([ALVES; BAZÁN; ARELLANO-VALLE, 2023](#)). Observe também que, se $Y \sim H_P$, então $-Y \sim H_{RP}$, ou seja, H_{RP} é chamada de função de distribuição acumulada reversa de H_P .

A partir das expressões 2.15 e 2.16, conseguimos obter a forma das funções densidades de probabilidade para as distribuições de probabilidade potência e reversa de potência. E de fato,

estas são dadas por

$$\begin{aligned} h_P(y|\lambda) &= \lambda H(y)^{\lambda-1} h(y), \quad y \in \mathbb{R} \\ e \\ h_{RP}(y|\lambda) &= \lambda H(-y)^{\lambda-1} h(-y), \quad y \in \mathbb{R}, \end{aligned} \quad (2.17)$$

em que $h(\cdot)$ é a função densidade da distribuição H . Convém observar que quando H pertence a uma família de distribuições simétricas, então para $\lambda = 1$, temos $H_P = H_{RP} = H$, ou seja, a distribuição de probabilidades H torna-se um caso especial de ambas as distribuições em 2.15 e 2.16.

Nos trabalhos de [Lemonte e Bazán \(2018\)](#), [Alves, Bazán e Arellano-Valle \(2023\)](#) e [Huayanay, Bazán e Russo \(2025\)](#), temos uma lista de distribuições que foram utilizadas como linhas de base para produzir novas funções de ligação, a saber: Cauchy, Logística, Gumbel Reversa, Normal, Laplace e t-Student. A distribuição de probabilidades t-Student apresenta um parâmetro extra que são os graus de liberdade ν . Como argumentam [Huayanay, Bazán e Russo \(2025\)](#), o parâmetro extra da referida distribuição de probabilidades não estaria ligado à função de ligação e, por fim, tornaria-se uma tarefa difícil buscar uma interpretação para ele no contexto de regressão binária. Por isso não utilizaremos a distribuição de probabilidades t-Student neste trabalho para produzir uma função de ligação.

As distribuições potência e reversa de potência estão sendo apresentadas neste trabalho em sua forma padronizada, isto é, introduzindo um parâmetro de localização $\mu \in \mathbb{R}$ e um parâmetro de escala $\sigma > 0$ nas expressões 2.15 e 2.16, então, estamos considerando $\mu = 0$ e $\sigma = 1$. A apresentação das distribuições em sua forma padrão é conveniente, pois a utilização destas como função de ligação se dará em sua forma padrão.

Tabela 1 – Distribuições potência e reversa de potência.

Tipo	Ligação	Notação
Potência	Normal Potência	N-P
	Logística Potência	L-P
	Laplace Potência	LA-P
	Cauchy Potência	C-P
	Gumbel Reversa Potência	GR-P
Reversa de Potência	Normal Reversa de Potência	N-RP
	Logística Reversa de Potência	L-RP
	Laplace Reversa de Potência	LA-RP
	Cauchy Reversa de Potência	C-RP
	Gumbel Reversa Reversa de Potência	GR-RP

Fonte: Elaborada pelo autor.

Na Tabela 1, apresentamos a lista de distribuições potência e reversa de potência que serão utilizadas neste trabalho, produzidas a partir das distribuições Cauchy, Logística, Laplace,

Normal e Gumbel Reversa, tomadas como linha de base. Vale mencionar que não utilizaremos a distribuição de probabilidades Gumbel como *baseline*, mas, apenas a Gumbel Reversa. O motivo é que, como argumentam [Huayanay, Bazán e Russo \(2025\)](#) e [Alves, Bazán e Arellano-Valle \(2023\)](#), quando escolhemos a distribuição de probabilidades Gumbel como linha de base, a assimetria permanece constante independentemente do valor de λ . O trabalho de [Huayanay, Bazán e Russo \(2025\)](#) oferece estudos empíricos que demonstram o porquê desse fato.

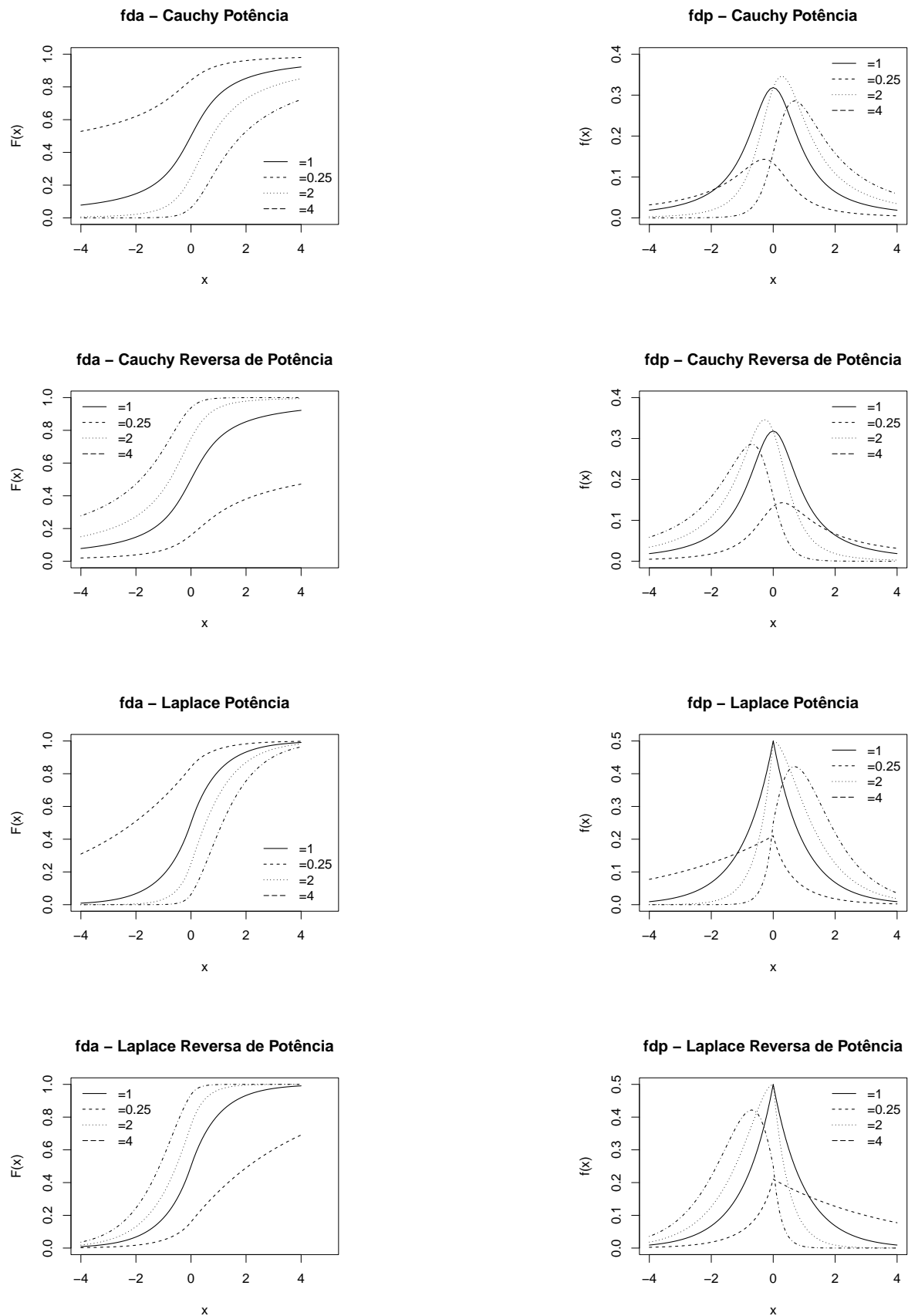
Na Tabela 2, apresentamos a forma das funções densidade de probabilidade (fdp) e as suas respectivas funções de distribuição acumulada (fda) para as distribuições apresentadas na Tabela 1 em sua forma padrão, ou seja, $\mu = 0$ e $\sigma = 1$. A inversa das fda's constitui a função de ligação para a construção dos modelos de regressão binária.

Tabela 2 – fda e fdp das distribuições usadas neste trabalho.

Modelo	fda $F_\lambda(y)$	fdp $f_\lambda(y)$
N-P	$[\Phi(y)]^\lambda$	$\lambda [\Phi(y)]^{\lambda-1} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\}$
L-P	$\left(\frac{1}{1 + \exp(-y)}\right)^\lambda$	$\lambda \left(\frac{1}{1 + \exp(-y)}\right)^{\lambda-1} \frac{\exp(-y)}{(1 + \exp(-y))^2}$
LA-P	$\left[\frac{1}{2} + \frac{1}{2} \text{sign}(y) \{1 - e^{- y }\}\right]^\lambda$	$\lambda \left[\frac{1}{2} + \frac{1}{2} \text{sign}(y) \{1 - e^{- y }\}\right]^{\lambda-1} \frac{1}{2} \exp(- y)$
C-P	$\left[\frac{1}{\pi} \arctan(y) + \frac{1}{2}\right]^\lambda$	$\frac{\lambda}{\pi} \left[\frac{1}{\pi} \arctan(y) + \frac{1}{2}\right]^{\lambda-1} [1 + y^2]^{-1}$
GR-P	$[1 - \exp\{-\exp\{y\}\}]^\lambda$	$\lambda \exp\{y - \exp\{y\}\} [1 - \exp\{-\exp\{y\}\}]^{\lambda-1}$
N-RP	$1 - [\Phi(-y)]^\lambda$	$\lambda [\Phi(-y)]^{\lambda-1} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\}$
L-RP	$1 - \left(\frac{1}{1 + \exp(y)}\right)^\lambda$	$\lambda \left(\frac{1}{1 + \exp(y)}\right)^{\lambda-1} \frac{\exp(-y)}{(1 + \exp(-y))^2}$
LA-RP	$1 - \left[\frac{1}{2} - \frac{1}{2} \text{sign}(y) \{1 - e^{- y }\}\right]^\lambda$	$\lambda \left[\frac{1}{2} - \frac{1}{2} \text{sign}(y) \{1 - e^{- y }\}\right]^{\lambda-1} \frac{1}{2} \exp(- y)$
C-RP	$1 - \left[\frac{1}{\pi} \arctan(-y) + \frac{1}{2}\right]^\lambda$	$\frac{\lambda}{\pi} \left[\frac{1}{\pi} \arctan(-y) + \frac{1}{2}\right]^{\lambda-1} [1 + y^2]^{-1}$
GR-RP	$1 - [1 - \exp\{-\exp\{-y\}\}]^\lambda$	$\lambda \exp\{-y - \exp\{-y\}\} \times [1 - \exp\{-\exp\{-y\}\}]^{\lambda-1}$

Fonte: Elaborada pelo autor.

Na Figura 2, temos as curvas das funções densidade de probabilidade ($f(x)$) e funções de

Figura 2 – Funções de distribuição acumulada e funções densidade de probabilidade para os modelos: C-P, C-RP, LA-P e LA-RP usando diferentes valores de λ .

Fonte: Elaborada pelo autor.

distribuição acumulada ($F(x)$) das distribuições C-P, C-RP, LA-P e LA-RP; as curvas das demais distribuições listadas na Tabela 1 estão no Apêndice C. Por meio dessas figuras, conseguimos destacar características interessantes. No tocante às distribuições potência e reversa de potência, percebe-se que as curvas das fdp's das distribuições potência são inclinadas para a direita quando $0 < \lambda < 1$, e inclinadas para a esquerda quando $\lambda > 1$. Quanto às distribuições reversa de potência, temos o oposto, as curvas das fdp's são inclinadas para a esquerda quando $0 < \lambda < 1$, e inclinadas para a direita quando $\lambda > 1$. E por fim, em ambos os casos, tanto para as distribuições potência quanto para as reversa de potência, obtemos a distribuição de probabilidades *baseline* quando $\lambda = 1$.

Com relação às funções de distribuição acumulada, constatamos que cada curva gerada por uma distribuição de probabilidades potência é um reflexo da curva gerada pela sua respectiva distribuição de probabilidades reversa de potência. Por exemplo, observe a distribuição de probabilidades Cauchy Potência. A curva gerada para $\lambda = 0,5$ está acima da curva da distribuição de probabilidades linha de base ($\lambda = 1$); quando observamos a distribuição de probabilidades Cauchy Reversa de Potência, a curva gerada para $\lambda = 0,5$ está abaixo da curva da distribuição de probabilidades linha de base, como se tivesse sido refletida num espelho. Para os interessados em estudar sobre as propriedades das distribuições potência e reversa de potência referente ao assunto de assimetria e curtose, indicamos a consulta de [Huayanay, Bazán e Russo \(2025\)](#).

2.4 Modelo de Regressão Binária Bayesiano

Em um problema de classificação binária, a variável de interesse possui duas categorias: uma que comumente chamamos de “sucesso” e a outra, de “fracasso”. Em análise prática, atribuímos valor 1 ao evento de interesse, que nesse caso seria a classe referente ao sucesso, e valor 0 ao evento complementar, ou seja, a classe que denotamos como fracasso. Um modelo de regressão binária consiste, portanto, em analisar o comportamento de uma variável binária em função de um conjunto de outras variáveis, as quais denominamos variáveis regressoras ou covariáveis.

Formalmente, considere $\mathbf{y} = (y_1, \dots, y_n)^\top$; um vetor $n \times 1$ de respostas binárias observadas, sendo $y_i = 1$ com probabilidade $p_i = P(Y_i = 1)$ e $y_i = 0$ com probabilidade $1 - p_i$, e $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^\top$ um vetor $k \times 1$ de covariáveis, para $i = 1, \dots, n$. \mathbf{X} será considerada a matriz de delineamento (*design matrix*) $n \times k$ cujas linhas são \mathbf{x}_i ; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$ um vetor $k \times 1$ de coeficientes de regressão. Assumindo que os dados observados são condicionalmente independentes, dado $\boldsymbol{\beta}$, temos que o modelo de regressão binária bayesiano pode ser assumido da seguinte forma:

$$\begin{aligned} Y_i | \boldsymbol{\beta} &\stackrel{\text{ind}}{\sim} \text{Bernoulli}(p_i), \\ p_i &= F(\eta_i) = F(\mathbf{x}_i^\top \boldsymbol{\beta}), \quad i = 1, \dots, n, \\ \boldsymbol{\beta} &\sim p(\boldsymbol{\beta}) \end{aligned} \tag{2.18}$$

em que $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ é o preditor linear, $p(\boldsymbol{\beta})$ é a distribuição de probabilidades a priori do vetor de coeficientes de regressão $\boldsymbol{\beta}$; $F(\cdot)$ denota uma função de distribuição acumulada (f.d.a), e $F^{-1}(\cdot)$ é tipicamente chamada de função de ligação (BAZÁN; BOLFARINE; BRANCO, 2010). A função de ligação $F^{-1}(\cdot)$ faz a transformação de $0 < p_i < 1$ para $i = 1, \dots, n$, tal que o preditor linear $F^{-1}(p_i) = \eta_i$ assumam valores em \mathbb{R} e, conseqüentemente, $F(\eta_i) = p_i$ faz a transformação do preditor linear, para qualquer valor em \mathbb{R} , considerando um valor no intervalo $(0, 1)$.

Quanto à função de ligação, esta desempenha um papel importantíssimo na construção do modelo de regressão binária. Tanto é, que esta pode impactar significativamente a qualidade do ajuste realizado se for mal escolhida, e para isso, existem até testes de adequacidade da função de ligação (COLLETT, 2002).

Para o modelo em 2.18, como já mencionado, quando $F(\cdot)$ é uma fda simétrica, a curva de resposta de probabilidade fornecida possui forma simétrica em torno de 0,5. Atualmente, as funções Logito, Cauchito e Probit são três exemplos de funções de ligação que podem ser consideradas simétricas, e dão origem aos modelos de regressão binária mais conhecidos na literatura. Porém, quando a fda provém de uma distribuição de probabilidades assimétrica, as funções de ligação resultantes também são assimétricas. Na literatura, temos o caso das funções de ligação Cloglog e Loglog.

Na Figura 3, apresentamos as curvas das fda's das distribuições Cauchy, Logística, Normal padrão, Gumbel Reversa e Gumbel, que dão origem, respectivamente, às funções de ligação Cauchito, Logito, Probit, Cloglog e Loglog. Podemos observar o comportamento de simetria das fda's Cauchy, Logística e Normal padrão quando $F(0) = 0,5$. Estas possuem ponto de inflexão exatamente neste ponto. Ao passo que as distribuições Gumbel Reversa e Gumbel apresentam ponto de inflexão distinto de $F(0) = 0,5$.

Dando continuidade, a forma da função de verossimilhança do modelo em 2.18 é expressa por

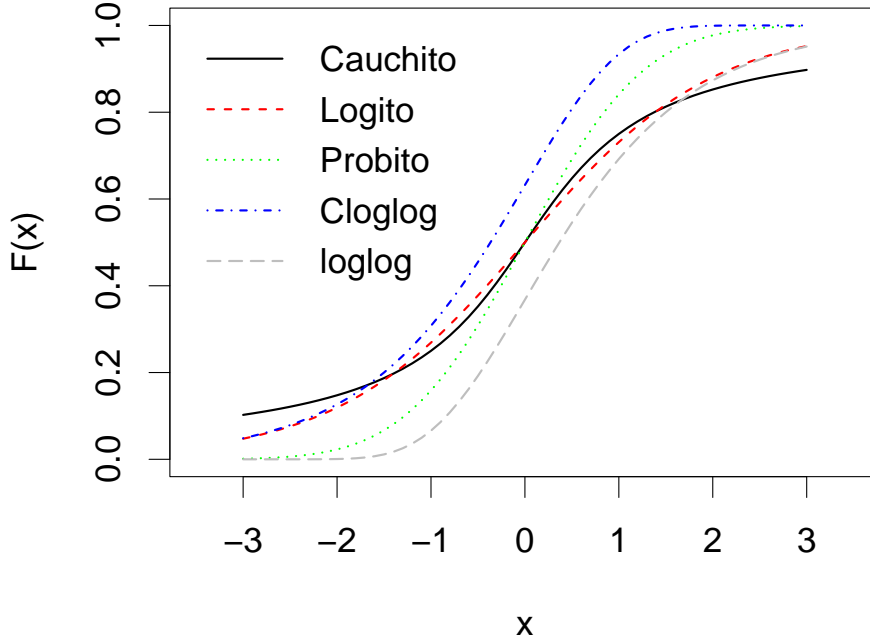
$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}) &= \prod_{i=1}^n p(y_i | \mathbf{x}_i, \boldsymbol{\beta}) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \\ &= \prod_{i=1}^n [F(\mathbf{x}_i^\top \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}_i^\top \boldsymbol{\beta})]^{1-y_i}. \end{aligned} \quad (2.19)$$

A respectiva função de log-verossimilhança é dada por

$$\begin{aligned} \log \mathcal{L}(\boldsymbol{\beta}) &= \log \prod_{i=1}^n [F(\mathbf{x}_i^\top \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}_i^\top \boldsymbol{\beta})]^{1-y_i} \\ &= \sum_{i=1}^n \{y_i \log [F(\mathbf{x}_i^\top \boldsymbol{\beta})] + (1 - y_i) \log [1 - F(\mathbf{x}_i^\top \boldsymbol{\beta})]\}. \end{aligned} \quad (2.20)$$

Atribuindo distribuição a priori $p(\boldsymbol{\beta}) \equiv N_k(\mathbf{0}, \boldsymbol{\Sigma})$, temos que a proporcional da distribuição log-posteriori resulta em

Figura 3 – Função de distribuição acumulada para as distribuições Logística, Normal padrão, Cauchy, Gumbel Reversa e Gumbel. Simetria em torno de $F(0) = 0,5$ para $x \in (-4,4)$.



Fonte: Elaborada pelo autor

$$\begin{aligned}
 \log p(\boldsymbol{\beta}|\mathbf{y}) &= \log \mathcal{L}(\boldsymbol{\beta}) + \log p(\boldsymbol{\beta}) + \text{constante} \\
 &= \sum_{i=1}^n \{y_i \log [F(\mathbf{x}_i^\top \boldsymbol{\beta})] + (1 - y_i) \log [1 - F(\mathbf{x}_i^\top \boldsymbol{\beta})]\} - \frac{1}{2} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta} + \text{constante}
 \end{aligned} \tag{2.21}$$

No caso em que assumimos $F(\cdot)$ advindo da distribuição de probabilidade Logística, origina-se a função de ligação chamada Logito. Porém, observe primeiro que:

$$p_i = \frac{1}{(1 + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta}))} = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \text{ e } 1 - p_i = \frac{1}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})}.$$

Então, chegamos ao seguinte resultado

$$\begin{aligned}
 \log p(\boldsymbol{\beta}|\mathbf{y}) &\propto \sum_{i=1}^n y_i \log \left[\frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \right] + (1 - y_i) \log \left[\frac{1}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \right] - \frac{1}{2} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta} \\
 &\propto \sum_{i=1}^n [y_i \mathbf{x}_i^\top \boldsymbol{\beta} - \log (1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}))] - \frac{1}{2} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}
 \end{aligned} \tag{2.22}$$

Considerando que $F(\cdot) \equiv \Phi(\cdot)$, em que $\Phi(\cdot)$ é a fda da distribuição de probabilidade Normal padrão, origina-se a função de ligação comumente chamada de Probit. Para esse caso, a proporcional da log-posteriori de β é dada por

$$\log p(\beta|\mathbf{y}) = \sum_{i=1}^n \{y_i \log [\Phi(\mathbf{x}_i^\top \beta)] + (1 - y_i) \log [1 - \Phi(\mathbf{x}_i^\top \beta)]\} - \frac{1}{2} \beta^\top \Sigma^{-1} \beta + \text{constante} \quad (2.23)$$

É perceptível que as proporcionais das distribuições em 2.22 e 2.23 não possuem a forma de uma distribuição de probabilidade conhecida, ou seja, elas não podem ser tratadas analiticamente. Para esse fim, faz jus utilizar os métodos de Monte Carlo via Cadeias de Markov para amostrar valores dos parâmetros de interesse.

2.4.1 Desbalanceamento de Classes em Classificação Binária

Em classificação binária, o problema de desbalanceamento de classes surge quando o número de observações em uma das classes é significativamente maior que o da outra. Quando isso ocorre, usar função de ligação simétrica nem sempre proporciona o melhor ajuste dos dados, afetando a parte inferencial do modelo. Em Wang e Dey (2011), por exemplo, os autores apresentam um problema cujo interesse é avaliar a probabilidade de uma empresa adotar (ou não) um sistema de pagamento eletrônico. Nesta situação, há uma grande divergência nas quantidades de observações nas categorias de resposta, pois a adoção de um sistema de pagamento eletrônico por parte de uma empresa é considerada um “evento raro”, e deve ocorrer com baixa probabilidade. Sobre isso, King e Zeng (2001) e Huayanay *et al.* (2019) argumentam que eventos raros são difíceis de prever. Por isso, utilizar modelos com função de ligação simétrica devem subestimar a probabilidade de eventos raros.

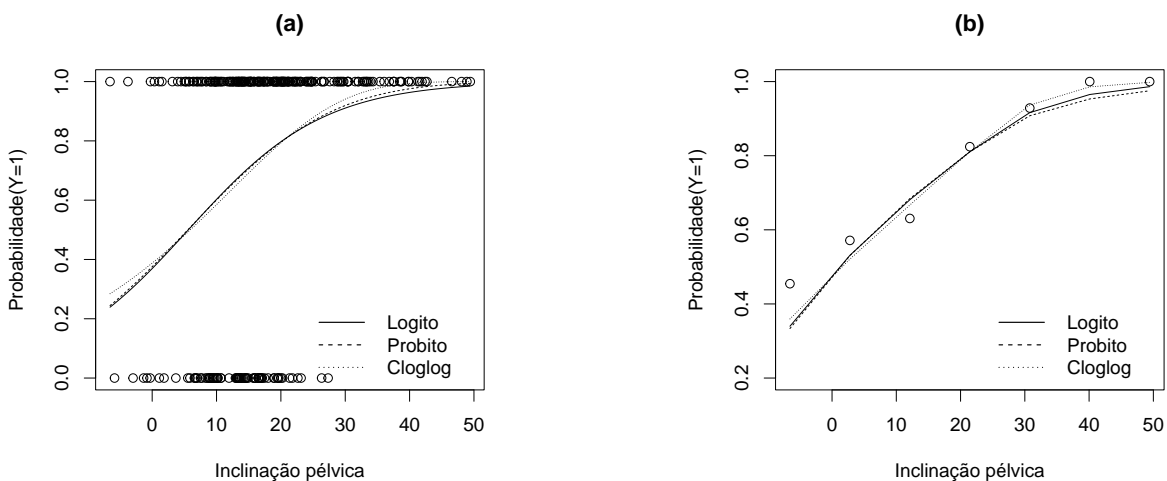
Para exemplificar, tomamos como base o conjunto de dados biomédicos (MA; ZHU; HSU, 2019), retirados do repositório UCI - *Machine Learning Repository* (DUA; GRAFF, 2017), cuja variável de interesse é a classificação dos pacientes ortopédicos em duas classes: “Normal=1” e “Anormal=0”. Para construir o exemplo, pegamos um subconjunto de dados de 290 observações, das quais 210 são “Normal=1”. Em seguida, ajustamos modelos de regressão binária com funções de ligação Logito, Probit e Cloglog a esta amostra. Os ajustes foram feitos utilizando apenas uma covariável, que neste caso é a inclinação pélvica, pois facilitará a visualização das curvas dos modelos ajustados.

Na Figura 4 (a) são apresentados os ajustes das curvas dos modelos Logito, Probit e Cloglog aos dados originais. Para facilitar a visualização e a comparação entre os modelos, a Figura 4 (b) mostra os dados agrupados em 7 categorias da covariável inclinação pélvica, conforme detalhado na Tabela 3. Em cada grupo, contabilizou-se o número total de observações n_i ($i = 1, \dots, 7$) e o número de observações x_i nas quais o evento de interesse ocorreu,

transformando assim os dados em variáveis binomiais.⁵

Na Tabela 3 temos o número estimado de “Normal=1” pelos modelos ajustados sob diferentes funções de ligação. Percebe-se, pelos resultados dos números estimados, que o modelo com ligação assimétrica (assimetria fixa) Cloglog se ajusta melhor aos dados quando comparado aos dois modelos com ligações simétricas. A assimetria (desbalanceamento) presente nesses dados, embora não tão intensa, foi suficiente para que um modelo com um pouco mais de assimetria, como o Cloglog, levasse vantagem. Porém, haverá muitas situações em que a presença de assimetria nos dados será muito maior, de tal forma que ligações com assimetria fixa não conseguirão suportá-la, levando-nos naturalmente a pensar que modelos dotados de assimetria flexível, que se adaptam de acordo com os dados, seriam uma opção melhor.

Figura 4 – Ajuste de modelos com funções de ligação Logito (linha sólida), Probit (linha tracejada), e Cloglog (linha pontilhada) ao subconjunto de dados ortopédicos com inclinação pélvica como uma covariável. Figura (a): ajuste dos modelos aos dados originais. Figura (b): dados agrupados em categorias.



Fonte: Elaborada pelo autor.

Tabela 3 – Número estimado de “Normal=1” sob diferentes funções de ligação.

Inclinação pélvica	[-6.55,1.45)	[1.45,9.44)	[9.44,17.4)	[17.4,25.4)	[25.4,33.4)	[33.4,41.4)	[41.4,49.4)
# Normal	5	24	70	61	26	16	7
Logito	3.67	22.29	75.97	59.98	25.41	15.24	6.82
Probit	3.73	22.30	75.65	60.03	25.67	15.44	6.90
Cloglog	3.96	21.90	74.05	59.99	26.21	15.77	6.98

Fonte: Elaborada pelo autor.

Um dos objetivos deste trabalho é justamente esse: avaliar o desempenho de modelos de regressão binária sob funções de ligação com assimetria flexível, usando as funções de ligação

⁵ Em outras palavras, seria: $X_i \sim \text{Binomial}(n_i, p_i)$ para $i = 1, \dots, 7$.

potência e reversa de potência, que possuem um parâmetro de forma λ que nos permite esta flexibilidade.

2.5 Comparação de Modelos

Após o ajuste de um determinado modelo, é intuitivo pensar na mensuração da qualidade do ajuste feito com o objetivo de validar o modelo proposto. Uma atitude também plausível seria compará-lo a certos modelos concorrentes como forma de buscar o modelo mais adequado, segundo alguns critérios, para o problema em análise. Nesta Seção, serão apresentadas algumas medidas para avaliar a qualidade do ajuste feito, bem como outros critérios que servirão de base para fazer seleção de modelos.

2.5.1 Critério de informação amplamente aplicável WAIC (*widely applicable information criterion*)

Neste trabalho, avaliaremos a performance preditiva de modelos concorrentes com base em seus desempenhos para predição de observações fora da amostra (*predicting out-of-sample observations*). Definamos primeiramente uma medida de acurácia preditiva apresentada em [Vehtari, Gelman e Gabry \(2017\)](#), chamada de log densidade preditiva pontual esperada (*expected log pointwise predictive density*). O nosso objetivo será, portanto, aproximar essa medida. Para um novo conjunto de n pontos de dados, tomados um de cada vez, a medida é definida da seguinte forma

$$\begin{aligned} \text{elpd} &= \text{expected log pointwise predictive density} \\ &= \sum_{i=1}^n \int p_t(\tilde{y}_i) \log p(\tilde{y}_i | \mathbf{y}) d\tilde{y}_i, \end{aligned} \quad (2.24)$$

em que $p_t(\tilde{y}_i)$ é a distribuição de probabilidade (desconhecida) que representa o verdadeiro processo gerador de dados para \tilde{y}_i , que é uma observação futura. A medida de acurácia preditiva definida em 2.24, segundo [Vehtari, Gelman e Gabry \(2017\)](#), foi assim definida para avaliar a densidade preditiva, e esta é intratável analiticamente, necessitando portanto de aproximação.

Uma outra quantidade que será bastante útil, principalmente quando formos definir o critério de informação amplamente aplicável WAIC (*widely applicable information criterion*), é a log densidade preditiva pontual definida como

$$\begin{aligned} \text{lpd} &= \text{log pointwise predictive density} \\ &= \sum_{i=1}^n \log p(y_i | \mathbf{y}) = \sum_{i=1}^n \log \int p(y_i | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}. \end{aligned} \quad (2.25)$$

Como argumenta [Vehtari, Gelman e Gabry \(2017\)](#), a quantidade dada em 2.25 para os pontos de dados observados \mathbf{y} é uma sobreestimativa da elpd para os dados futuros em 2.24. De

posse de uma amostra $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$ da distribuição a posteriori $p(\boldsymbol{\theta}|\mathbf{y})$, podemos calcular a lpd dada em 2.25 da seguinte forma:

$$\hat{\text{lpd}} = \sum_{i=1}^n \log \left(\frac{1}{N} \sum_{l=1}^N p(y_i|\boldsymbol{\theta}^{(l)}) \right). \quad (2.26)$$

Como a quantidade acima é uma sobreestimativa da elpd dada em (2.24), uma forma de correção será necessária para que possamos fornecer uma estimativa mais razoável para a elpd. Uma proposta de correção, baseada nos dados, para o número efetivo estimado de parâmetros foi apresentada em Watanabe (2010). Essa proposta utiliza as variâncias das log densidades preditivas e é definida como

$$p = \sum_{i=1}^n \text{Var}(\log p(y_i|\boldsymbol{\theta})), \quad (2.27)$$

em que as variâncias em 2.27 são tomadas em relação à distribuição a posteriori de $\boldsymbol{\theta}$. Especificamente, de posse de uma amostra $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$ da distribuição a posteriori de $\boldsymbol{\theta}$, na prática calculamos as log densidades preditivas $\log p(y_i|\boldsymbol{\theta}^{(1)}), \dots, \log p(y_i|\boldsymbol{\theta}^{(N)})$ para $i = 1, \dots, n$, e posteriormente obtemos as variâncias amostrais dessas quantidades que nos fornecem a seguinte estimativa:

$$\hat{p} = \sum_{i=1}^n V_{l=1}^N(\log p(y_i|\boldsymbol{\theta}^{(l)})), \quad (2.28)$$

em que $V_{l=1}^N$ representa a variância amostral, $V_{l=1}^N a_l = \frac{1}{N-1} \sum_{l=1}^N (a_l - \bar{a})^2$.

Dessa forma, podemos definir o então chamado critério de informação amplamente aplicável (ou critério de informação de Watanabe) WAIC, que é uma estimativa para a elpd, por

$$\hat{\text{elpd}} = \hat{\text{lpd}} - \hat{p}, \quad (2.29)$$

ou equivalentemente defini-lo por $-2 \hat{\text{elpd}} = -2 \hat{\text{lpd}} + 2 \hat{p}$ para fornecê-lo na escala deviance. É necessário destacar que a quantidade WAIC, assim como outros critérios de informação, procura se apresentar como uma medida que concilie a adequacidade do modelo em detrimento de sua complexidade. Como argumenta Watanabe (2010), a medida WAIC se torna mais estável que o critério de informação deviance, que também possui uma versão similar de cálculo para o número efetivo de parâmetros baseado em variância como em 2.29 (SPIEGELHALTER *et al.*, 2002) por poder calcular a variância separadamente para cada ponto de dado. Finalmente, o WAIC é assintoticamente equivalente à abordagem de validação cruzada *leave-one-out*.

2.5.2 Avaliação de Modelos via Medida L

Apresentamos a seguir o critério de avaliação e seleção de modelos baseado em funções de perda (ou medidas de perda *-loss measure*). A abordagem bayesiana avalia a performance de modelos ao comparar aquilo que é predito com aquilo que tem sido observado, e o critério de

seleção de modelos baseado em medidas de perda, que será apresentado adiante, utilizará esse fundamento.

A estrutura geral de uma medida de perda L é dada por, para o caso mais simples, $L(\tilde{y}, a; \mathbf{y})$, em que a seria a “ação” tomada para prever \tilde{y} , e este último seria a observação a ser predita, e \mathbf{y} seriam os dados observados. Dada a medida de perda, o procedimento plausível e intuitivo, como mostram [Gelfand e Ghosh \(1998\)](#), seria calcular

$$\min_a E_{\tilde{y}|\mathbf{y}} L(\tilde{y}, a; \mathbf{y}), \quad (2.30)$$

ou seja, a obtenção da menor perda esperada. A esperança em 2.30 é tomada com relação à distribuição preditiva associada a \tilde{y} . No contexto de seleção de modelos, ajustamos $m = 1, 2, \dots, M$ modelos aos dados observados, e para cada um deles teremos uma estimativa da medida de perda esperada, que é minimizada sobre a . O modelo selecionado, em tese, será aquele que produzir o menor mínimo em 2.30. A expressão em 2.30 pode ser obtida por integração de Monte Carlo.

Sob o contexto de modelos de regressão, de posse de uma matriz de delineamento \mathbf{X} , e dos dados observados \mathbf{y} , a medida $L(\tilde{\mathbf{y}}, \mathbf{a}; \mathbf{y})$ de perda quadrática, definida como em [Gelfand e Ghosh \(1998\)](#), é dada por

$$L(\tilde{\mathbf{y}}, \mathbf{a}; \mathbf{y}, \mathbf{X}) = E[(\tilde{\mathbf{y}} - \mathbf{a})'(\tilde{\mathbf{y}} - \mathbf{a})|\mathbf{y}, \mathbf{X}] + k(\mathbf{y} - \mathbf{a})'(\mathbf{y} - \mathbf{a}), \quad k \geq 0, \quad (2.31)$$

em que a primeira parte da soma no lado direito de 2.31 é vista como um termo que captura a precisão da estimação, e o segundo termo, como uma forma de avaliar a bondade do ajuste. Para os n componentes de $\tilde{\mathbf{y}}$, podemos desenvolver a expressão em 2.31 e chegar ao seguinte resultado:

$$\begin{aligned} L(\tilde{\mathbf{y}}, \mathbf{a}; \mathbf{y}, \mathbf{X}) &= \sum_{i=1}^n E[\tilde{y}_i^2 - 2\tilde{y}_i a_i + a_i^2] + k \sum_{i=1}^n (y_i - a_i)^2 \\ &= \sum_{i=1}^n E(\tilde{y}_i^2) - 2E(\tilde{y}_i) a_i + a_i^2 + k \sum_{i=1}^n (y_i - a_i)^2 \\ &= \sum_{i=1}^n \text{Var}(\tilde{y}_i|\mathbf{y}, \mathbf{X}) + \mu_i^2 - 2\mu_i a_i + a_i^2 + k \sum_{i=1}^n (y_i - a_i)^2 \\ &= \sum_{i=1}^n \text{Var}(\tilde{y}_i|\mathbf{y}, \mathbf{X}) + \sum_{i=1}^n (\mu_i^2 - a_i)^2 + k \sum_{i=1}^n (y_i - a_i)^2, \end{aligned} \quad (2.32)$$

em que $\mu_i = E(\tilde{y}_i|\mathbf{y}, \mathbf{X})$. Conforme [Gelfand e Ghosh \(1998\)](#), selecionamos a_i como sendo aquele que minimiza a expressão em 2.32, o que nos leva a $a_i = (\mu_i + ky_i)(k + 1)^{-1}$. Fazendo $\nu = k/(k + 1)$, e substituindo o valor de a_i em 2.32, obtemos

$$L(\tilde{\mathbf{y}}, \mathbf{a}; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^n \text{Var}(\tilde{y}_i|\mathbf{y}, \mathbf{x}_i) + \nu \sum_{i=1}^n (\mu_i - y_i)^2. \quad (2.33)$$

Para $\nu = 1$, 2.33 torna-se

$$L(\tilde{\mathbf{y}}, \mathbf{a}; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^n \text{Var}(\tilde{y}_i|\mathbf{y}, \mathbf{x}_i) + \sum_{i=1}^n (\mu_i - y_i)^2. \quad (2.34)$$

Modelos que produzirem os menores valores de 2.34 são indicados como sendo os melhores modelos.

2.6 Diagnóstico de Modelo e Detecção de Outliers

Nesta Seção, abordaremos a estratégia apresentada em Goh e Dey (2014) para avaliar a qualidade do ajuste de modelos bayesianos baseada na divergência de Bregman funcional para comparar distribuições a posteriori. A proposta de Goh e Dey (2014) é útil tanto para mensurar a qualidade do ajuste do modelo avaliado, como para construir um método para detecção de *outliers*. No entanto, nesta tese a utilizaremos com o objetivo de construir uma métrica de detecção de *outliers*. Especificamente, a divergência de Bregman funcional é utilizada para medir a dissimilaridade entre a distribuição a posteriori $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})$ e a distribuição a posteriori após sofrer algum tipo de perturbação $p_v(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})$.

Formalmente, seja ψ uma função estritamente convexa e diferenciável, e ψ' a derivada de ψ . A divergência de Bregman funcional para comparar distribuições de probabilidade a posteriori é definida da seguinte forma

$$D_\psi = \int \{ \psi(p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})) - \psi(p_v(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})) - [p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) - p_v(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})] \psi'(p_v(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})) \} \times p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) d\boldsymbol{\theta}. \quad (2.35)$$

Neste trabalho, à semelhança do que foi abordado no artigo seminal de Goh e Dey (2014), a classe de funções convexas ψ que será utilizada na análise de diagnósticos para os modelos propostos nesta tese é a seguinte:

$$\psi_\alpha(x) = \begin{cases} x \log(x) - x + 1, & \alpha = 1 \\ \log(x) + x - 1, & \alpha = 0 \\ \frac{x^\alpha - \alpha x + \alpha - 1}{\alpha(\alpha - 1)}, & \text{caso contrário.} \end{cases}$$

Para esta classe de funções convexas, a divergência de Bregman funcional em 2.35 possui formas de medidas de divergências conhecidas, dependendo da escolha de α .

Direcionando-se para o contexto de detecção de *outliers*, uma medida também pode ser construída. Para isso, considere as seguintes definições de perturbação, levando em conta a independência entre as observações y_1, \dots, y_n :

$$v_1(\boldsymbol{\theta}, \mathbf{y}, \mathbf{X}) = \frac{p(\mathbf{y}_{(i)}|\mathbf{X}_{(i)}, \boldsymbol{\theta})}{p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})} = \frac{\prod_{j \neq i}^n p(\mathbf{y}_j|\mathbf{x}_j, \boldsymbol{\theta})}{\prod_{j=1}^n p(\mathbf{y}_j|\mathbf{x}_j, \boldsymbol{\theta})} = \frac{1}{p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta})}, \quad (2.36)$$

e

$$v_2(\boldsymbol{\theta}, \mathbf{y}, \mathbf{X}) = \frac{p(\mathbf{y}|\mathbf{X}_{[i(j)]}, \boldsymbol{\theta})}{p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})} = \frac{p(\mathbf{y}_i|\mathbf{x}_{i(j)}, \boldsymbol{\theta})}{p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta})}, \quad (2.37)$$

em que y_i e \mathbf{x}_i são, respectivamente, a i -ésima observação e o i -ésimo vetor de covariáveis; $\mathbf{y}_{(i)}$ designa um vetor de dados aleatórios sem a i -ésima observação; $\mathbf{X}_{(i)}$ denota a matriz de covariáveis com a i -ésima linha removida; $\mathbf{X}_{[i(j)]}$ refere-se a matriz de covariáveis obtida ao deletar a j -ésima componente do i -ésimo vetor de covariáveis de \mathbf{X} , enquanto $\mathbf{x}_{i(j)}$ é o i -ésimo vetor de covariáveis com a j -ésima componente deletada. As perturbações definidas acima, de acordo com Goh e Dey (2014), podem ser usadas para medir o efeito na i -ésima observação e no i -ésimo vetor de covariáveis.

Então, de posse de uma amostra $\{\boldsymbol{\theta}\}_{l=1}^N$ da distribuição a posteriori de $\boldsymbol{\theta}$ para mensurar o efeito da i -ésima observação, o primeiro passo é calcular a perturbação média para i -ésima observação da seguinte forma:

$$\bar{v}_i = \frac{1}{N} \sum_{l=1}^N \frac{1}{p(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}^{(l)})}, \quad i = 1, \dots, n, \quad (2.38)$$

e logo em seguida, fornecer uma estimativa para a divergência de Bregman para a i -ésima observação ao fazer

$$\hat{D}_{i,\psi} = \frac{1}{N} \sum_{l=1}^N \left[\psi \left(\frac{1}{\bar{v}_i p(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}^{(l)})} \right) \right], \quad i = 1, \dots, n. \quad (2.39)$$

A medida dada em 2.39 é positiva, e seus valores muito altos indicam fortes indícios de possíveis *outliers*.

No trabalho de Goh e Dey (2014), os autores propuseram avaliar a dissimilaridade $p_v(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})$ e $p(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})$ de forma direta, usando a divergência de Bregman funcional. Essa abordagem representa uma forma alternativa mais estável (computacionalmente falando), e também serve para detecção de *outliers*. Assim, tal abordagem é dada por

$$D_\psi = \int \{ \psi(p(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})) - \psi(p_v(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})) - [p(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X}) - p_v(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})] \psi'(p_v(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})) \} d\boldsymbol{\theta}, \quad (2.40)$$

que é intratável analiticamente e aproximações serão necessárias. Como alternativa, os autores sugerem fazer aproximações da quantidade em 2.40 através da abordagem por estimação de densidade marginal de Importância ponderada (*Weighted Marginal Density Estimation-IWMDE*). Essa abordagem também é a mesma que será adotada neste trabalho para obter aproximações de 2.40. Utilizando uma distribuição de probabilidade normal multivariada no IWMDE e para $\alpha = 1$, correspondendo à divergência de Kullback-Leibler, pode-se mostrar que a divergência de Bregman funcional para a i -ésima observação é aproximada por

$$\hat{D}_{i,\psi} = \frac{1}{N} \sum_{l=1}^N \left[-\log \left(\frac{1}{\bar{v}_i p(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}^{(l)})} \right) \right], \quad i = 1, \dots, n. \quad (2.41)$$

Santos e Bolfarine (2016), também apresentaram uma abordagem de detecção de *outliers* no contexto de regressão quantílica. Por um método similar ao deles, pode-se propor avaliar as quantidades apresentadas em 2.41 em relação às outras para cada observação. Basicamente, isso é feito computando

quantas vezes cada valor é maior que os outros ao longo das simulações MCMC. Dito isso, dada uma amostra $\{\boldsymbol{\theta}\}_{l=1}^N$ da distribuição a posteriori de $\boldsymbol{\theta}$, obtemos as estimativas de divergência como

$$D_{i,\psi}^{(l)} = -\log\left(\frac{1}{\bar{v}_i p(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}^{(l)})}\right), \quad i = 1, \dots, n, \quad l = 1, \dots, N, \quad (2.42)$$

e para cada i calculamos as proporções amostrais

$$P_i = \frac{1}{N} \sum_{l=1}^N I\left(D_{i,\psi}^{(l)} > \max_{j \neq i} D_{j,\psi}^{(l)}\right) \quad (2.43)$$

que providencia a estimativa da i -ésima observação ser um *outlier*. Note que P_i , obviamente, dependerá da função de ligação que será utilizada no modelo quando este for ajustado. Logo, um gráfico de P_i versus cada observação i será uma ferramenta bastante útil para comparação de modelos.

2.6.1 Divergência de Bregman Funcional Normalizada

Na Seção anterior, mostramos como usar a divergência de Bregman para poder avaliar a detecção de observações influentes no conjunto de dados. No entanto, ao utilizar a divergência de Bregman $D_{i,\psi} \in \mathbb{R}$ para $i = 1, \dots, n$ em sua escala original, podem surgir dúvidas sobre um ou mais valores serem substancialmente maiores que outros. Como definir o que seria uma observação significativamente influente? O que seria uma diferença relevante entre dois valores $D_{i,\psi}$ e $D_{j,\psi}$ para $i \neq j$?

Ainda na Seção anterior, apontamos uma forma de contornar essa questão, construindo uma metodologia baseada em Santos e Bolfarine (2016) para poder avaliar a chance de uma determinada observação i ser influente com base na medida apresentada em 2.43, que seria interpretada como a probabilidade da observação i ser um *outlier*. Porém, dentro desse contexto, Danilevicz e Ehlers (2022) propuseram uma metodologia para poder comparar duas densidades, reescalando a divergência de Bregman funcional para que ela possa variar no intervalo $(0, 1)$, facilitando assim a comparação e a interpretação.

Formalmente, dadas n divergências entre as densidades f_0 e f_1, \dots, f_n , que escrevemos como $D_\psi(f_0, f_1), D_\psi(f_0, f_2), \dots, D_\psi(f_0, f_n)$, se existe uma função $\|\psi\|$ para as quais as somas das n divergências $D_{\|\psi\|}(f_0, f_1), D_{\|\psi\|}(f_0, f_2), \dots, D_{\|\psi\|}(f_0, f_n)$ somam um, então $D_{\|\psi\|}(\cdot)$ é chamada de divergência de Bregman normalizada. Apresentado esse conceito, os autores definem o chamado operador \mathcal{B} que transforma qualquer divergência de Bregman em uma divergência de Bregman normalizada ao fazer

$$D_{\|\psi\|}(f_0, f_q) = \mathcal{B}(D_\psi(f_0, f_q)) = \frac{D_\psi(f_0, f_q)}{\sum_{i=1}^n D_\psi(f_0, f_i)}, \quad q = 1, \dots, n. \quad (2.44)$$

A vantagem dessa abordagem, como já mencionado, é que $0 \leq D_{\|\psi\|}(f_0, f_q) \leq 1$ para $q = 1, \dots, n$, facilitando tanto a comparação entre divergências, como a identificação de observações influentes. Danilevicz e Ehlers (2022) definem como ponto de partida para investigar possíveis observações candidatas à *outliers* o limite $1/n$, ou seja, se qualquer observação retorna $D_{\|\psi\|}(f_0, f_i) > 1/n$, então ela é uma candidata natural a ser um ponto influente.

2.7 Comparação de Modelos Baseada na Predição

Na Seção anterior, apresentamos algumas medidas de performance e seleção de modelos que eram baseadas na qualidade do ajuste. Nesta Seção, entretanto, apresentaremos medidas que também

servem para avaliar performance, bem como fazer seleção de modelos, mas que são construídas com base no poder de predição destes.

No contexto de classificação binária, a construção das métricas de avaliação de desempenho é baseada na chamada matriz de confusão, como apresentada na Tabela 4 (VUJOVIĆ *et al.*, 2021). Nessa Tabela, as linhas representam a classe predita e as colunas representam a classe (atual) observada. Fazendo todas as combinações possíveis, temos o surgimento das quantidades: TP (verdadeiro positivo, do inglês *True Positive*), a qual ocorre quando uma saída positiva é predita pelo modelo como positiva; FN (falso negativo, do inglês *False Negative*) quando o modelo prediz uma saída como negativa, mas na verdade ela é positiva; FP (falso positivo, do inglês *False Positive*) ocorre quando é predita uma saída como positiva, quando na verdade é negativa; e TN (verdadeiro negativo, do inglês *True Negative*) ocorre quando o modelo prediz uma observação negativa como negativa.

Tabela 4 – Matriz de confusão.

Valor predito	Valor observado	
	Positivo (1)	Negativo (0)
Positivo (1)	TP (Verdadeiro Positivo)	FP (Falso Positivo)
Negativo (0)	FN (Falso Negativo)	TN (Verdadeiro Negativo)

Fonte: Elaborada pelo autor.

Das quantidades TP, FP, FN e TN, medidas de desempenho podem ser construídas para avaliar modelos de classificação. Entre elas, podemos citar: Acurácia, Taxa de erro, Sensibilidade, Especificidade, Taxa de Verdadeiros Positivos e Taxa de Falsos Positivos. Essas métricas são detalhadas a seguir:

- 1) *Acurácia (ACC)*: é a soma de predições acuradas (TP+TN) sobre o total de observações.
- 2) *Taxa de erro (ER)*: calcula a proporção de predições incorretas pelo modelo. É uma métrica complementar a acurácia, ou seja $1-ACC$.
- 3) *Sensibilidade (SN)*: é calculada como o total de predições positivas corretas (TP) dividida pelo total de observações positivas (TP+FN) (VUJOVIĆ *et al.*, 2021).
- 4) *Especificidade (ES)*: é calculada como o total de predições negativas corretas (TN) dividida pelo total de observações negativas (TN+FP).
- 5) *Taxa de Verdadeiros Positivos (PRC)*: também chamada de precisão, essa métrica é calculada como o total de predições positivas corretas (TP) dividida pelo total de predições positivas (TP+FP).
- 6) *Taxa de Falsos Positivos (FPR)*: é calculada como o total de predições negativas corretas (TN) dividida pelo total de predições negativas (TN+FN). Essa métrica é complementar à especificidade, ou seja $1-ES$.

Dentro do cenário de desbalanceamento das classes, a escolha da métrica de avaliação é extremamente importante. Hossin e Sulaiman (2015) argumentam que quando há desbalanceamento de

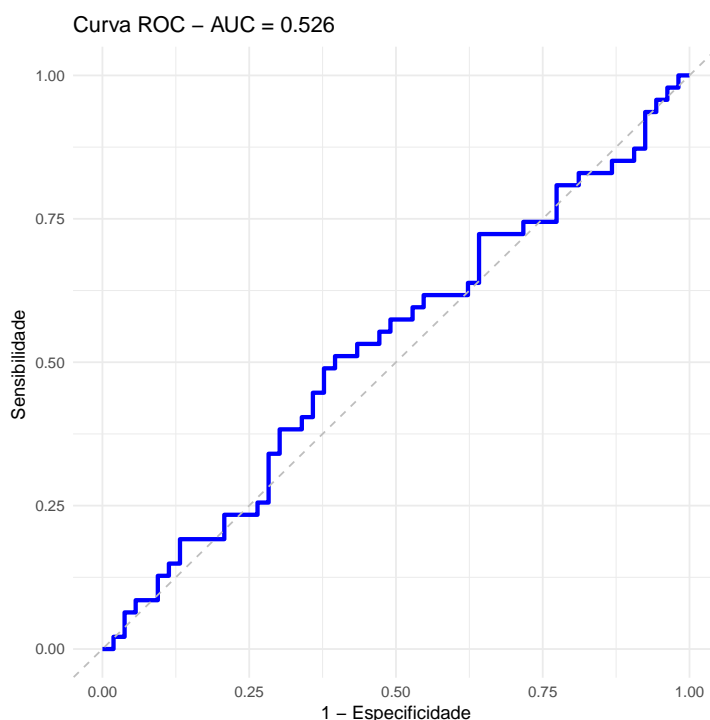
classes, a acurácia, por exemplo, não é tão informativa para avaliar o desempenho de classificadores. [Zou et al. \(2016\)](#) argumentam que a acurácia força os classificadores a minimizar a taxa de erro, e como resultado, classificadores tendem a demonstrar bom desempenho na classe majoritária, e mau desempenho na classe minoritária. Entretanto, em muitos problemas de classificação desbalanceada do mundo real, o erro de classificação da classe minoritária é muito mais caro que o erro de classificação da classe majoritária. Exemplo disso seria um diagnóstico de doença de câncer. É muito mais grave diagnosticar alguém como não tendo câncer, dado que ele tem (falso negativo), do que diagnosticar como tendo câncer, quando não tem (falso positivo), pois o primeiro caso poderia acarretar a morte do indivíduo ([ZOU et al., 2016](#)).

Portanto, métricas mais adequadas de desempenho para cenários de dados desbalanceados são necessárias. Nesse ínterim, medidas como: Índice de Jaccard (CSI) (também chamado de índice crítico de sucesso; CSI, do inglês *critical success index*), Índice de Sokal & Sneath (SSI, do inglês *Sokal & Sneath index*) e Índice de confiança (FAITH - do inglês *FAITH index*) são apresentadas em [Choi et al. \(2010\)](#) para medir a similaridade entre as classes observadas e as previstas pelo classificador. A medida F_β , Coeficiente Kappa de Cohen (KAPPA) e o Coeficiente de correlação de Matthews (MCC) ([HUAYANAY; BAZÁN; RUSSO, 2025](#)) também são métricas recomendadas quando há cenário de dados desbalanceados. Na Tabela 5, temos a descrição das métricas que serão utilizadas neste trabalho. Para todas as métricas nessa Tabela, as faixas de valores variam no intervalo $[0, 1]$, e quanto maior o valor para cada métrica, em tese, melhor o desempenho do classificador.

Em relação a métodos gráficos para avaliação de classificadores, a curva ROC (*Receiver Operating Characteristic*) é uma curva que descreve o *trade-off* de compensação entre benefícios (PRC) e custos (FRP). Nesse gráfico, o eixo x é igual a 1-Especificidade e o eixo y representa a Sensibilidade ([ZOU et al., 2016](#)), e os pontos são obtidos calculando a 1-Especificidade e a Sensibilidade para cada ponto de corte em $(0, 1)$. A Figura 5 apresenta a curva ROC de um modelo ingênuo, o qual classifica cada observação como positiva ou negativa com igual probabilidade. O modelo com discriminação perfeita corresponde àquele com Especificidade e Sensibilidade iguais a 1, ou seja $(x = 0, y = 1)$, e consequentemente valores de pontos de corte localizados próximo ao canto superior esquerdo do gráfico indicam que o modelo ajustado produz o maior percentual de acerto ([ZOU et al., 2016](#)). Através desse gráfico, surge uma outra métrica para avaliação de classificadores em cenário de dados desbalanceados, que é a área sob a curva ROC, chamada de AUC, que quanto mais próxima de 1 maior o poder preditivo do modelo. Na Tabela 5 apresentamos as métricas preditivas que usaremos neste trabalho.

Para finalizar esta Seção, convém frisar a importância da escolha do ponto de corte mencionado acima. O ponto de corte em classificação binária é, basicamente, um ponto p_c no intervalo $(0, 1)$, tal que a probabilidade predita pelo classificador sendo maior que este ponto, ou seja $P(\hat{Y} = 1) \geq p_c$, atribui a observação que está sendo predita à classe positiva (1), e caso contrário, à classe negativa (0). Percebe-se que, se o ponto de corte não for bem escolhido, o desempenho de determinado classificador poderá estar sendo mal avaliado e não sendo condizente com a realidade. Em muitos cenários, é comum fazer a escolha do ponto de corte para 0,5. No entanto, como argumentam [Zou et al. \(2016\)](#), fazer a escolha do ponto de corte para 0,5 em situações de classes desbalanceadas é inadequado. Por isso, usar uma metodologia que nos permita fazer a escolha apropriada para essa tarefa é essencial. [Assunção, Izbicki e Prates \(2024\)](#) mostraram que, em cenários de dados desbalanceados, a calibração do ponto de corte contribui para melhorar o desempenho preditivo do classificador, mesmo quando a probabilidade do evento de interesse

Figura 5 – Curva ROC associado a um modelo ingênuo.



Fonte: Elaborada pelo autor

Tabela 5 – Métricas de desempenho para cenário de desbalançamento. t_0 e t_1 são valores de ponto de corte em $(0, 1)$, sendo neste caso, $t_0 = 0$ e $t_1 = 1$.

Métrica	Fórmula
Sensibilidade (SN)	$\frac{TP}{TP+FN}$
Especificidade (ES)	$\frac{TN}{TN+FP}$
Precisão (PRC)	$\frac{TP}{TP+FP}$
Índice de Jaccard (CSI)	$\frac{TP}{TP+FP+FN}$
Índice de Sokal (SSI)	$\frac{TP}{TP+2 \times FP+2 \times FN}$
Índice de confiança (FAITH)	$\frac{TP+0,5 \times TN}{TP+FP+TN+FN}$
Coefficiente Kappa de Cohen (KAPPA)	$\frac{2 \times (TP \times TN - FP \times FN)}{(TP+FP)(FP+TN) + (TP+FN)(FN+TN)}$
Coefficiente de correlação de Matthews (MCC)	$\frac{(TP \times TN - FP \times FN)}{\sqrt{(FN+FP)(TP+FN)(TN+FP)(TN+FN)}}$
Medida - F_β (β)	$\frac{(\beta^2+1)PRC \times SN}{\beta^2 PR+SN}$
Área sob à curva ROC (AUC)	$\int_{t_0}^{t_1} ROC(t) dt$

Fonte: Elaborada pelo autor.

não é estimada corretamente — situação que ocorre, por exemplo, quando se utiliza uma função de ligação simétrica nesse contexto. Avaliaremos este resultado em nossas aplicações nos próximos Capítulos (Capítulos 3 e 4).

Nesta tese, a metodologia que utilizaremos será a mesma que foi brevemente mencionada em [Zou et al. \(2016\)](#), na qual o ponto de corte ideal seria aquele que otimizasse o valor de alguma métrica. Em nosso caso, usaremos a medida F_1 ($F_\beta(\beta)$ para $\beta = 1$). Por fim, vale destacar que a escolha do ponto de corte não altera o valor do AUC.

MODELOS DE REGRESSÃO BINÁRIA ESPACIAL

No presente Capítulo, apresentamos o primeiro dos modelos espaciais que serão trabalhados nesta tese. Iniciamos abordando os tipos de dados espaciais existentes na literatura (Seção 3.1), e dentre estes, destacando os dados de área, o tipo de dado espacial foco deste trabalho. Logo em seguida, descrevemos brevemente os modelos espaciais para dados de área, em especial apresentando a distribuição a priori G-Wishart para ser avaliada frente ao modelo CAR tradicional (Seção 3.2). Na Seção 3.3, apresentamos o nosso modelo de regressão binária espacial. Elaboramos um estudo de simulação para o modelo proposto na Seção 3.4, e na Seção 3.5, destacamos uma aplicação do modelo a dados reais. Por fim, apresentamos algumas conclusões e discussões sobre trabalhos futuros na Seção 3.6.

3.1 Tipos de Dados Espaciais

Dentre os muitos fenômenos que ocorrem nas mais diversas áreas do campo científico, não é difícil perceber que alguns deles apresentam variabilidade das observações no espaço e no tempo. Problemas de saúde pública, meio ambiente, geologia, epidemiologia, ecologia, processamento de imagens são algumas das áreas que podem ser citadas (REIS *et al.*, 2013; CRESSIE, 1993). Para esses problemas, é de suma importância construir modelos que possam não apenas indicar, como também quantificar a dependência espacial entre as medidas observadas.

Um problema contextualizado seria, por exemplo, o desejo de avaliar a relação entre a taxa de incidência de câncer de pulmão por município e ano, em um estado específico, com o hábito de fumar e outras informações importantes que pudessem estar disponíveis (BANERJEE; CARLIN; GELFAND, 2003). Para dados pluviométricos de uma dada região, um estado por exemplo, saber a previsão da quantidade de chuva que poderá ocorrer em determinados municípios, e quais os fatores que estão relacionados a esses fenômenos é de suma importância (CRESSIE, 1993).

Para que uma análise de dados espaciais possa ser realizada, os componentes básicos para tal tarefa constituem-se das localizações espaciais $\{\mathbf{s}_1, \dots, \mathbf{s}_m\}$, em que $\mathbf{s}_i \in \mathbb{R}^d$, um espaço euclidiano d -dimensional, e dos dados observados nessas localizações $\{Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_m)\}$, em que $Y(\mathbf{s}_i)$ representa

a variável aleatória observada na localização \mathbf{s}_i . Dependendo do tipo de aplicação que se busca avaliar, não somente os dados poderão ser considerados aleatórios, mas também as próprias localizações (CRESSIE, 1993).

Dito isso, para a construção de uma análise estatística mais robusta e consistente de dados espaciais, formalizamos a seguir uma definição intuitiva, sem rigor matemático, de um processo espacial. Essa definição é baseada em Cressie (1993), a qual os leitores interessados podem consultar, verificando as referências lá contidas para buscar detalhes matemáticos mais formais.

Basicamente, um processo espacial é um processo estocástico cujo domínio de variação é o espaço. Especificamente, seja $\mathbf{s} \in \mathbb{R}^d$ uma localização genérica contida em um espaço euclidiano d -dimensional, $D \subset \mathbb{R}^d$ um subconjunto d -dimensional, e $Y(\mathbf{s})$ uma medida observada na localização \mathbf{s} . Então, um processo espacial é definido por

$$\{Y(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}, \quad (3.1)$$

em que D representará um conjunto de índices sob o qual \mathbf{s} está variando.

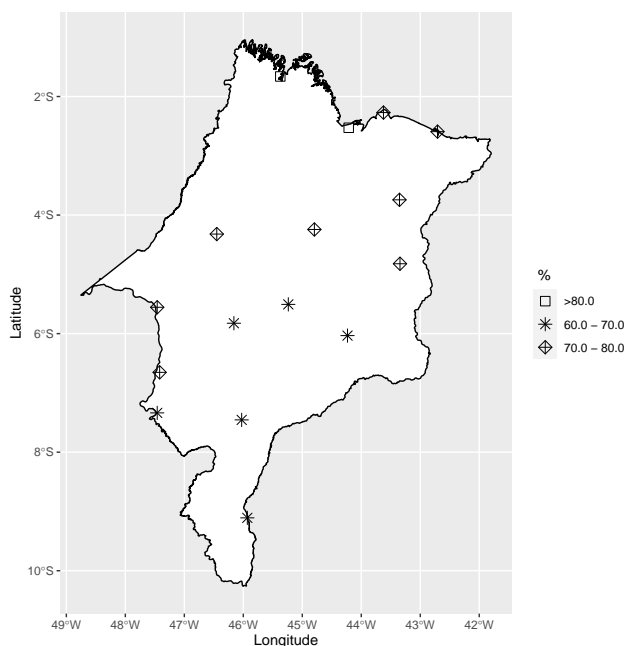
Através da formulação do processo espacial acima, podemos classificar conjuntos de dados espaciais em um dos três tipos básicos (BANERJEE; CARLIN; GELFAND, 2003):

1. **Dados Geoestatísticos:** o dado observado $Y(\mathbf{s})$ compõe um vetor aleatório sobre uma localização $\mathbf{s} \in \mathbb{R}^d$, onde \mathbf{s} varia continuamente sobre D , um subconjunto de \mathbb{R}^d considerado fixo.
2. **Dados de Área:** D é novamente considerado fixo, porém, agora particionado em um número finito de unidades de área (sub-regiões) com fronteiras bem definidas. A forma do subconjunto d -dimensional D pode ser regular ou irregular (como a área de um país por exemplo).
3. **Dados de Padrão Pontual (arranjos pontuais):** agora D é em si aleatório, apresentando apenas os índices das localizações de eventos aleatórios, que são os padrões de pontos. $Y(\mathbf{s})$ se limitará a indicar, basicamente, a ocorrência do evento, assumindo por exemplo valor igual a 1 para todo $\mathbf{s} \in D$. Para esse tipo de conjunto de dado, ainda há a possibilidade de que $Y(\mathbf{s})$ apresente informação de alguma outra variável disponível, representando o que chamamos de processo pontual marcado (*marked point pattern process*).

Exemplos de dados do primeiro tipo podem ser medições de volume de chuva em estações meteorológicas (REIS *et al.*, 2013), mapeamento de solo, previsão de concentrações de contaminantes em águas subterrâneas etc. Na Figura 6, temos uma ilustração para esse tipo de dado, cujo conteúdo é a apresentação do mapa das localizações de 16 estações de monitoramento de dados meteorológicos para o estado do Maranhão - Brasil no ano de 2010. A medida que está sendo plotada refere-se à média anual, para o ano de 2010, da umidade relativa do ar registrada nessas estações.

Para o segundo tipo de dado, exemplos como: número de casos de uma doença por municípios de um estado (REIS *et al.*, 2013), número de acidentes por bairros de um determinado município, taxa de incidência de uma determinada doença por estado no Brasil etc, podem ser citados. Para ilustrar esse segundo tipo de dado, na Figura 7 apresentamos o mapa da distribuição do PIB (Produto Interno Bruto) de 2019 por estado na região Nordeste - Brasil. Perceba que, diferentemente da ilustração apresentada na

Figura 6 – Mapa de estações (16 estações) de monitoramento de umidade relativa do ar para o estado brasileiro do Maranhão. A medida de plotagem refere-se à média da umidade relativa do ar para o ano de 2010.



Fonte: Instituto Nacional de Meteorologia - INMET

Figura 6, agora tem-se áreas bem definidas (cada estado da região nordeste) e uma medida agregada para toda aquela área (valor do PIB em 2019).

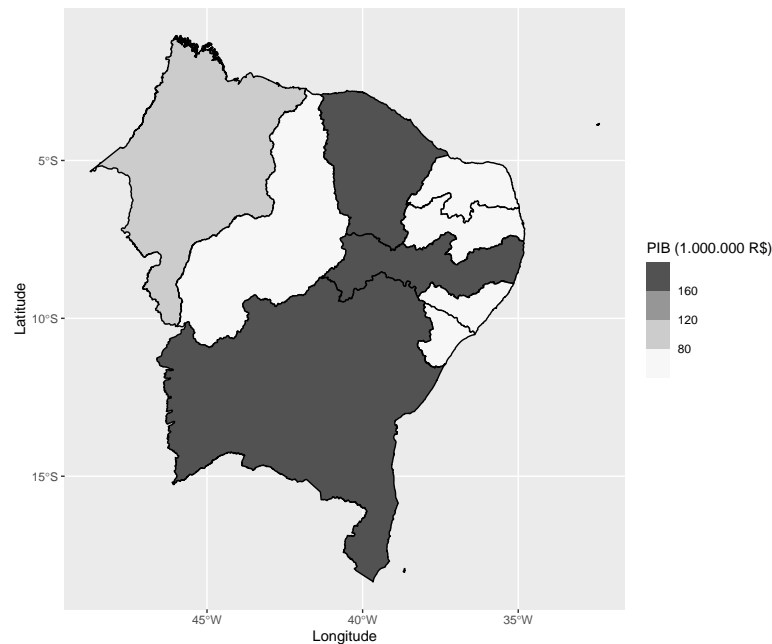
A diferença básica que distingue problemas de dados de área de problemas de dados geoestatísticos encontra-se no fato de que, neste último caso, a localização \mathbf{s} (o índice espacial) varia continuamente sobre D (CRESSIE, 1993), ao passo que, para o primeiro, as localizações são compostas por áreas com fronteiras bem definidas. O que há de comum nesses dois tipos de dados espaciais é que $Y(\mathbf{s})$ é considerado aleatório. Pensando em termos de modelagem, o objetivo de uma análise geoestatística seria fazer inferência sobre a distribuição espacial da superfície D a partir de informações fornecidas pela associação pareada entre os pontos amostrais $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_m)$, expressa em função da distância que os separa (ANSELIN, 2002).

De acordo com o tipo de dado espacial que se tem em mãos, e também levando em consideração os objetivos que se quer alcançar, haverá um conjunto de técnicas de análise mais apropriadas para a análise dos dados. Neste trabalho, exploraremos modelos para dados de área. O leitor interessado pode consultar a literatura adequada para avaliar as técnicas de análise para os outros tipos de dados espaciais.

3.2 Modelos Hierárquicos Espaciais para Dados de Área

O uso dos modelos espaciais são mais voltados para especificação de distribuições a priori do que para modelar os dados diretamente em si. Isto é, os efeitos espaciais aleatórios entram em um segundo estágio de uma especificação hierárquica bayesiana. Nesta tese, eles estarão sendo usados com esse objetivo. Na prática, especificamos uma distribuição Normal multivariada conjunta como distribuição a

Figura 7 – Mapa do PIB das unidades federativas da Região Nordeste (Brasil) para o ano de 2019.



Fonte: IBGE

priori para os efeitos aleatórios espaciais latentes com vetor de médias nulo e matriz de precisão Ω , que incorpora a dependência espacial entre as sub-regiões. No final das contas, precisamos construir uma forma de especificar a matriz Ω , e os modelos espaciais que aqui serão apresentados têm essa finalidade.

A especificação de uma distribuição Normal multivariada, para os efeitos aleatórios espaciais centrada em $\mathbf{0}$ e matriz de precisão Ω , advém de alguns resultados em modelagem espacial para dados de área. Especificamente, seja n_i o conjunto de sub-regiões adjacentes (ou vizinhas) à sub-região i . Assim, o fundamento de campo aleatório de Markov Gaussiano (GMRF - *Gaussian Markov random field*) permite-nos especificar a distribuição conjunta de Y_1, \dots, Y_m por meio das especificações condicionais $p(y_i | y_j, j \neq i) = p(y_i | y_j, j \in n_i)$, ou seja, as distribuições condicionais das sub-regiões $i = 1, \dots, m$ dependendo apenas de seus vizinhos, e o lema de Brook (BESAG, 1974; CRESSIE, 1993) garante a existência da distribuição conjunta de $p(y_1, \dots, y_m)$.

Em primeiro lugar, para poder construir um modelo para dados de área, o passo inicial seria definir associação espacial entre as unidades de área, que pode ser obtida por meio de uma estrutura de vizinhança baseada no arranjo das subáreas no mapa (veja a Figura 7 para fins de ilustração).¹ Antes de prosseguirmos, a partir de agora nos referiremos, sem prejuízo notacional, à variável aleatória $Y(\mathbf{s}_i)$ por apenas Y_i , para $i = 1, \dots, m$.

No caso de dados de área, considere, formalmente, as seguintes medidas observadas Y_1, Y_2, \dots, Y_m relacionadas às sub-regiões $1, 2, \dots, m$. Chamamos \mathbf{W} de matriz de proximidade quando as entradas w_{ij} em \mathbf{W} conectam espacialmente as unidades de área i e j de alguma forma (normalmente w_{ii} é configurado

¹ Em dados geoestatísticos, por exemplo, a estrutura de associação entre variáveis aleatórias em duas localizações distintas, digamos \mathbf{s} e \mathbf{s}' , se daria por meio de sua covariância, construída através de uma função que dependeria da distância entre as localizações.

para 0). De acordo com [Banerjee, Carlin e Gelfand \(2003\)](#), algumas opções para definição de \mathbf{W} podem ser consideradas, e entre elas citamos algumas:

- $w_{ij} = 1$ se i e j compartilham alguma fronteira comum, e nesse caso usamos a notação $i \sim j$; e 0 caso contrário. \mathbf{W} definido dessa forma é o que chamamos de matriz de adjacência;
- w_{ij} poderia refletir a “distância” entre as unidades de área, por meio, digamos, de distâncias intercentroidais, à semelhança do que é feito em dados geoestatísticos;
- w_{ij} poderia ser 1 para todo i, j dentro de uma certa distância especificada.

As formas de definição para \mathbf{W} são muitas. Uma das mais utilizadas considera \mathbf{W} com entradas w_{ij} como apresentado no primeiro ponto acima, ou seja, na forma de uma matriz de adjacência. Através da formalização de \mathbf{W} , podemos construir um mecanismo de associação espacial que será importante na modelagem formal dos dados. Para obter mais *insights* relacionados à matrizes de proximidade, os leitores podem consultar [Anselin \(2002\)](#), que revisa algumas questões importantes sobre a especificação de modelos espaciais no contexto de regressão, e entre essas questões, aponta algumas considerações relacionadas à definição de matrizes de proximidade (as quais o autor chama de *Spatial weights*).

Agora, considere y_i como sendo a resposta binária observada na sub-região i (para $i = 1, \dots, m$); $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^\top$, o vetor de covariáveis referente à sub-região i ; e $\boldsymbol{\beta}$, um vetor de coeficientes de regressão $k \times 1$. Sob a estrutura de um modelo de regressão binária espacial, $Y_i = 1$ com $p_i = F(\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i)$ e $1 - p_i$ para $Y_i = 0$, com $i = 1, \dots, m$, e o modelo hierárquico podendo ser especificado como

$$\prod_{i=1}^m p(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \phi_i) \times N_m(\boldsymbol{\phi} | \mathbf{0}, \boldsymbol{\Omega}) \times p(\boldsymbol{\beta}) = \prod_{i=1}^m [F(\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i)]^{y_i} [1 - F(\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i)]^{1-y_i} \times N_m(\boldsymbol{\phi} | \mathbf{0}, \boldsymbol{\Omega}) \times p(\boldsymbol{\beta}), \quad (3.2)$$

em que $p(\boldsymbol{\beta})$ refere-se à distribuição a priori para $\boldsymbol{\beta}$ e ϕ_i , com $i = 1, \dots, m$, aos efeitos aleatórios espaciais latentes.

O primeiro modelo que poderíamos propor para $\boldsymbol{\Omega}$ seria o modelo condicionalmente autoregressivo - CAR (*Conditional autoregressive*), introduzido por [Besag \(1974\)](#). Esse modelo é bem atraente, principalmente no contexto bayesiano, por sua conveniência computacional na implementação de alguns métodos de amostragem MCMC.

Esse modelo funciona baseado na intuição de que para se modelar os efeitos aleatórios espaciais, em termos da distribuição condicional de ϕ_i , $i = 1, \dots, m$, usaríamos apenas as informações das sub-regiões vizinhas atreladas à sub-região i ([BANERJEE; CARLIN; GELFAND, 2003](#)). Com base nesse fundamento, o modelo CAR descreve o relacionamento entre os efeitos aleatórios associados às unidades de área por meio da reunião de informações de suas unidades de área vizinha ([MORRIS et al., 2019](#)). A intuição de tais modelos é muito similar a dos modelos autoregressivos utilizados na modelagem de dados de séries temporais, nos quais observações feitas num determinado momento dependem do(s) momento(s) anterior(es).

Fazendo-se os ajustes necessários para que a distribuição conjunta de $\boldsymbol{\phi}$ seja própria, e a matriz de precisão seja não singular (os leitores podem consultar o Capítulo 3 de [Banerjee, Carlin e Gelfand \(2003\)](#)

para obter mais informações sobre esses detalhes), uma forma de especificação seria

$$\boldsymbol{\phi} \sim N_m(\mathbf{0}, \mathbf{D} - \rho \mathbf{W}), \quad (3.3)$$

em que $\mathbf{D} = \text{diag}(n_1, \dots, n_m)$ é uma matriz diagonal, e n_i é a quantidade de sub-regiões vizinhas à sub-região i (para $i = 1, \dots, m$); ρ é um parâmetro inserido tal que a matriz de precisão seja não singular; \mathbf{W} é a matriz de adjacência. Banerjee, Carlin e Gelfand (2003) argumentam que o intervalo de valores em que ρ pode ser escolhido tal que $\boldsymbol{\Omega}$ seja não singular é $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(m)})$, em que $\lambda_{(1)}, \dots, \lambda_{(m)}$ são os autovalores ordenados de $\mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$. Os leitores interessados em analisar uma forma de implementação do modelo CAR, podem consultar o link: <<https://mc-stan.org/users/documentation/case-studies/mbjoseph-CARStan.html>>. No Apêndice D, disponibilizamos uma breve introdução ao modelo espacial SAR.

Em 3.3, retirando ρ , teríamos uma matriz de precisão singular e a distribuição de $\boldsymbol{\phi}$ já não seria própria. Esse modelo passa a ser conhecido como ICAR (*intrinsic conditional autoregressive*). Porém, apesar de a distribuição desse modelo não ser própria, quando o usamos como distribuição a priori, não há prejuízos em termos de modelagem, pois à luz dos dados (y_1, \dots, y_m) , a distribuição a posteriori se tornaria própria.

Até agora a especificação de $\boldsymbol{\Omega}$ se deu por meio de formas paramétricas, utilizando os modelos CAR ou ICAR. No entanto, um tratamento totalmente bayesiano pode ser oferecido ao problema quando, ao invés de assumirmos uma forma paramétrica conhecida para a matriz de precisão, atribuímos uma distribuição a priori para $\boldsymbol{\Omega}$ para representar a incerteza sobre a relação espacial entre os efeitos aleatórios.

Nesse enfoque, a distribuição G-Wishart é uma das alternativas disponíveis como distribuição a priori para modelos de grafos Gaussianos (ROVERATO, 2002). Essa distribuição é uma priori conjugada e possui aplicação em distintos problemas, tais como Física, Biomedicina e Tecnologia. Em outras palavras, estaríamos fazendo $\boldsymbol{\Omega} \sim \text{G-Wishart}_{\mathbf{W}}(\kappa, \mathbf{S})$, em que G-Wishart significa uma distribuição G-Wishart, κ é o parâmetro referente aos graus de liberdade, e \mathbf{S} é uma matriz escala restrita a ter entradas nulas para cada zero na matriz de adjacência \mathbf{W} . Essa distribuição possui a seguinte forma:

$$p(\boldsymbol{\Omega} | \mathbf{W}) = \frac{1}{Z_{\mathbf{W}}(\kappa, \mathbf{S})} |\boldsymbol{\Omega}|^{(\kappa-2)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{S}\boldsymbol{\Omega}) \right\} I(\boldsymbol{\Omega} \in M_{\mathbf{W}}), \quad \kappa > 2, \quad (3.4)$$

na qual $M_{\mathbf{W}}$ representa o conjunto de matrizes simétricas positivas definidas restritas a terem entradas nulas fora da diagonal referente aos zeros em \mathbf{W} , e $Z_{\mathbf{W}}(\kappa, \mathbf{S})$ é a constante de normalização.

Segundo Bandyopadhyay e Canale (2016), o modelo CAR é restritivo em não permitir não estacionariedade, autocorrelação espacialmente variável e parâmetros de suavização. Contudo, a especificação G-Wishart permite flexibilidade ao eliminar a suposição de estacionariedade, ou seja, considera que possam haver áreas dentro de uma determinada região com dinâmica espacial própria, e facilita a implementação computacional por meio da conjugação, uma vez que usamos uma estrutura bayesiana. Neste e no próximo Capítulo, apresentaremos os modelos espaciais focando na distribuição a priori G-Wishart $_{\mathbf{W}}(\kappa, \mathbf{S})$.

3.3 Modelo de Regressão Binária Espacial

A partir da apresentação em 3.2, o modelo de regressão binária espacial com função de ligação flexível é apresentado da seguinte forma:

$$Y_i | \phi_i \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p_i), \quad i = 1, \dots, m,$$

$$p_i = F_\delta(\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i) \quad (3.5)$$

$$\boldsymbol{\beta} \sim p(\boldsymbol{\beta}), \quad \delta \sim p(\delta)$$

$$\boldsymbol{\phi} \sim N_m(\mathbf{0}, \boldsymbol{\Omega})$$

$$\boldsymbol{\Omega} \sim \text{G-Wishart}_{\mathbf{W}}(\kappa, \mathbf{S}), \quad (3.6)$$

em que $p(\boldsymbol{\beta})$ e $p(\delta)$ denotam as densidades a priori, e $F_\delta(\cdot)$ será uma das fda's apresentadas na Tabela 2. O subscrito δ em $F_\delta(\cdot)$ indica que, nesta tese, o modelo em 3.5 e os modelos que serão apresentados ao longo dos capítulos estão sendo implementados usando uma reparametrização conveniente, a saber, $\delta = \log(\lambda)$, como abordado em Alves, Bazán e Arellano-Valle (2023). Para \mathbf{S} , a matriz escala da distribuição G-Wishart em 3.5, faremos $\mathbf{S} = \mathbf{D} - \rho \mathbf{W}$, ou seja, essa especificação corresponde à distribuição a priori própria CAR apresentada em 3.3. Na prática, valores típicos de ρ giram em torno de 0,80-0,90, então aqui fixamos $\rho = 0,90$. O valor do parâmetro de graus de liberdade κ foi fixado em $\kappa = m$, conforme adotado em Bandyopadhyay e Canale (2016).

Assumindo que $\boldsymbol{\beta}$, δ e $\boldsymbol{\phi}$ são a priori independentes, a distribuição a posteriori conjunta é dada por

$$p(\boldsymbol{\beta}, \delta, \boldsymbol{\phi}, \boldsymbol{\Omega} | \mathbf{y}, \mathbf{X}) \propto \prod_{i=1}^m p(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \delta, \phi_i) p(\boldsymbol{\phi} | \boldsymbol{\Omega}) p(\boldsymbol{\Omega} | \mathbf{W}) p(\boldsymbol{\beta}) p(\delta). \quad (3.7)$$

Aqui, atribuímos distribuições a priori normais independentes centradas em torno de zero para todos os coeficientes de regressão, ou seja, $\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_\beta^2 \mathbf{I}_k)$, em que $\sigma_\beta^2 > 0$, e $\delta \sim U(-2, 2)$ de acordo com Bazán *et al.* (2017) e Alves, Bazán e Arellano-Valle (2023). Percebe-se que essa distribuição a posteriori conjunta é analiticamente intratável e recorreremos a aproximações baseadas em técnicas de Monte Carlo via Cadeias de Markov (MCMC) para realizar uma análise totalmente bayesiana.

A seguir, obtemos as distribuições a posteriori condicionais completas para os blocos de parâmetros associados, isto é: i) $(\boldsymbol{\beta}, \delta)$, ii) $\boldsymbol{\Omega}$ e iii) $\boldsymbol{\phi}$. Antes disso, cabe ressaltar que todos os algoritmos dos amostradores das distribuições condicionais completas, com exceção da distribuição condicional completa de $\boldsymbol{\Omega}$, para os modelos apresentados neste Capítulo e no Capítulo 4, foram implementados via linguagem de programação C++, para obtermos ganhos de rapidez computacional através dos pacotes Rcpp e RcppArmadillo, usando o software R como interface. No Apêndice A, apresentamos uma breve introdução ao Rcpp, para implementações computacionais iniciais.

A proporcional da distribuição condicional completa conjunta de $(\boldsymbol{\beta}, \delta)^2$ é dada por

$$p(\boldsymbol{\beta}, \delta | \boldsymbol{\phi}, \boldsymbol{\Omega}, \mathbf{y}, \mathbf{X}) \propto \prod_{i=1}^m \left[F_\delta(\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i) \right]^{y_i} \prod_{i=1}^m \left[1 - F_\delta(\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i) \right]^{1-y_i} \prod_{j=1}^k \exp \left\{ -\frac{\beta_j^2}{2 \times 10^2} \right\} \frac{1}{4}. \quad (3.8)$$

² veja que $\sigma_\beta^2 = 10^2$ em 3.8

Para amostrar valores da distribuição condicional completa conjunta de $(\boldsymbol{\beta}, \delta)$, usaremos o método de Monte Carlo Hamiltoniano com a métrica de Riemann, o RMHMC, apresentado na Seção 2.2.3.1. As derivadas parciais requeridas e matrizes de informação de Fisher necessárias podem ser adquiridas no Apêndice B, fazendo $r = 1$. Nesse Apêndice, as derivadas parciais e matrizes de informação foram calculadas para o modelo mais geral apresentado no Capítulo 4, que fornece o modelo em 3.5 como caso particular quando $r = 1$.

A distribuição a posteriori condicional completa de $\boldsymbol{\Omega}$ é obtida observando-se que

$$p(\boldsymbol{\phi}|\boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\phi}^\top \boldsymbol{\Omega} \boldsymbol{\phi} \right\} \propto |\boldsymbol{\Omega}|^{1/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\phi}^\top \boldsymbol{\Omega} \boldsymbol{\phi}) \right\} \quad (3.9)$$

$$\propto |\boldsymbol{\Omega}|^{1/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\phi} \boldsymbol{\phi}^\top \boldsymbol{\Omega}) \right\} \propto |\boldsymbol{\Omega}|^{1/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\phi} \boldsymbol{\phi}^\top \boldsymbol{\Omega}) \right\}, \quad (3.10)$$

e como $\boldsymbol{\phi}|\boldsymbol{\Omega} \sim N_m(\mathbf{0}, \boldsymbol{\Omega})$ segue que

$$p(\boldsymbol{\Omega}|\boldsymbol{\beta}, \delta, \boldsymbol{\phi}, \mathbf{y}, \mathbf{X}) \propto |\boldsymbol{\Omega}|^{(1+\kappa-2)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[(\mathbf{S} + \boldsymbol{\phi} \boldsymbol{\phi}^\top) \boldsymbol{\Omega} \right] \right\}. \quad (3.11)$$

Então, a distribuição condicional completa é dada por $\boldsymbol{\Omega}|\boldsymbol{\beta}, \delta, \boldsymbol{\phi}, \mathbf{y}, \mathbf{X} \sim \text{G-Wishart}_{\mathbf{W}}(\kappa + 1, \mathbf{S} + \boldsymbol{\phi} \boldsymbol{\phi}^\top)$, com a matriz de escala $\mathbf{S} + \boldsymbol{\phi} \boldsymbol{\phi}^\top$ restrita a ter entradas nulas fora da diagonal, correspondendo a zeros na matriz de adjacência \mathbf{W} . A matriz de precisão $\boldsymbol{\Omega}$ conterá muitos zeros dependendo da estrutura de vizinhança espacial do problema em análise, ou seja, é uma matriz esparsa. Por isso, para simular os valores da distribuição G-Wishart, adotaremos o esquema de amostragem exata proposto por Mohammadi e Wit (2015), que leva em conta a esparsidade da matriz de precisão. Nesse trabalho, os autores propõem um método de Monte Carlo via cadeias de Markov transdimensional baseado em um processo de nascimento e morte de tempo contínuo, cuja eficiência foi avaliada por meio de vários experimentos simulados e também por meio de análise de dados reais. Os resultados desse método estão disponíveis no pacote R Bdgraph (MOHAMMADI; WIT, 2019).

A distribuição condicional a posteriori completa do vetor de efeitos aleatórios espaciais $\boldsymbol{\phi}$ é dada por

$$\begin{aligned} p(\boldsymbol{\phi}|\boldsymbol{\beta}, \delta, \boldsymbol{\Omega}, \mathbf{y}, \mathbf{X}) &\propto p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}) p(\boldsymbol{\phi}|\boldsymbol{\Omega}) \\ &\propto \prod_{i=1}^m \left[F_\delta(\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i) \right]^{y_i} \prod_{i=1}^m \left[1 - F_\delta(\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i) \right]^{1-y_i} \exp \left\{ -\frac{1}{2} \boldsymbol{\phi}^\top \boldsymbol{\Omega} \boldsymbol{\phi} \right\}. \end{aligned} \quad (3.12)$$

Para amostrar valores da distribuição do vetor de efeitos aleatórios espaciais, usaremos o método HMC, apresentado na sua forma padrão (ou seja, a matriz de massas é dada por $\mathbf{M} = \mathbf{I}_m$) na Seção 2.2.3. Os vetores de derivadas parciais dos efeitos aleatórios espaciais também são fornecidos no Apêndice B quando $r = 1$.

Ao escolher amostrar valores dos parâmetros do modelo particionado em blocos (por meio das distribuições condicionais completas) ao invés de usar um software de programação probabilística como o Stan (TEAM *et al.*, 2018), para amostrar valores da distribuição a posteriori de forma completa, há a vantagem de poder escolher o método que melhor proporcione eficiência computacional de acordo com as características da distribuição condicional completa daquele parâmetro.

Para finalizar esta Seção, obtemos a média Preditiva a posteriori do modelo de regressão binária espacial 3.5 para a observação \tilde{y}_i da seguinte forma:

$$\mu_i = E(\tilde{y}_i | \mathbf{y}, \mathbf{x}_i) = E[E(\tilde{y}_i | \boldsymbol{\theta}, \mathbf{y}, \mathbf{x}_i)] = E[F_\delta((\mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i), \delta) | \mathbf{y}, \mathbf{x}_i], \quad (3.13)$$

que, de posse de uma amostra da distribuição a posteriori de $\{\boldsymbol{\beta}^{(l)}, \delta^{(l)}, \boldsymbol{\phi}^{(l)}\}_{l=1}^N$, pode ser aproximada como

$$\tilde{\mu}_i = \frac{1}{N} \sum_{l=1}^N [F_\delta(\mathbf{x}_i^\top \boldsymbol{\beta}^{(l)} + \phi_i^{(l)}; \delta^{(l)})]. \quad (3.14)$$

3.4 Estudo de Simulação

Nesta Seção, avaliamos o desempenho dos modelos e algoritmos propostos descritos nas Seções anteriores por meio de um estudo de simulação. O estudo de simulação foi dividido em duas partes. Primeiro, realizamos uma bateria de simulações para o modelo proposto neste trabalho, utilizando as funções de ligação flexíveis da Tabela 2 sob diferentes configurações do parâmetro de forma λ e para diferentes tamanhos de amostra. Na segunda parte, avaliamos o cenário no qual geramos dados do modelo de ligação simétrico Logito e avaliamos o desempenho dos modelos de ligação assimétrico propostos para verificar sua robustez em um cenário de especificação incorreta.

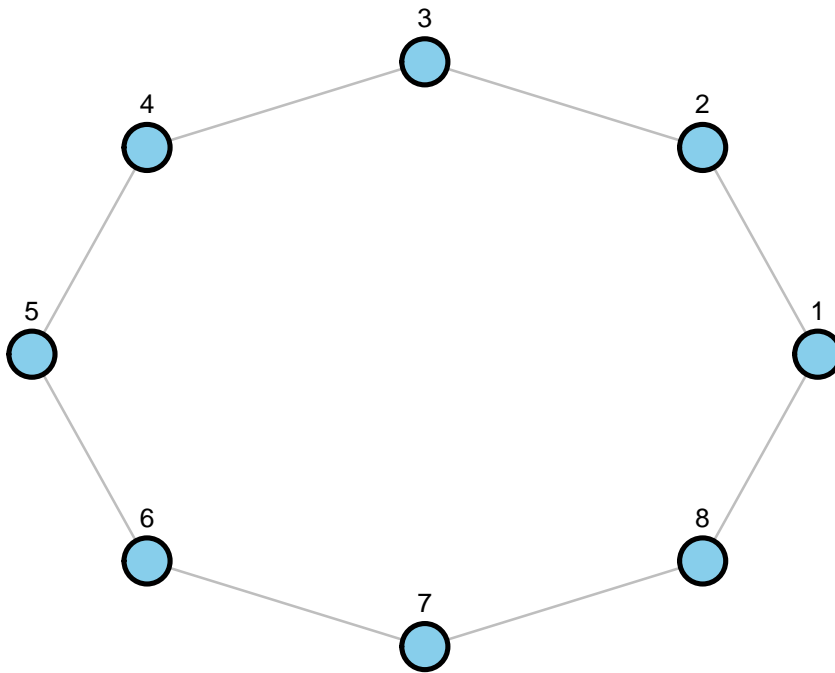
Para o primeiro contexto, geramos $M = 100$ replicações de conjuntos de dados binários com tamanhos de amostra de $m = 200$ e $m = 500$. Os valores de duas covariáveis, x_1 e x_2 , foram gerados independentemente de uma distribuição normal padrão, enquanto os coeficientes de regressão foram fixados em $\beta_1 = -0,5$ e $\beta_2 = 0,5$. Um vetor de respostas binárias y_i foi então gerado de acordo com as probabilidades $p_i = F_\delta(\beta_1 x_{i1} + \beta_2 x_{i2} + \phi_i)$ (modelo em 3.5) para cada um dos modelos na Tabela 2, em que $p_i = P(y_i = 1)$, $i = 1, \dots, m$. O parâmetro de forma foi fixado em $\lambda = 0,5$ e $\lambda = 4$ (respectivamente, $\delta = -0.693$ e $\delta = 1.386$) para investigar a flexibilidade do nosso modelo para diferentes graus de assimetria. Finalmente, os efeitos aleatórios espaciais foram gerados de acordo com uma distribuição Normal multivariada com matriz de precisão G-Wishart esparsa, ou seja, $\boldsymbol{\phi} \sim N(\mathbf{0}, \boldsymbol{\Omega}^{-1})$ com $\boldsymbol{\Omega} \sim \text{G-Wishart}_{\mathbf{W}}(\kappa = m, \mathbf{S})$. Para todas as configurações e tamanhos de amostra, em cada replicação extraímos 5000 amostras MCMC, descartando as primeiras 2500 como *burn-in*.

Na segunda metade do estudo de simulação, a ideia é usar a função de ligação simétrica Logito para gerar um cenário de especificação incorreta, ou seja, $p_i = 1 / (1 + \exp(-(\beta_1 x_{i1} + \beta_2 x_{i2} + \phi_i)))$, e assim avaliar a robustez dos modelos e algoritmos propostos. Para esse cenário, geramos $M = 100$ réplicas de conjuntos de dados binários com tamanhos de amostra $m = 200$, duas covariáveis x_1 e x_2 geradas independentemente de uma distribuição normal padrão e coeficientes $\beta_1 = -0,5$ e $\beta_2 = 0,5$. Efeitos aleatórios espaciais também foram gerados como antes, a partir de uma distribuição Normal multivariada, e com matriz de precisão segundo uma distribuição G-Wishart $_{\mathbf{W}}$ e $\kappa = m$. Finalmente, respostas binárias foram geradas com probabilidades p_i . Similarmente, 5000 amostras MCMC foram coletadas, descartando as primeiras 2500 como *burn-in* para cada réplica e tamanho de amostra.

A estrutura espacial das sub-regiões, para os dois cenários de simulação, foi gerada de um grafo circular, tal que, cada sub-região possui duas sub-regiões vizinhas. A Figura 8 ilustra essa estrutura de adjacência para oito sub-regiões.

Foram feitas simulações iniciais com o objetivo de otimizar os parâmetros pertinentes ao HMC, a saber, o tamanho do passo ϵ e a quantidade de passos *leapfrog* L . Basicamente, os critérios utilizados para a escolha desses valores baseou-se em avaliar a taxa de aceitação a novas propostas pelo algoritmo, e em avaliar o gráfico de convergência das cadeias dos parâmetros a serem estimados. Assim, para o bloco de parâmetros (β, δ) , sob todas as funções de ligação avaliadas neste trabalho, configuramos o valor do tamanho do passo para $\epsilon = 0.1$ e a quantidade de passos para $L = 22$; para o vetor de efeitos aleatórios espaciais ϕ , fizemos $\epsilon = 0.02$ e $L = 22$.

Figura 8 – Grafo circular para 8 sub-regiões.



Fonte: Elaborada pelo autor.

O desempenho do esquema MCMC abordado nesta tese foi avaliado em termos da qualidade de estimativa de parâmetros. A raiz do erro quadrático médio (RMSE - *root mean square error*), o viés relativo (RBias - *relative bias*) e a probabilidade de cobertura (CP - *coverage probability*) são definidos abaixo:

$$RMSE^2 = \frac{1}{M} \sum_{i=1}^M (\hat{\theta}^{(i)} - \theta)^2$$

$$RBias = \left(\frac{1}{M} \sum_{i=1}^M \hat{\theta}^{(i)} - \theta \right) / \theta$$

$$CP = \frac{1}{M} \sum_{i=1}^M I(\theta \in [\hat{\theta}_{0.025}^{(i)} - \hat{\theta}_{0.975}^{(i)}]).$$

$\hat{\theta}^{(i)}$ e $\hat{\theta}_{\alpha}^{(i)}$ são, respectivamente, a estimativa pontual e o α -percentil estimado de um parâmetro θ na i -ésima replicação, $i = 1, 2, \dots, M$. Usamos as médias a posteriori como estimativas pontuais, e também incluímos uma versão normalizada do RMSE (NRMSE) simplesmente dividindo pela média da amostra

dos valores RMSE. Além dessas métricas de desempenho, incluímos, para cada modelo, o tempo médio de execução (em segundos) em cada replicação.

Os resultados da primeira parte das simulações são mostrados nas Tabelas 6 e 7. Para $m = 200$, obtivemos bons resultados para todos os modelos em termos de viés relativo, RMSE, NRMSE e probabilidades de cobertura para todas as configurações de λ . Para a configuração $m = 500$ e independentemente do valor de λ , pode-se observar probabilidades de cobertura ligeiramente menores para o modelo GR-RP quando comparado aos outros modelos. Em geral, os modelos tiveram bom desempenho em todas as configurações testadas.

Analisando o tempo computacional, notamos que, como esperado, para um tamanho amostral menor, o tempo de execução dos modelos em cada réplica é menor quando comparado a um tamanho amostral maior. Também constatamos que para $m = 200$ o tempo de execução de cada modelo em cada réplica é mais estável, porém, para $m = 500$, notamos maiores diferenças de tempo em cada réplica para todos os modelos, com destaque para os modelos C-RP e GR-RP.

A Tabela 8 contém os resultados das análises de robustez dos modelos com funções ligação flexíveis baseado na qualidade do ajuste. Além das métricas de qualidade de ajuste apresentadas no Capítulo 2, incluímos as métricas EAIC (*expected Akaike information criterion*), EBIC (*expected bayesian information criterion*) (ASL; RASHIDI; ABAD, 2021) e LPML (*log-pseudo verossimilhança marginal*) (GEISSER; EDDY, 1979). Pela Tabela, notamos que os modelos se adaptaram muito bem ao cenário atípico, e apresentaram métricas de ajuste muito próximas às apresentadas pelo modelo de dados original, ou seja, o modelo Logito. Através desse estudo, podemos constatar o poder da flexibilidade desses modelos, mostrando que eles podem oferecer bons ajustes em um cenário adverso.

Vale destacar que, dos modelos considerados, os modelos L-P e L-RP são tidos como generalizações do modelo Logito, e este último torna-se um caso particular dos primeiros quando $\lambda = 1$ (equivalentemente, $\delta = 0$). Neste estudo de robustez, essa capacidade de identificação da assimetria por meio do λ (δ) também pôde ser observada, pois a média das estimativas de δ em 100 replicações foi $\hat{\delta} = -0.060$ (intervalo de credibilidade de 95% $(-0.264; 0.141)$) para o modelo L-P, e $\hat{\delta} = 0.033$ (intervalo de credibilidade de 95% $(-0.161; 0.227)$) para o modelo L-RP.

Antes de finalizar esta Seção, vale destacar que no Apêndice D (Seção D.3), apresentamos um estudo das métricas de predição para cada modelo, no qual fornecemos mais *insights*.

Tabela 6 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $m = 200$ e $m = 500$. 5000 iterações, descartando 2500 como *burn-in*. Modelos: C-P, C-RP, GR-P, GR-RP e LA-P

Ligação	Parâmetro	$m = 200$						$m = 500$					
		Rbias	RMSE	NRMSE	CP	Tempo médio	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	
C-P	$\lambda = 0,5$	β_1	0.256	0.331	-0.527	0.95		β_1	0.109	0.183	-0.330	0.96	
		β_2	0.133	0.368	0.649	0.85	76.95 (8.22)	β_2	0.058	0.194	0.367	0.91	3238.71 (38.05)
		δ	0.083	0.187	-0.249	0.91		δ	0.049	0.135	-0.186	0.92	
C-P	$\lambda = 4$	β_1	0.145	0.252	-0.441	0.90		β_1	0.009	0.153	-0.304	0.91	
		β_2	0.202	0.282	0.469	0.90	540.40 (812.56)	β_2	0.084	0.162	0.300	0.93	3253.80 (42.67)
		δ	0.104	0.505	0.33	0.89		δ	0.202	0.744	0.446	0.84	
C-RP	$\lambda = 0,5$	β_1	0.102	0.291	-0.529	0.95		β_1	0.316	0.326	-0.495	0.91	
		β_2	0.166	0.287	0.492	0.97	19.44 (3.08)	β_2	0.208	0.259	0.429	0.95	537.51 (567.66)
		δ	0.054	0.152	-0.209	0.94		δ	0.059	0.143	-0.194	0.97	
C-RP	$\lambda = 4$	β_1	0.147	0.242	-0.423	0.93		β_1	0.412	0.4	-0.567	0.85	
		β_2	0.184	0.259	0.437	0.93	19.03 (2.07)	β_2	0.458	0.44	0.603	0.81	658.036 (288.69)
		δ	0.034	0.241	0.168	0.87		δ	0.09	0.269	0.178	0.84	
GR-P	$\lambda = 0,5$	β_1	0.010	0.154	-0.306	0.91		β_1	0.037	0.101	-0.195	0.95	
		β_2	0.099	0.162	0.296	0.92	485.14 (2.03)	β_2	0.007	0.101	0.2	0.93	3242.99 (10.44)
		δ	0.066	0.147	-0.198	0.96		δ	0.008	0.104	-0.148	0.95	
GR-P	$\lambda = 4$	β_1	0.008	0.087	0.173	0.91		β_1	0.026	0.061	-0.120	0.89	
		β_2	0.011	0.095	0.188	0.90	489.06 (4.30)	β_2	0.020	0.066	0.129	0.86	3235.3 (10.89)
		δ	0.02	0.137	0.097	0.91		δ	0.006	0.09	0.065	0.91	
GR-RP	$\lambda = 0,5$	β_1	0.075	0.156	-0.229	0.92		β_1	0.026	0.106	-0.206	0.97	
		β_2	0.134	0.177	0.312	0.85	484.21 (5.09)	β_2	0.009	0.117	0.231	0.89	9969.41 (5398.81)
		δ	0.105	0.181	-0.237	0.90		δ	0.042	0.114	-0.158	0.95	
GR-RP	$\lambda = 4$	β_1	0.012	0.081	-0.159	0.94		β_1	0.006	0.06	-0.119	0.89	
		β_2	-0.008	0.07	0.142	0.97	555.988 (203.01)	β_2	0.01	0.056	0.11	0.94	3496.373 (774.82)
		δ	0.012	0.113	0.081	0.96		δ	0.017	0.077	0.055	0.99	
LA-P	$\lambda = 0,5$	β_1	0.098	0.167	-0.304	0.97		β_1	0.07	0.125	-0.234	0.92	
		β_2	0.122	0.199	0.354	0.93	490 (14.61)	β_2	0.094	0.136	0.249	0.93	3246.36 (50.04)
		δ	0.051	0.158	-0.216	0.91		δ	0.033	0.104	-0.145	0.96	
LA-P	$\lambda = 4$	β_1	0.136	0.175	-0.308	0.88		β_1	0.061	0.105	-0.198	0.92	
		β_2	0.127	0.165	0.293	0.95	486.24 (1.79)	β_2	0.077	0.115	0.213	0.93	3246.49 (29.64)
		δ	0.057	0.179	0.122	0.94		δ	0.02	0.117	0.083	0.92	

Fonte: Elaborada pelo autor.

Tabela 7 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $m = 200$ e $m = 500$. 5000 iterações, descartando 2500 como *burn-in*. Modelos: L-P, L-RP, N-P, N-RP e LA-RP

Ligação	$m = 200$						$m = 500$						
	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	
L-P	$\lambda = 0,5$	β_1	0.091	0.252	-0.462	0.96	488.4818 (5.13)	β_1	0.047	0.153	-0.292	0.91	
		β_2	0.04	0.23	0.442	0.95		β_2	0.002	0.155	0.309	0.93	3248.05 (9.31)
		δ	0.068	0.162	-0.219	0.92		δ	0.032	0.1	-0.14	0.93	
	$\lambda = 4$	β_1	0.125	0.156	-0.278	0.94	484.2 (1.74)	β_1	0.011	0.109	-0.216	0.95	
		β_2	0.048	0.159	0.303	0.94		β_2	0.053	0.119	0.226	0.94	3253.58 (25.7)
		δ	0.026	0.115	0.081	0.98		δ	0.017	0.094	0.066	0.91	
L-RP	$\lambda = 0,5$	β_1	0.093	0.23	-0.421	0.92	489 (4.56)	β_1	0.06	0.167	-0.315	0.94	
		β_2	0.124	0.205	0.365	0.96		β_2	0.045	0.14	0.269	0.98	3243.46 (33.88)
		δ	0.031	0.149	-0.208	0.93		δ	0.041	0.111	-0.154	0.89	
	$\lambda = 4$	β_1	0.077	0.154	-0.285	0.94	487.29 (4.48)	β_1	0.027	0.105	-0.204	0.98	
		β_2	0.058	0.177	0.335	0.93		β_2	0.052	0.116	0.22	0.95	3247.41 (11.77)
		δ	0.029	0.143	0.1	0.93		δ	0.023	0.097	0.069	0.91	
N-P	$\lambda = 0,5$	β_1	0.076	0.143	-0.265	0.95	480 (4.61)	β_1	0.051	0.098	-0.186	0.97	
		β_2	0.027	0.171	0.333	0.9		β_2	0.029	0.108	0.209	0.92	3235.85 (11.53)
		δ	0.045	0.137	-0.189	0.93		δ	0.05	0.12	-0.165	0.88	
	$\lambda = 4$	β_1	0.035	0.102	-0.196	0.96	484.1 (1.76)	β_1	0.027	0.072	-0.14	0.92	
		β_2	0.06	0.112	0.212	0.9		β_2	0.05	0.078	0.148	0.93	3247.38 (26.65)
		δ	0.046	0.156	0.107	0.95		δ	0.018	0.109	0.077	0.91	
N-RP	$\lambda = 0,5$	β_1	0.053	0.142	-0.269	0.95	483 (3.94)	β_1	0.035	0.09	-0.174	0.98	
		β_2	0.068	0.158	0.296	0.92		β_2	0.033	0.111	0.215	0.91	3233.24 (42.3)
		δ	0.032	0.157	-0.22	0.91		δ	0.009	0.102	-0.146	0.91	
	$\lambda = 4$	β_1	0.03	0.103	-0.2	0.91	482.36 (2.59)	β_1	0.03	0.07	-0.137	0.97	
		β_2	0.003	0.107	0.212	0.9		β_2	0.039	0.071	0.137	0.94	3249.16 (34.29)
		δ	0.026	0.144	0.101	0.9		δ	0.015	0.097	0.069	0.94	
LA-RP	$\lambda = 0,5$	β_1	0.136	0.197	-0.348	0.94	488.7543 (3.35)	β_1	0.102	0.131	-0.238	0.95	
		β_2	0.124	0.208	0.371	0.94		β_2	0.062	0.14	0.264	0.91	3245.56 (12)
		δ	0.079	0.17	-0.227	0.89		δ	0.036	0.099	-0.138	0.97	
	$\lambda = 4$	β_1	0.153	0.156	-0.271	0.92	487.3 (4.1)	β_1	0.069	0.099	-0.186	0.96	
		β_2	0.118	0.155	0.277	0.94		β_2	0.054	0.1	0.189	0.94	3243.77 (9.84)
		δ	0.044	0.184	0.127	0.86		δ	0.035	0.121	0.085	0.87	

Fonte: Elaborada pelo autor.

Tabela 8 – Comparação de modelos para diferentes conjuntos de dados de tamanho 100, gerados a partir de um modelo Logito.

Ligação	DIC	EAIC	EBIC	WAIC	Lm	LPML
Logito	205.9301	247.4760	254.0727	245.4157	83.62801	-122.7085
C-P	216.9481	249.3431	255.9397	248.3443	83.52640	-124.1728
C-RP	218.8798	249.3306	255.9272	248.3271	83.51953	-124.1642
G-RP	212.7507	249.0943	255.6909	248.1044	83.97274	-124.0654
GR-RP	208.2805	248.1019	254.6985	247.0171	83.54351	-123.5147
LA-P	212.6670	248.9947	255.5913	247.9484	83.49726	-123.9739
LA-RP	210.9074	249.0308	255.6275	248.0285	83.26746	-124.0157
L-P	214.1913	248.5341	255.1307	247.4799	83.46165	-123.7419
L-RP	214.1748	248.4794	255.0760	247.4021	83.52575	-123.7041
N-P	211.1455	248.4200	255.0166	247.3509	83.58346	-123.6795
N-RP	208.5697	248.5259	255.1226	247.4201	83.55728	-123.7131

Fonte: Elaborada pelo autor.

3.5 Aplicação

Nesta Seção, ilustramos a utilização dos modelos propostos em um conjunto de dados reais. Os dados originais são um compilado de mais de 30 anos de observação e consistem, para cada município do estado brasileiro da Bahia, na informação do ‘trimestre menos chuvoso’, resultando em 417 observações. Os dados foram extraídos e adaptados de [Alvares et al. \(2013\)](#), e disponibilizados na base de dados espaciais da Embrapa (<<https://geoinfo.dados.embrapa.br/>>).

A resposta observada original é categórica, ou seja, podendo ser uma de seis respostas possíveis: i) “AGOSTO-SETEMBRO-OUTUBRO”; ii) “DEZEMBRO-JANEIRO-FEVEREIRO”; iii) “JULHO-AGOSTO-SETEMBRO”; iv) “JUNHO-JULHO-AGOSTO”; v) “NOVEMBRO-DEZEMBRO-JANEIRO” e vi) “SETEMBRO-OUTUBRO-NOVEMBRO”. Transformamos a variável observada original em dicotômica, atribuindo valor 1 aos municípios cujo ‘trimestre menos chuvoso’ foi ‘AGOSTO-SETEMBRO-OUTUBRO’ e zero caso contrário, resultando em 119 observações do evento de interesse (proporção de 0.2853); ou seja, um desbalanceamento considerável. Veja na Figura 9 a distribuição da variável resposta por município. Na Tabela 9, apresentamos a descrição do conjunto de dados por inteiro, incluindo as covariáveis usadas na análise. Estas foram extraídas de outra fonte, a saber, do IBGE (Instituto Brasileiro de Geografia e Estatística).

As covariáveis foram padronizadas subtraindo a média dos valores da covariável e dividindo pelo desvio padrão dos valores, e o valor do parâmetro referente aos graus de liberdade κ , na distribuição a priori $G\text{-Wishart}_{\mathcal{W}}(\kappa, \mathcal{S})$, foi fixado em $\kappa = 417$ (quantidade de sub-regiões). Os resultados exibidos nesta Seção são baseados em 10.000 iterações após descartar as primeiras 5.000 como *burn-in*. A avaliação da convergência das cadeias de Markov foi baseada na inspeção visual de gráficos de traços e gráficos de autocorrelação.

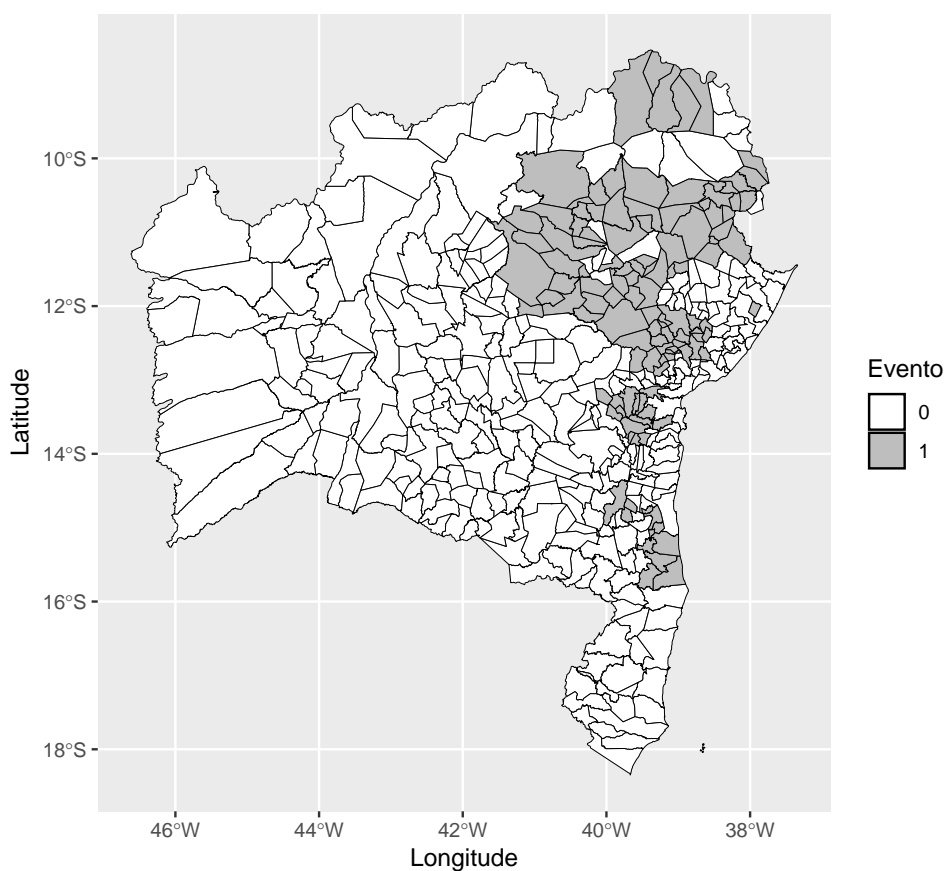
De início, usamos todas as covariáveis (X_1, X_2, X_3 e X_4) para fazer os ajustes de todos os modelos, porém percebemos que nem todas elas apresentaram significância estatística para a variável resposta que foi definida. Isso pode ser observado na Figura 10, na qual contém as estimativas pontuais e

Tabela 9 – Características do conjunto de dados ‘trimestre menos chuvoso’ para os municípios do estado da Bahia.

Variável	Notação	Descrição
CD_MUNI	–	código do município estabelecido pelo Instituto Brasileiro de Geografia e Estatística - IBGE
NM_MUNICIP	–	nome do município
latitude	X_1	latitude do município
longitude	X_2	longitude do município
altitude	X_3	altitude do município
semiárido	X_4	se o município pertence à região do semiárido brasileiro: 1 - sim; 0 - não

Fonte: Elaborada pelo autor.

Figura 9 – Distribuição da variável resposta para cada município do estado da Bahia.

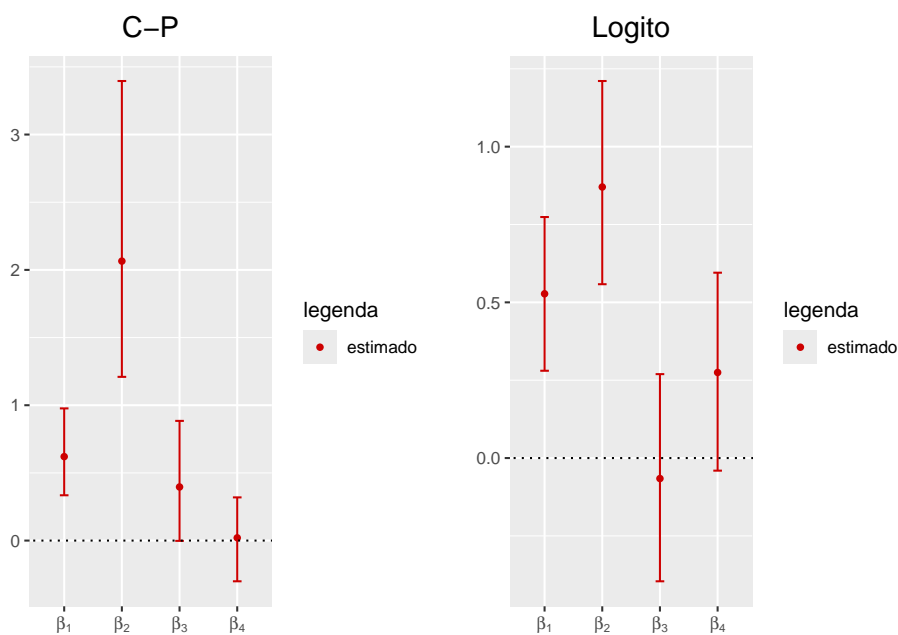


Fonte: Elaborada pelo autor.

intervalos de credibilidade para todas as covariáveis sob os modelos Logito e C-P. Percebe-se que o zero está contido nos intervalos de credibilidade dos coeficientes de duas covariáveis, X_3 e X_4 , para os dois modelos considerados. Como segundo passo, decidimos remover essas covariáveis e refazer os ajustes para avaliar os efeitos resultantes.

Após os novos ajustes, os resultados parecem apontar para a significância das covariáveis restantes, como pode ser visto na Tabela 10, que contém as estimativas pontuais e intervalos de credibilidade de 95% dos coeficientes de regressão das covariáveis usadas. O modelo C-RP apresentou problemas de convergência das cadeias de Markov, e portanto não será incluído no decorrer das análises. Pela inspeção do *trace plot* (gráfico de traço) e do gráfico de autocorrelação, parece haver indicação de que as cadeias atingiram a estacionariedade. Como ilustração, na Figura 11 apresentamos os gráficos de traços e autocorrelação dos modelos C-P e Logito; os outros modelos não foram incluídos por questões de espaço e estarão disponíveis no Apêndice D.

Figura 10 – Estimativas pontuais e intervalos de credibilidade de 95% a posteriori, usando todas as covariáveis. Modelos C-P e Logito.

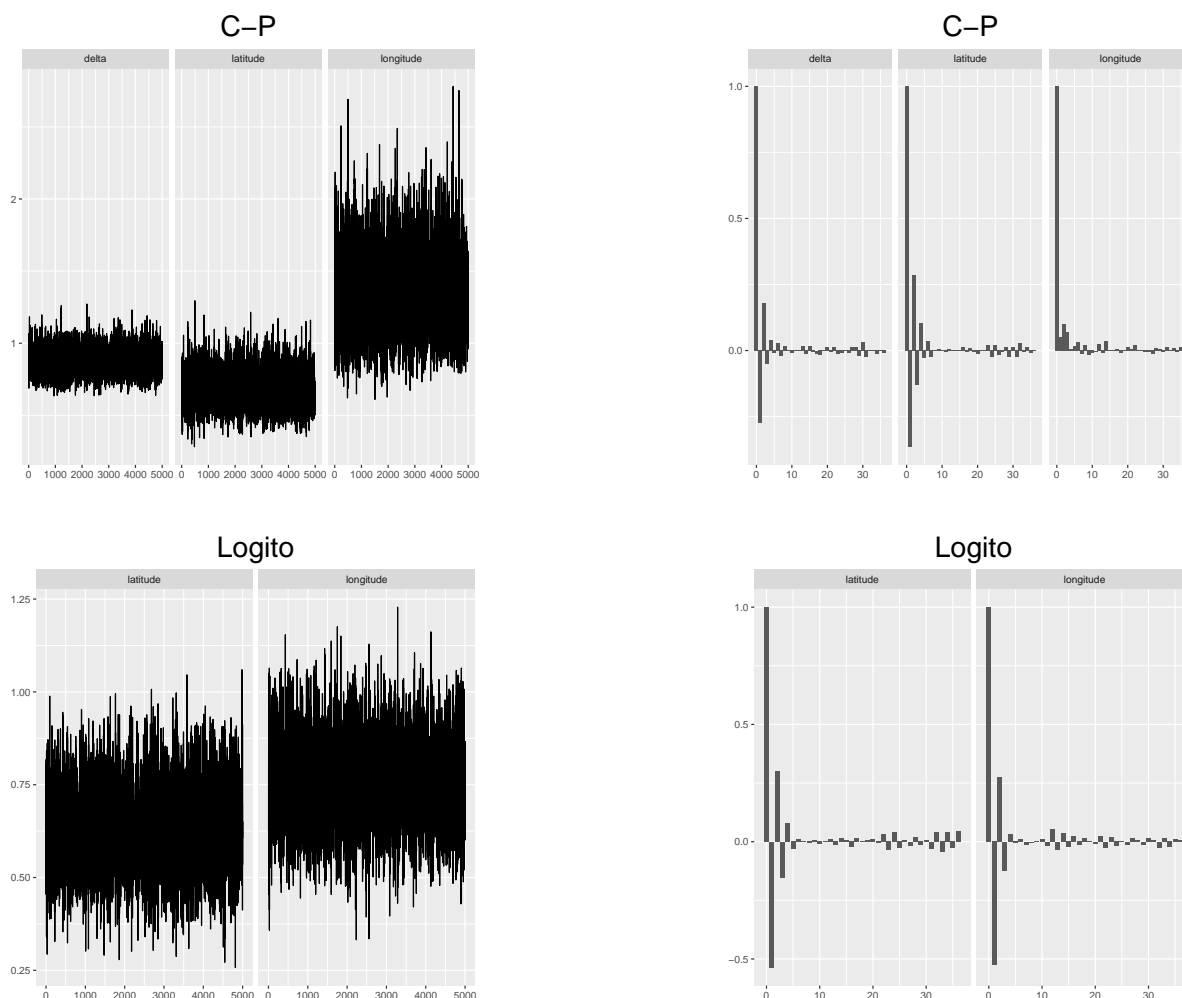


Fonte: Elaborada pelo autor.

Na Tabela 11, realizamos as comparações de modelos sob as diversas funções de ligações consideradas neste trabalho, baseadas na qualidade do ajuste. Além das métricas já vistas anteriormente, calculamos também o CPO (*conditional predictive ordinate*) para cada modelo. A partir da Tabela, constatamos que os modelos com função de ligação flexível possuem desempenho melhor quando comparados aos modelos Logito, Probit e Cloglog. Dos modelos com ligação assimétrica estudados, o modelo de ligação Gumbel Reversa Reversa de Potência (GR-RP) apresenta o melhor desempenho, ficando atrás em apenas uma das seis métricas consideradas.

Na Figura 12, utilizamos a divergência de Bregman para detecção de *outliers*. Usamos a perturbação

Figura 11 – Gráficos de traço e autocorrelação dos modelos C-P e Logito.



Fonte: Elaborada pelo autor.

de caso excluído dada em (2.36) para todas as funções de ligação apresentadas na Tabela 2. Em seguida, calculamos as probabilidades de ser um *outlier* (gráfico superior na Figura 12) conforme dado em 2.43 para todas as observações, com uma função convexa ψ_1 (divergência de Kullback-Leibler), e as divergências de Bregman normalizada (gráfico inferior na Figura 12) em 2.44. Os valores (as probabilidades de *outliers* e divergência de Bregman normalizada) estão sendo apresentados *versus* cada observação (município) do conjunto de dados.

No gráfico superior da Figura 12, percebe-se, pelo resultado, que as maiores probabilidades de *outliers* são fornecidas pelos modelos com função de ligação assimétrica. O mesmo comportamento também ocorre no gráfico inferior, para os valores da divergência de Bregman normalizada. Em relação às probabilidades de possíveis *outliers* (gráfico superior), não há uma única observação que tenha uma probabilidade maior que 0,30; a observação que apresentou a maior probabilidade foi verificada para o município de Heliópolis sob o modelo com ligação C-P, com um valor de 0.218. Para os modelos Logito, Probit e Cloglog, os comportamentos foram semelhantes, nenhuma observação com probabilidade de ser um *outlier* maior que 0.1, exceto para a ligação Probit que apresentou uma observação com um valor de

Tabela 10 – Estimativas dos parâmetros usando diferentes funções de ligação: C-P, C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, Probit, Logito e Cloglog. As estimativas são a média a posteriori, desvio padrão (em parênteses) e intervalo de credibilidade de 95%.

	C-P	N-P	N-RP	L-P	L-RP
Latitude	0.685 (0.126) (0.455 , 0.953)	0.364 (0.059) (0.248 , 0.481)	0.651 (0.114) (0.437 , 0.883)	0.609 (0.1) (0.417 , 0.814)	1.292 (0.268) (0.829 , 1.875)
Longitude	1.353 (0.274) (0.889 , 1.952)	0.551 (0.074) (0.408 , 0.701)	1.02 (0.178) (0.705 , 1.394)	0.952 (0.135) (0.7 , 1.229)	2.317 (0.574) (1.453 , 3.693)
δ	0.885 (0.089) (0.718 , 1.067)	0.791 (0.074) (0.646 , 0.939)	-1.086 (0.133) (-1.362 , -0.843)	0.802 (0.075) (0.661 , 0.952)	-1.206 (0.167) (-1.57 , -0.908)
	GR-P	GR-RP	Probit	Logito	Cloglog
Latitude	0.296 (0.049) (0.2 , 0.394)	0.666 (0.101) (0.479 , 0.871)	0.376 (0.067) (0.245 , 0.506)	0.631 (0.111) (0.425 , 0.862)	0.282 (0.049) (0.188 , 0.378)
Longitude	0.423 (0.055) (0.319 , 0.534)	1.096 (0.102) (0.885 , 1.26)	0.449 (0.068) (0.317 , 0.585)	0.754 (0.114) (0.537 , 0.99)	0.31 (0.047) (0.217 , 0.406)
δ	1.218 (0.071) (1.078 , 1.354)	-0.595 (0.097) (-0.781 , -0.417)	–	–	–

Fonte: Elaborada pelo autor.

probabilidade de 0,10.

Quando avaliamos as divergências de Bregman normalizadas (gráfico inferior), constatamos que os modelos C-P e LA-RP apresentam muitas observações com valores altos. Destes dois modelos, o LA-RP é o que apresenta o maior valor da divergência de Bregman normalizada, a saber, 0.076 para o município de Valente. Utilizando o critério de partida de [Danilevicz e Ehlers \(2022\)](#) (veja Seção 2.6.1) para classificar observações como possíveis *outliers*, inserimos na Figura 13 os municípios candidatos a observações influentes para pelo menos 8 dos 12 modelos ajustados.

Indiretamente, apesar de não serem métricas de comparação entre modelos, os resultados advindos da Figura 12 são interessantes porque, assumindo que modelos com assimetria flexível se ajustam melhor a dados com a presença de desbalanceamento, é um pensamento plausível esperar que eles sejam mais sensíveis a observações discrepantes.

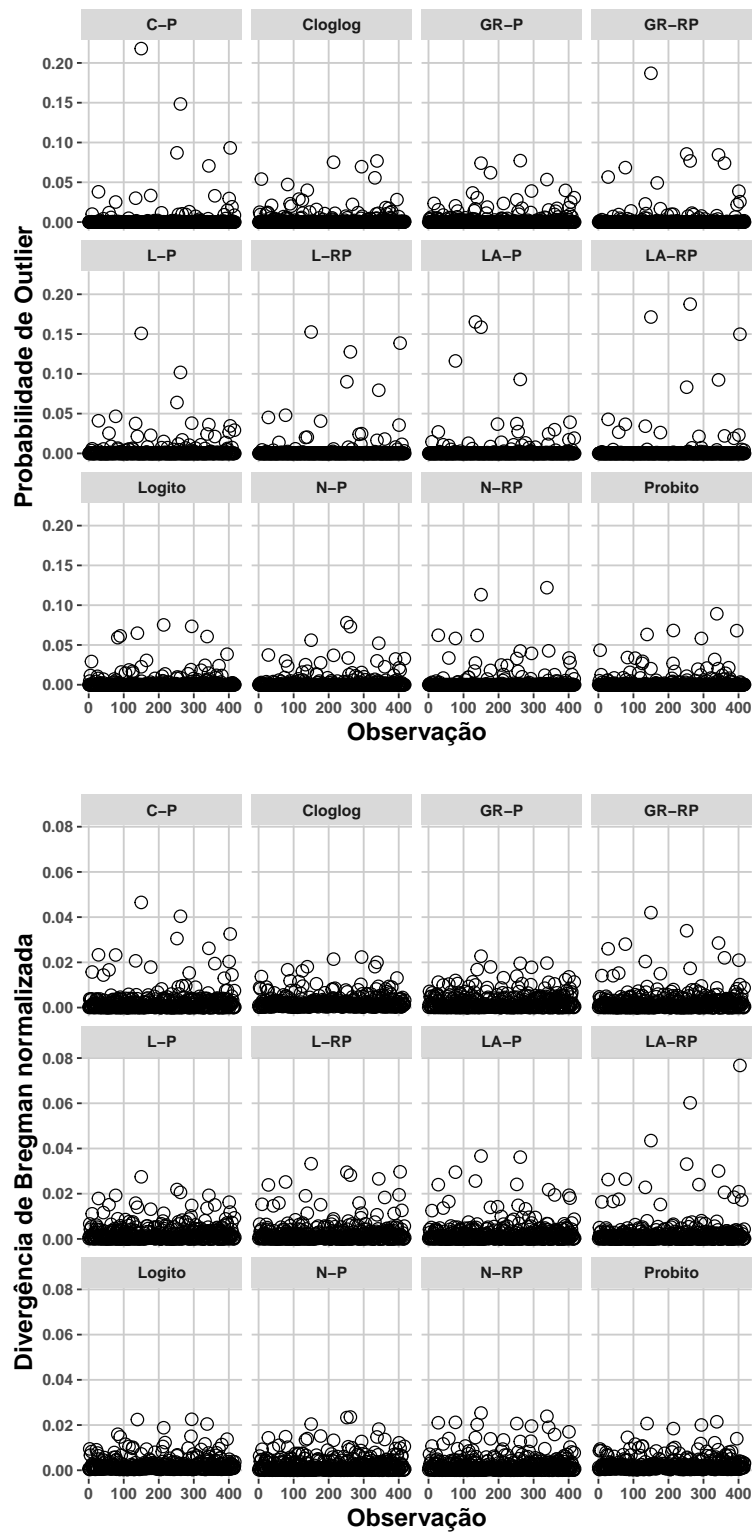
Na Tabela 12, comparamos o desempenho dos modelos considerando duas abordagens:

1. usando o amostrador exato para a distribuição condicional completa a posteriori de Ω (GW-E), a saber, de 4.5, $\Omega \sim \text{G-Wishart}_{\mathcal{W}}(\kappa + 1, \mathcal{S} + \phi\phi^{\top})$;
2. atribuindo priori paramétrica CAR para Ω .

Os resultados mostram que os modelos com priori CAR apresentam, na maioria das métricas avaliadas, o melhor desempenho. A única exceção é o modelo N-P, no qual os ajustes com priori G-Wishart alcançam desempenho semelhante ao CAR, mas não superior. Quanto ao tempo de execução, os modelos com priori CAR são os mais rápidos, enquanto os modelos com o esquema GW-E são os mais demorados.

Na Tabela 13, temos os resultados, para cada modelo, das métricas de predição apresentadas na Seção 2.7. Usamos as distribuições preditivas do nosso modelo espacial em 3.13 para estimar as

Figura 12 – Probabilidades a posteriori estimadas de serem *outliers* com função convexa ψ_1 e divergência de Bregman normalizada com função convexa ψ_1 , contra observações deletadas para os dados de ‘trimestre menos chuvoso’ para cada município baiano.



Fonte: Elaborada pelo autor.

Tabela 11 – Comparação de modelos usando diferentes funções de ligação: C-P, N-P, N-RP, L-P, L-RP, LA-P, LA-RP, Logito, Probit, Cloglog, GR-P e GR-RP.

	C-P	N-P	N-RP	L-P	L-RP
WAIC	382.8400	394.9800	391.8300	391.930	387.6100
DIC	382.8200	395.3100	392.2200	392.290	387.7400
EAIC	384.7700	394.9900	394.2000	394.160	390.6600
EBIC	396.8700	407.0900	406.3000	406.260	402.7600
CPO	-0.4600	-0.4700	-0.4700	-0.470	-0.4600
Lm	124.2562	129.1524	128.3681	128.827	125.6173
	LA-P	LA-RP	GR-P	GR-RP	Logito
WAIC	384.3600	380.8500	401.9800	378.5800	500.2600
DIC	384.7600	380.8500	401.9600	378.9300	500.3400
EAIC	385.1400	383.5100	400.3300	381.9000	501.4700
EBIC	397.2400	395.6100	412.4300	394.0000	509.5300
CPO	-0.4600	-0.4600	-0.4800	-0.4500	-0.6000
Lm	125.9833	123.7936	130.7879	126.0634	169.1756
	Probit	Cloglog			
WAIC	502.5800	650.4800			
DIC	502.6200	650.0100			
EAIC	501.9200	648.8400			
EBIC	509.9900	656.9100			
CPO	-0.6000	-0.7800			
Lm	169.5807	204.5569			

Fonte: Elaborada pelo autor.

probabilidades. Pelos resultados, vemos que os modelos potência e reversa de potência apresentam resultados similares, com destaque para os modelos LA-P e LA-RP, que apresentam resultados muito bons em boa parte das métricas consideradas. Dentre os modelos tradicionais (Logito, Probit e Cloglog), estes chegaram a apresentar desempenho tão bom quanto os demais em algumas métricas, a saber, CSI, FAITH, AUC e F1. De fato, conforme [Assunção, Izbicki e Prates \(2024\)](#), a calibração do ponto de corte melhorou o desempenho preditivo dos modelos tradicionais Logito, Probit e Cloglog diante de um cenário de dados desbalanceados. Também é interessante observar que os pontos de corte estimados pelos modelos potência e reversa de potência, exceto para o modelo L-RP, segundo o critério apresentado na Seção 2.7, ficaram entre 0.21 e 0.26, nas proximidades da proporção real que era de 0.28.

Na Tabela 14, comparamos os resultados das métricas de qualidade da predição sob os modelos espaciais G-Wishart usando o amostrador exato para a distribuição condicional completa a posteriori para Ω (GW-E), e atribuindo distribuição a priori CAR para Ω . Percebe-se, pela Tabela, que os ajustes realizados usando distribuição a priori CAR proporcionaram, para a maioria das métricas consideradas, os melhores valores. Nesse quesito, o modelo GR-P foi o que mais se destacou.

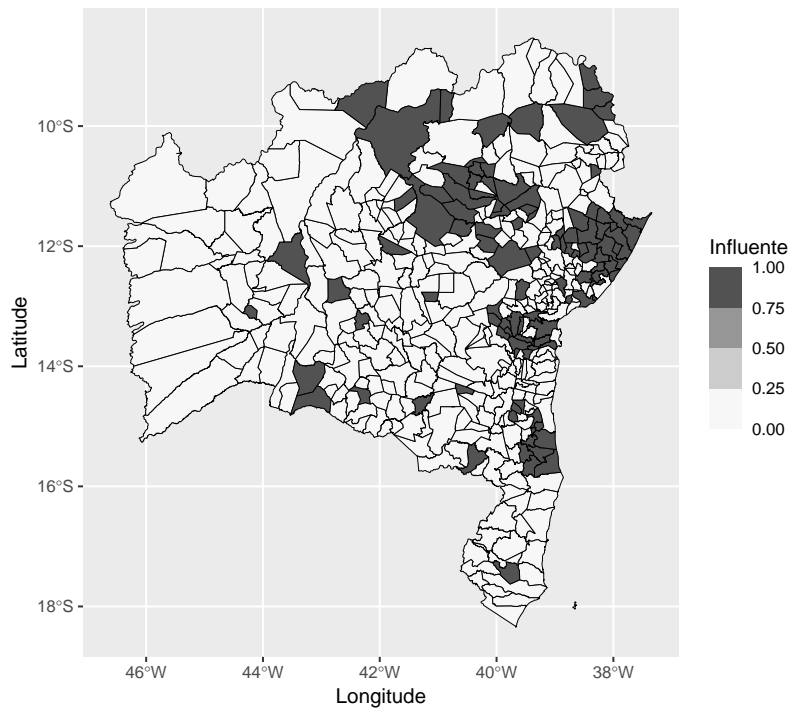
Finalmente, na Figura 14 apresentamos as probabilidades preditivas estimadas do evento de interesse. Decidimos apresentar os resultados do modelo GR-P, Logito, Probit e Cloglog, todos ajustados

Tabela 12 – Comparação de modelos usando diferentes funções de ligação: C-P, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato de [Mohammadi e Wit \(2015\)](#) (GW-E); ii) atribuindo priori CAR para a estrutura espacial (CAR).

Métrica	Método	C-P	N-P	N-RP	L-P	L-RP	LA-P	LA-RP	GR-P	GR-RP
WAIC	GW-E	382.8400	394.9800	391.8300	391.930	387.6100	384.3600	380.8500	401.9800	378.5800
	CAR	335.1300	394.9800	341.4500	328.1200	367.7700	304.14000	355.7500	202.12000	334.9400
DIC	GW-E	382.8200	395.3100	392.2200	392.290	387.7400	384.7600	380.8500	401.9600	378.9300
	CAR	338.2400	395.3100	344.4800	332.5900	368.1500	312.36000	358.9700	212.17000	339.9300
EAIC	GW-E	384.7700	394.9900	394.2000	394.160	390.6600	385.1400	383.5100	400.3300	381.9000
	CAR	324.5800	394.9900	329.2500	317.3000	366.9500	276.41000	351.6500	197.19000	323.0500
EBIC	GW-E	396.8700	407.0900	406.3000	406.260	402.7600	397.2400	395.6100	412.4300	394.0000
	CAR	336.6700	407.0900	341.3500	329.4000	379.0500	288.51000	363.7400	209.29000	335.1500
CPO	GW-E	-0.4600	-0.4700	-0.4700	-0.470	-0.4600	-0.4600	-0.4600	-0.4800	-0.4500
	CAR	-0.4000	-0.4700	-0.4100	-0.3900	-0.4400	-0.37000	-0.4300	-0.24000	-0.4000
Lm	GW-E	124.2562	129.1524	128.3681	128.827	125.6173	125.9833	123.7936	130.7879	126.0634
	CAR	110.0552	129.1524	112.5902	117.8311	121.6751	94.24025	118.0443	64.70567	112.0592
Tempo (em seg.)	GW-E	5999.028	5941.929	5872.868	6014.606	5987.995	5979.229	5984.164	6036.917	5992.514
	CAR	87.219	77.733	80.65	108.153	106.06	93.862	94.101	96.817	86.68

Fonte: Elaborada pelo autor.

Figura 13 – Divergência de Bregman normalizada com função convexa ψ_1 de possíveis municípios influentes para pelo menos 8 modelos.



Fonte: Elaborada pelo autor.

sob distribuição a priori G-Wishart (sob esquema de amostragem exata). No Apêndice D (Seção D.5), fornecemos gráficos de probabilidades preditivas para as demais funções de ligação assimétrica. Na Figura, fica nítido, pelos resultados, que as probabilidades estimadas do modelo GR-P são mais plausíveis com a realidade observada na Figura 9, revelando que, em termos de predição, um modelo mais flexível é mais preferível.

Tabela 13 – Comparação de modelos segundo métricas de qualidade da predição, usando diferentes funções de ligação: C-P, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P, LA-RP, Logito, Probit e Cloglog. P_cmax - estimativa do ponto de corte.

	C-P	N-P	N-RP	L-P	L-RP	LA-P	LA-RP
P _c max	0.210	0.230	0.220	0.230	0.170	0.260	0.140
ACC	0.734	0.739	0.739	0.739	0.719	0.758	0.717
SN	0.916	0.891	0.891	0.882	0.941	0.866	0.983
ES	0.661	0.678	0.678	0.681	0.631	0.715	0.611
PRC	0.519	0.525	0.525	0.525	0.505	0.548	0.502
FPR	0.339	0.322	0.322	0.319	0.369	0.285	0.389
CSI	0.495	0.493	0.493	0.491	0.489	0.505	0.498
SSI	0.329	0.327	0.327	0.325	0.324	0.338	0.331
FAITH	0.498	0.496	0.496	0.495	0.494	0.502	0.499
KAPPA	0.469	0.470	0.470	0.468	0.454	0.494	0.461
MCC	0.521	0.514	0.514	0.509	0.518	0.527	0.540
AUC	0.371	0.412	0.394	0.405	0.376	0.399	0.358
F1	0.663	0.660	0.660	0.658	0.657	0.671	0.665
	GR-P	GR-RP	Logito	Probit	Cloglog		
P _c max	0.220	0.210	0.500	0.500	0.630		
ACC	0.746	0.724	0.741	0.739	0.734		
SN	0.891	0.916	0.899	0.899	0.891		
ES	0.688	0.648	0.678	0.674	0.671		
PRC	0.533	0.509	0.527	0.525	0.520		
FPR	0.312	0.352	0.322	0.326	0.329		
CSI	0.500	0.487	0.498	0.495	0.488		
SSI	0.333	0.322	0.331	0.329	0.323		
FAITH	0.500	0.493	0.499	0.498	0.494		
KAPPA	0.481	0.455	0.476	0.472	0.463		
MCC	0.523	0.509	0.521	0.518	0.508		
AUC	0.433	0.357	0.425	0.435	0.437		
F1	0.667	0.655	0.665	0.663	0.656		

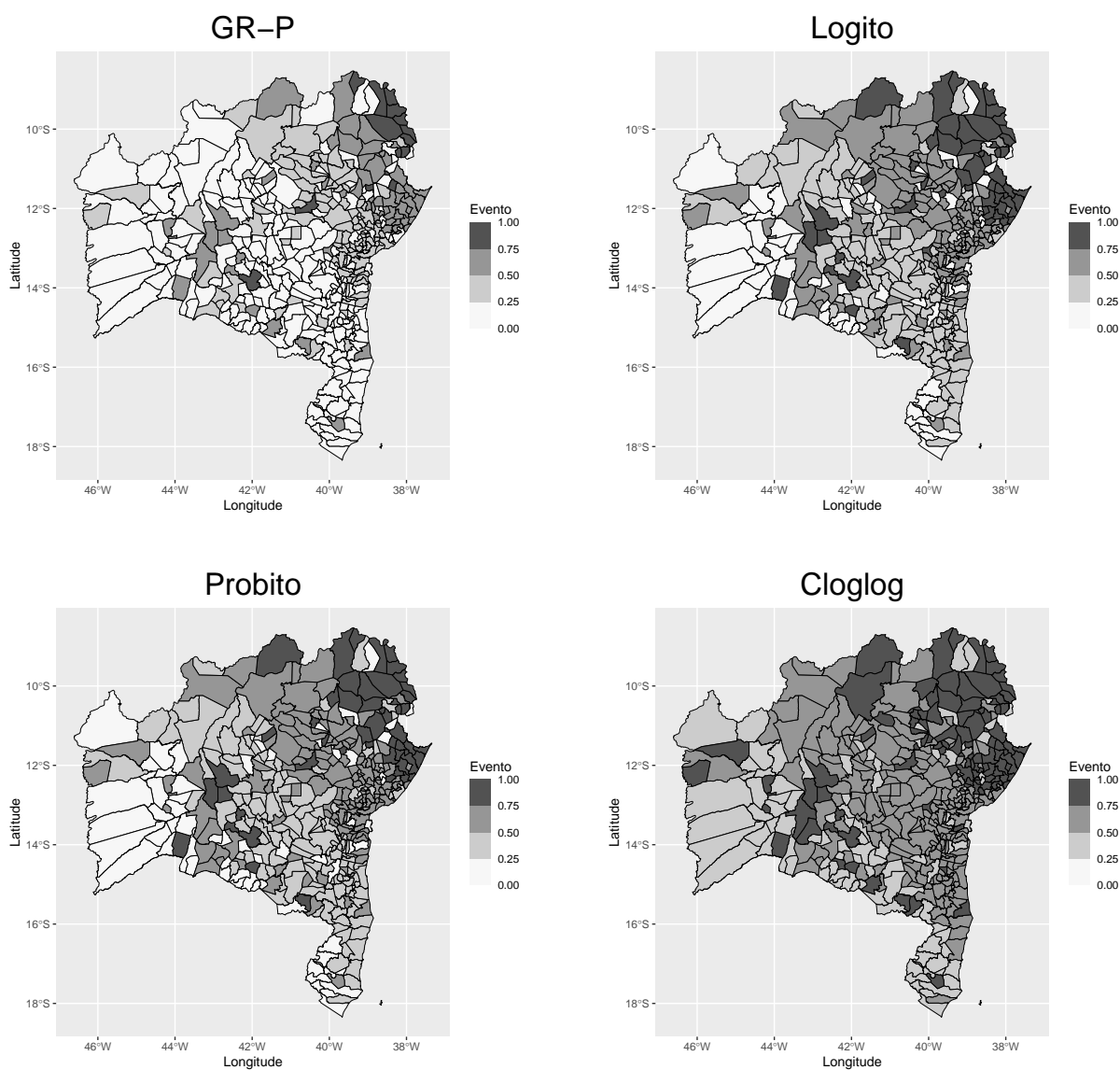
Fonte: Elaborada pelo autor.

Tabela 14 – Comparação de modelos segundo métricas de qualidade da predição, usando diferentes funções de ligação: C-P, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP; sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato de [Mohammadi e Wit \(2015\)](#) (GW-E); ii) atribuindo priori CAR para a estrutura espacial (CAR)

Métrica	C-P		N-P		N-RP		L-P		L-RP		LA-P		LA-RP		GR-P		GR-RP	
	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR
Pmax	0.210	0.240	0.230	0.370	0.220	0.280	0.230	0.390	0.170	0.210	0.260	0.410	0.140	0.250	0.220	0.430	0.210	0.380
ACC	0.734	0.801	0.739	0.918	0.739	0.803	0.739	0.835	0.719	0.746	0.758	0.911	0.717	0.784	0.746	0.928	0.724	0.825
SN	0.916	0.992	0.891	0.933	0.891	0.966	0.882	0.824	0.941	0.941	0.866	0.933	0.983	0.966	0.891	0.891	0.916	0.908
ES	0.661	0.725	0.678	0.913	0.678	0.738	0.681	0.839	0.631	0.668	0.715	0.903	0.611	0.711	0.688	0.943	0.648	0.792
PRC	0.519	0.590	0.525	0.810	0.525	0.596	0.525	0.671	0.505	0.531	0.548	0.793	0.538	0.502	0.572	0.533	0.862	0.509
FPR	0.339	0.275	0.322	0.087	0.322	0.262	0.319	0.161	0.369	0.332	0.285	0.097	0.389	0.289	0.312	0.057	0.352	0.208
CSI	0.495	0.587	0.493	0.766	0.493	0.584	0.491	0.587	0.489	0.514	0.505	0.750	0.498	0.561	0.500	0.779	0.487	0.597
SSI	0.329	0.415	0.327	0.620	0.327	0.412	0.325	0.415	0.324	0.346	0.338	0.600	0.335	0.331	0.390	0.333	0.639	0.322
FAITH	0.498	0.542	0.496	0.592	0.496	0.540	0.495	0.535	0.494	0.507	0.502	0.589	0.499	0.530	0.500	0.591	0.493	0.542
KAPPA	0.469	0.595	0.470	0.809	0.470	0.594	0.468	0.620	0.454	0.494	0.494	0.793	0.461	0.562	0.481	0.825	0.455	0.620
MCC	0.521	0.648	0.514	0.813	0.514	0.638	0.509	0.627	0.518	0.550	0.527	0.799	0.540	0.613	0.523	0.826	0.509	0.643
AUC	0.371	0.419	0.412	0.361	0.394	0.440	0.405	0.416	0.376	0.399	0.399	0.389	0.358	0.409	0.433	0.336	0.357	0.424
F1	0.663	0.740	0.660	0.867	0.660	0.737	0.658	0.740	0.657	0.679								

Fonte: Elaborada pelo autor.

Figura 14 – Probabilidades preditivas do evento de interesse para cada município da Bahia sob os modelos GR-P, Probit, Logito e Cloglog.



Fonte: Elaborada pelo autor.

3.6 Discussões Finais

Neste Capítulo, apresentamos a estrutura do nosso primeiro modelo de regressão binária espacial com função de ligação flexível. Esta possui um parâmetro de forma para lidar com o cenário de desbalanceamento dos dados espaciais. Desenvolvemos métodos de amostragem para os parâmetros do modelo que fossem eficientes em termos computacionais, para garantir escalabilidade em problemas com muitos dados.

Através do nosso estudo de simulação, mostramos que o modelo proposto, sob as funções de ligação flexíveis adotadas, e os métodos de amostragem das distribuições a posteriori, apresentou bom desempenho em termos de recuperação do verdadeiro valor dos parâmetros nas diversas combinações de tamanho de amostra e valores de λ considerados. Também verificou-se bons resultados deste modelo (sob todas as funções de ligação flexíveis) em um cenário de má especificação, tanto em termos de critérios de bondade de ajuste quanto em termos de critérios de qualidade da predição.

No tocante à parte da aplicação aos dados reais, alguns resultados podem ser destacados. Os modelos com funções de ligação flexíveis apresentaram resultados melhores em termos de qualidade de ajuste quando comparados aos modelos com funções de ligação Logito, Probit e Cloglog. Na parte do diagnóstico de *outliers*, usando a divergência de Bregman, os modelos flexíveis foram os que identificaram um maior número de possíveis *outliers*. Como comentado antes, acreditamos que isso advenha do fato de que estes modelos com assimetria flexível se ajustam melhor a dados com presença de desbalanceamento e, conseqüentemente, são mais sensíveis a observações discrepantes.

Também comparamos os modelos com funções de ligação flexíveis ajustados sob distribuição a priori G-Wishart e distribuição a priori CAR. Os resultados em termos de qualidade de ajuste mostraram que os ajustes realizados sob distribuição a priori CAR apresentam melhor desempenho; e quando avaliamos sob critérios de qualidade da predição, os resultados apontam que os ajustes com distribuição a priori CAR são melhores para a maioria das métricas consideradas. Com isso, percebe-se a robustez do modelo CAR em termos de aplicação prática, apesar de ser mais antigo que o modelo G-Wishart. No Apêndice D (Seção D.6), fornecemos um estudo de simulação sob distribuição a priori CAR, considerando os mesmos cenários apresentados na Seção 3.4, e os resultados apontam que os modelos desempenham tão bem quanto sob especificação G-Wishart, em termos de recuperação dos verdadeiros valores dos parâmetros.

MODELOS DE REGRESSÃO BINÁRIA ESPACIAL: UMA GENERALIZAÇÃO

Neste Capítulo, apresentamos um modelo que pode ser visto tanto como um caso especial, quanto como uma generalização do modelo espacial apresentado no Capítulo 3. No capítulo anterior, o foco era modelar dados espaciais provenientes de uma única região, composta por m sub-regiões. Agora, estendemos essa abordagem para considerar r regiões, assumidas exatamente iguais, todas com a mesma estrutura de adjacência e contendo m sub-regiões cada. Assim, o modelo anterior é recuperado como um caso especial quando $r = 1$.

A proposta deste modelo generalizado foi motivada por um conjunto de dados que será analisado posteriormente (Seção 4.4). Adiante, iniciamos este Capítulo descrevendo os dados motivacionais do modelo (Seção 4.1); logo em seguida apresentamos a forma do nosso modelo generalizado (Seção 4.2), e um estudo de simulação na Seção 4.3. A aplicação aos dados motivacionais será apresentada na Seção 4.4, e por fim, apresentamos algumas conclusões e discutimos sobre trabalhos futuros (Seção 4.5).

4.1 Dados Motivacionais

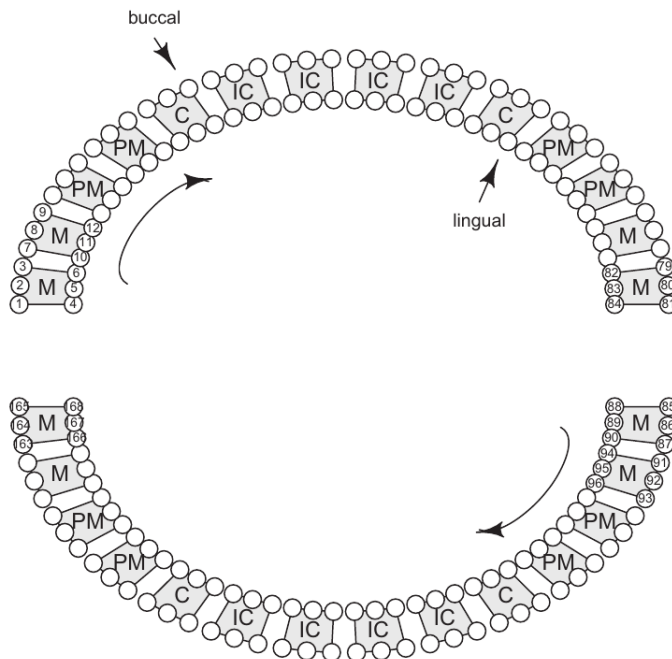
A doença periodontal (DP) é uma doença crônica comumente conhecida e uma das principais causas de perda dentária em humanos (BANDYOPADHYAY; CANALE, 2016). Segundo Negrato *et al.* (2013), a doença periodontal é uma inflamação bacteriana que afeta tanto a gengiva quanto o osso que sustenta os dentes, e é causada por microrganismos anaeróbicos que estão presentes na placa bacteriana que adere aos dentes.

Segundo os mesmos autores, pode-se considerar que a DP tem certa prevalência, o que pode ser demonstrado por meio de algumas estatísticas. Por exemplo, eles mencionam em seu trabalho que mais da metade da população dos EUA com mais de 18 anos tem DP em algum estágio, e quando a idade considerada é de 35 anos ou mais, esse percentual aumenta para 75%, podendo este número ser muito maior em países pobres e subdesenvolvidos. Hugoson *et al.* (1989), Brown, Oliver e Loe (1990), na mesma linha, relatam que cerca de 10-15% da população em países ocidentais desenvolve periodontite grave.

As consequências decorrentes da DP, quando não tratadas, levam à perda óssea progressiva ao redor do dente, levando ao afrouxamento dos maxilares (superior) ou mandibular (inferior), com eventual perda [Bandyopadhyay e Canale \(2016\)](#).

Para avaliar a progressão da periodontite em um indivíduo, [Bandyopadhyay e Canale \(2016\)](#) explicam brevemente o procedimento mais comum realizado por higienistas dentais, e resumiremos esse procedimento a seguir. Basicamente, uma sonda periodontal calibrada é usada para medir o nível de inserção clínica (CAL - *clinical attachment level*), que é o biomarcador mais popular da DP. O CAL é definido como a distância até a raiz de um dente que não está mais preso ao osso circundante pelo ligamento periodontal. O CAL é medido em seis localizações dentárias pré-especificadas para cada um dos 28 dentes de um indivíduo, excluindo os quatro terceiros molares (dentes do siso), resultando em 168 localizações diferentes para cada indivíduo, assumindo que não há perda dentária. A sonda periodontal normalmente mede a perda de inserção em milímetros, arredondando para o milímetro mais próximo, ou seja, os valores resultantes são valores inteiros. A Figura 15 representa os diferentes tipos de dentes (molares, pré-molares, caninos e incisivos), as localizações (bucal versus lingual) e a numeração das medições nos dentes (como 1–6 e 7–12), de um sujeito hipotético sem dentes ausentes. As medições 1–84 estão localizadas no maxilar superior, enquanto as medições 85–168 estão na mandíbula inferior. “Lingual” refere-se às sub-regiões que estão em contato direto com a língua, enquanto as opostas são as sub-regiões “bucais”.

Figura 15 – Tipos de dentes (M, molar; PM, pré-molar; C, canino; IC, incisivo), localizações (bucal, ou lado da bochecha, versus lingual, ou lado da língua) e numeração das medições (como 1–6 para o segundo molar no quadrante superior esquerdo), para um sujeito hipotético sem dentes ausentes.



Extraído de [Bandyopadhyay e Canale \(2016\)](#).

Os dados motivacionais para este trabalho são resultados de um estudo clínico conduzido na *Medical University of South Carolina* para determinar o status de DP. Esses dados também foram usados

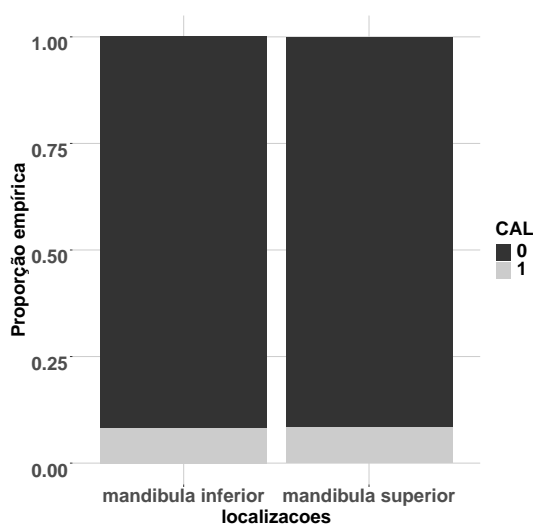
por [Bandyopadhyay e Canale \(2016\)](#), mas de uma perspectiva diferente; e, como argumentado no trabalho acima mencionado, assumimos a hipótese de que a progressão do CAL pode ser referenciada espacialmente, ou seja, localizações dentárias próximas compartilham status de CAL semelhante em comparação com localizações distantes.

Em relação ao banco de dados à nossa disposição, temos informações de medidas de CAL para 168 locais de 288 indivíduos, e o evento de interesse sendo em cada local $Y = 1$ se $CAL > 5$ mm ou $Y = 0$ caso contrário. E vinculados a cada indivíduo, temos informações sobre 5 covariáveis, a saber:

- X_1 – idade: idade em anos do paciente;
- X_2 – sexo: 1 - feminino, 0 - masculino;
- X_3 – IMC: índice de massa corporal (em kg/m^2);
- X_4 – fumante: 1 - se fuma, 0 - caso contrário;
- X_5 – HbA1c: marcador de diabetes: 1 - alto, 0 - controlado.

Para esse banco de dados, podemos citar também que a proporção do evento de interesse, ou seja, a perda de inserção clínica (CAL) observada no paciente, é muito diferente do evento complementar, isto é, a não presença de CAL. Para visualizar isso, na Figura 16 temos a proporção constante de CAL nas mandíbulas superior e inferior para os 288 indivíduos.

Figura 16 – Proporção empírica de CAL em relação às mandíbulas superior e inferior para os 288 indivíduos.



Fonte: Elaborada pelo autor.

4.2 O Modelo Hierárquico Espacial

Considere agora que y_{si} seja uma resposta binária observada para a unidade observacional s (em nosso caso, serão os pacientes), para $s = 1, \dots, r$, em alguma sub-região i , para $i = 1, \dots, m$,

dentro da unidade observacional s , com probabilidade p_{si} para $Y_{si} = 1$, $1 - p_{si}$ para $Y_{si} = 0$. Para cada unidade observacional, temos um vetor de covariáveis \mathbf{x}_s , e $\boldsymbol{\beta}$, um vetor de coeficientes de regressão $k \times 1$. Especificamos o modelo espacial generalizado da seguinte forma:

$$\begin{aligned} Y_{si} | \phi_{si} &\stackrel{\text{ind}}{\sim} \text{Bernoulli}(p_{si}), \quad s = 1, \dots, r, \text{ e } i = 1, \dots, m, \\ p_{si} &= F_{\delta}(\mathbf{x}_i^{\top} \boldsymbol{\beta} + \phi_{si}). \end{aligned} \quad (4.1)$$

Agora \mathbf{y} é uma matriz $r \times m$ de respostas observadas e $\boldsymbol{\phi}_s$ é o vetor de efeitos aleatórios espaciais para cada unidade observacional $s = 1, \dots, r$. Aqui, como já mencionado na Seção 3.2, estaremos usando distribuição G-Wishart $\mathcal{W}(\kappa, \mathbf{S})$ como priori para $\boldsymbol{\Omega}$. Dessa forma, assumindo as mesmas distribuições a priori independentes para $\boldsymbol{\beta}$ e δ apresentadas no Capítulo 3, e assumindo a distribuição de $\boldsymbol{\phi}_s$ como na especificação 3.5, derivada da suposição de campo aleatório de Markov Gaussiano, temos a seguinte distribuição a posteriori conjunta:

$$p(\boldsymbol{\beta}, \delta, \boldsymbol{\phi}, \boldsymbol{\Omega} | \mathbf{y}, \mathbf{X}) \propto \prod_{s=1}^r p(\mathbf{y}_s | \mathbf{x}_s, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_s) p(\boldsymbol{\phi}_s | \boldsymbol{\Omega}) p(\boldsymbol{\Omega} | \mathbf{W}) p(\boldsymbol{\beta}) p(\delta). \quad (4.2)$$

Obtendo a proporcional das distribuições a posteriori condicionais completas, chegamos ao seguinte resultado para a proporcional da distribuição condicional completa conjunta de $(\boldsymbol{\beta}, \delta)$:

$$p(\boldsymbol{\beta}, \delta | \boldsymbol{\phi}, \boldsymbol{\Omega}, \mathbf{y}, \mathbf{X}) \propto \prod_{s=1}^r \prod_{i=1}^m [F_{\delta}(\mathbf{x}_i^{\top} \boldsymbol{\beta} + \phi_{si})]^{y_{si}} \prod_{s=1}^r \prod_{i=1}^m [1 - F_{\delta}(\mathbf{x}_i^{\top} \boldsymbol{\beta} + \phi_{si})]^{1-y_{si}} \prod_{j=1}^k \exp \left\{ -\frac{\beta_j^2}{2 \times 10^2} \right\} \frac{1}{4}. \quad (4.3)$$

Para amostrar valores da distribuição condicional completa conjunta de $(\boldsymbol{\beta}, \delta)$, usaremos o mesmo método que adotamos para a distribuição a posteriori 3.8, o Monte Carlo Hamiltoniano com a métrica de Riemann. As derivadas parciais requeridas e matrizes de informação de Fisher necessárias podem ser consultadas no Apêndice B.

A distribuição posterior condicional completa de $\boldsymbol{\Omega}$ é similar à apresentada em 4.5, porém com leves alterações. Note primeiro que agora

$$\begin{aligned} \prod_{s=1}^r p(\boldsymbol{\phi}_s | \boldsymbol{\Omega}) &\propto |\boldsymbol{\Omega}|^{r/2} \exp \left\{ -\frac{1}{2} \sum_{s=1}^r \boldsymbol{\phi}_s^{\top} \boldsymbol{\Omega} \boldsymbol{\phi}_s \right\} \propto |\boldsymbol{\Omega}|^{r/2} \exp \left\{ -\frac{1}{2} \sum_{s=1}^r \text{tr}(\boldsymbol{\phi}_s^{\top} \boldsymbol{\Omega} \boldsymbol{\phi}_s) \right\} \\ &\propto |\boldsymbol{\Omega}|^{r/2} \exp \left\{ -\frac{1}{2} \sum_{s=1}^r \text{tr}(\boldsymbol{\phi}_s \boldsymbol{\phi}_s^{\top} \boldsymbol{\Omega}) \right\} \propto |\boldsymbol{\Omega}|^{r/2} \exp \left\{ -\frac{1}{2} \sum_{s=1}^r \text{tr}(\boldsymbol{\phi}_s \boldsymbol{\phi}_s^{\top} \boldsymbol{\Omega}) \right\} \\ &\propto |\boldsymbol{\Omega}|^{r/2} \exp \left\{ -\frac{1}{2} \text{tr} \left(\sum_{s=1}^r \boldsymbol{\phi}_s \boldsymbol{\phi}_s^{\top} \boldsymbol{\Omega} \right) \right\}, \end{aligned} \quad (4.4)$$

e pelo fato de $\boldsymbol{\phi}_s | \boldsymbol{\Omega} \sim N_m(\mathbf{0}, \boldsymbol{\Omega})$, segue que

$$p(\boldsymbol{\Omega} | \boldsymbol{\beta}, \delta, \boldsymbol{\phi}, \mathbf{y}, \mathbf{X}) \propto |\boldsymbol{\Omega}|^{(r+\kappa-2)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\left(\mathbf{S} + \sum_{s=1}^r \boldsymbol{\phi}_s \boldsymbol{\phi}_s^{\top} \right) \boldsymbol{\Omega} \right] \right\}. \quad (4.5)$$

Assim, a distribuição a posteriori de $\boldsymbol{\Omega}$ é dada por $\boldsymbol{\Omega} | \boldsymbol{\beta}, \delta, \boldsymbol{\phi}, \mathbf{y}, \mathbf{X} \sim \text{G-Wishart}_{\mathcal{W}}(\kappa + r, \mathbf{S} + \sum_{s=1}^r \boldsymbol{\phi}_s \boldsymbol{\phi}_s^{\top})$, com a matriz de escala $\mathbf{S} + \sum_{s=1}^r \boldsymbol{\phi}_s \boldsymbol{\phi}_s^{\top}$ restrita a ter entradas nulas fora da diagonal, correspondendo a zeros na matriz de adjacência \mathbf{W} , que é única para cada uma das r regiões, uma vez que assumimos que

todas elas são iguais. Como mencionado no Capítulo 3, aqui também usaremos o esquema de amostragem exata proposto por Mohammadi e Wit (2015) para amostrar valores de Ω .

A distribuição a posteriori condicional completa do vetor de efeitos aleatórios espaciais ϕ_s , $s = 1, \dots, r$, é dada por

$$\begin{aligned} p(\phi_s | \beta, \delta, \Omega, \mathbf{y}_s, \mathbf{x}_s) &\propto p(\mathbf{y}_s | \mathbf{x}_s, \beta, \delta, \phi_s) p(\phi_s | \Omega) \\ &\propto \prod_{i=1}^m [F_\delta(\mathbf{x}_s^\top \beta + \phi_{si})]^{y_{si}} \prod_{i=1}^m [1 - F_\delta(\mathbf{x}_s^\top \beta + \phi_{si})]^{1-y_{si}} \exp \left\{ -\frac{1}{2} \phi_s^\top \Omega \phi_s \right\}. \end{aligned} \quad (4.6)$$

Usaremos o método HMC padrão para amostrar valores da distribuição a posteriori dos efeitos aleatórios espaciais.

Finalizamos esta Seção apresentando a distribuição preditiva do modelo proposto, e em seguida, a respectiva expressão da medida L resultante, apresentada anteriormente em 2.33. Dito isso, de posse de uma amostra da distribuição a posteriori de $\{\beta^{(l)}, \delta^{(l)}, \phi^{(l)}\}_{l=1}^N$, a expressão da média preditiva a posteriori para o modelo generalizado fica

$$\tilde{\mu}_{si} = \frac{1}{N} \sum_{l=1}^N [F_\delta((\mathbf{x}_s^\top \beta^{(l)} + \phi_{si}^{(l)}); \delta^{(l)})], \quad (4.7)$$

e a variância preditiva a posteriori $Var(\tilde{y}_i | \mathbf{y}, \mathbf{X})$ pode ser aproximada como $\tilde{\mu}_{si}(1 - \tilde{\mu}_{si})$. De posse dessas informações, assumindo que $\nu = 1$, uma estimativa de 2.33 para o modelo de regressão espacial generalizado é dada por

$$L(\tilde{\mathbf{y}}, \mathbf{a}; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^m \sum_{s=1}^r \tilde{\mu}_{si}(1 - \tilde{\mu}_{si}) + \sum_{i=1}^m \sum_{s=1}^r (\tilde{\mu}_{si} - y_{si})^2. \quad (4.8)$$

4.3 Estudo de Simulação

Nesta Seção, propomos um estudo de simulação para avaliar o desempenho do modelo apresentado em 4.1 sob as diversas funções de ligação flexíveis adotadas nesta tese. Assim como na Seção 3.4, construímos esse estudo em duas partes. A primeira consiste em avaliar a qualidade da recuperação das estimativas dos parâmetros do modelo usando as métricas apresentadas antes (veja a Seção 3.4), e na segunda parte, avaliamos o desempenho dos modelos em um cenário de má especificação.

Para o primeiro contexto, novamente, geramos $M = 100$ replicações de conjuntos de dados binários multivariados com tamanhos de amostra de $r = 50$ e $r = 200$, e $m = 70$ unidades de área. Os valores de duas covariáveis, x_1 e x_2 , foram gerados independentemente de uma distribuição normal padrão, enquanto os coeficientes das covariáveis foram fixados em $\beta_1 = -0,7$ e $\beta_2 = 0,7$. Uma matriz de respostas binárias y_{si} foi então gerada de acordo com as probabilidades $p_{si} = F_\delta(\beta_1 x_{s1} + \beta_2 x_{s2} + \phi_{si})$ para cada um dos modelos na Tabela 2, em que $p_{si} = P(y_{si} = 1)$, $i = 1, \dots, m$ e $s = 1, \dots, r$. O parâmetro de forma, como antes, foi fixado em $\lambda = 0,5$ e $\lambda = 4$ ($\delta = -0.693$ e $\delta = 1.386$). Finalmente, os efeitos aleatórios espaciais foram gerados de acordo com uma distribuição Normal multivariada com matriz de precisão G-Wishart esparsa, ou seja, $\phi_s \sim N_m(\mathbf{0}, \Omega^{-1})$, $s = 1, \dots, r$, com $\Omega \sim \text{G-Wishart}_{\mathcal{W}}(\kappa = m, \mathcal{S})$. Para cada replicação, extraímos 5000 amostras MCMC descartando as primeiras 2500 como *burn-in*.

Na segunda parte do estudo de simulação, geramos, novamente, dados de um modelo de regressão binária Logito, e buscamos avaliar a robustez dos modelos e algoritmos propostos sob esse cenário de má especificação. Para esse cenário, geramos $M = 100$ réplicas de conjuntos de dados binários multivariados com tamanhos de amostra $n = 100$, duas covariáveis x_1 e x_2 geradas independentemente de uma distribuição normal padrão e coeficientes $\beta_1 = -0,7$ e $\beta_2 = 0,7$. Efeitos aleatórios espaciais também foram gerados como na Seção 3.4. 5000 amostras MCMC foram coletadas, descartando as primeiras 2500 como *burn-in* para cada réplica e tamanho de amostra. Uma submatriz de dimensão $m \times m$ foi extraída da matriz de adjacência dos dados motivacionais (Seção 4.1) para gerar a estrutura de vizinhança das sub-regiões.

Como antes, foram feitas simulações iniciais com o objetivo de otimizar os parâmetros pertinentes ao HMC, baseados nos critérios da taxa de aceitação a novas propostas pelo algoritmo, e em avaliar o gráfico de convergência das cadeias dos parâmetros a serem estimados. Para o bloco de parâmetros (β, δ) , sob todas as funções de ligação avaliadas, configuramos o valor do tamanho do passo para $\epsilon = 0.1$ e quantidade de passos para $L = 22$; para o vetor de efeitos aleatórios espaciais ϕ_s , para $s = 1, \dots, r$, fizemos $\epsilon = 0.001$ e $L = 22$.

Os resultados da primeira parte das simulações são exibidos nas Tabelas 15 e 16 . As Tabelas apontam, para todos os modelos, bons resultados em termos de viés relativo, RMSE, NRMSE e probabilidades de cobertura (CP) para todas as configurações de λ e tamanho de amostras. Para a configuração $r = 200$ e $\lambda = 0,5$ e 4, pode-se observar probabilidades de cobertura ligeiramente menores para os modelos GR-RP, LA-P e LA-RP quando comparados aos outros modelos. Em geral, os modelos tiveram bom desempenho em todas as configurações testadas.

Analisando o tempo computacional, constatamos o que tínhamos observado antes no estudo de simulação do Capítulo 3. Para um tamanho amostral menor, o tempo de execução dos modelos em cada réplica é menor quando comparado a um tamanho amostral maior. Para tamanhos amostrais menores, o tempo de execução de cada modelo em cada réplica é mais estável, ao passo que para amostras maiores, as diferenças de tempo em cada réplica são maiores para todos os modelos.

A Tabela 17 contém os resultados das análises de robustez dos modelos. A partir dos resultados, notamos que os modelos corresponderam muito bem ao cenário atípico, apresentando métricas de qualidade de ajuste muito próximas às apresentadas pelo modelo que gerou os dados.

Por fim, percebe-se que todas as características anteriormente observadas no modelo apresentado no Capítulo 3 — como o bom desempenho dos métodos de amostragem, dos algoritmos propostos e a flexibilidade a cenários atípicos — estendem-se, de forma natural, a este modelo mais geral.

Tabela 15 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostra $r = 50$. 5000 iterações, descartando 2500 como *burn-in*.

Ligação	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	Ligação	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	
C-P	$\lambda = 0,5$	β_1	0.015	0.065	-0.092	0.93	C-RP	$\lambda = 0,5$	β_1	0.020	0.067	-0.094	0.95	
		β_2	0.016	0.062	0.088	0.95			β_2	0.024	0.066	0.092	0.94	293.71 (6.62)
		δ	0.004	0.04	-0.058	0.94			δ	0	0.034	-0.048	0.94	
	$\lambda = 4$	β_1	0.017	0.055	-0.077	0.93		β_1	0.020	0.059	-0.083	0.88		
		β_2	0.016	0.048	0.068	0.96		β_2	0.009	0.048	0.068	0.97	295.649 (8.55)	
		δ	0.005	0.042	0.03	0.95		δ	0.004	0.038	0.027	0.92		
GR-P	$\lambda = 0,5$	β_1	-0.004	0.039	-0.056	0.96	GR-RP	$\lambda = 0,5$	β_1	-0.005	0.040	-0.058	0.94	
		β_2	0.005	0.0506	0.071	0.94			β_2	0.001	0.047	0.067	0.91	355.721 (9.80)
		δ	0.004	0.042	-0.06	0.94			δ	-0.004	0.046	-0.066	0.91	
	$\lambda = 4$	β_1	0.003	0.024	-0.035	0.95		β_1	0.002	0.022	-0.032	0.96		
		β_2	-0.0001	0.026	0.037	0.92		β_2	0.001	0.025	0.036	0.93	366.338 (15.35)	
		δ	0	0.03	0.022	0.96		δ	0.006	0.031	0.022	0.94		
L-P	$\lambda = 0,5$	β_1	0.009	0.055	-0.078	0.92	L-RP	$\lambda = 0,5$	β_1	0.011	0.061	-0.086	0.94	
		β_2	-0.001	0.055	0.079	0.94			β_2	-0.001	0.055	0.079	0.91	385.259 (8.57)
		δ	0.003	0.034	-0.049	0.95			δ	-0.001	0.037	-0.053	0.89	
	$\lambda = 4$	β_1	0.0001	0.045	-0.064	0.94		β_1	0.0009	0.040	-0.058	0.97		
		β_2	0.005	0.040	0.057	0.97		β_2	0.005	0.040	0.057	0.98	400.481 (14.82)	
		δ	0.002	0.033	0.024	0.93		δ	0.002	0.035	0.025	0.89		
N-P	$\lambda = 0,5$	β_1	-0.003	0.035	-0.050	0.94	N-RP	$\lambda = 0,5$	β_1	0.0002	0.037	-0.052	0.94	
		β_2	0.003	0.037	0.052	0.92			β_2	-0.007	0.036	0.052	0.95	275.464 (7.69)
		δ	0.003	0.037	-0.053	0.92			δ	-0.001	0.036	-0.052	0.94	
	$\lambda = 4$	β_1	0.0008	0.027	-0.038	0.97		β_1	0.0003	0.029	-0.042	0.93		
		β_2	0.002	0.026	0.038	0.96		β_2	0.002	0.025	0.035	0.98	275.243 (9.32)	
		δ	0.003	0.038	0.027	0.95		δ	0	0.043	0.031	0.89		
LA-P	$\lambda = 0,5$	β_1	0.01	0.05	-0.07	0.93	LA-RP	$\lambda = 0,5$	β_1	0.016	0.043	-0.06	0.98	
		β_2	0.002	0.046	0.065	0.94			β_2	0.012	0.047	0.066	0.97	348 (18.15)
		δ	0.005	0.037	-0.054	0.92			δ	0.003	0.033	-0.048	0.94	
	$\lambda = 4$	β_1	0.01	0.04	-0.056	0.95		β_1	-0.027	0.027	-0.04	0.78		
		β_2	0.012	0.038	0.053	0.92		β_2	0.016	0.041	0.058	0.94	1347.61 (307.2)	
		δ	0.006	0.04	0.029	0.92		δ	0.001	0.036	0.026	0.97		

Fonte: Elaborada pelo autor.

Tabela 16 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostra $r = 200$. 5000 iterações, descartando 2500 como *burn-in*.

Ligação	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	Ligação	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	
C-P	$\lambda = 0,5$	β_1	0.0123	0.029	-0.041	0.98	C-RP	$\lambda = 0,5$	β_1	0.004	0.030	-0.043	0.96	
		β_2	0.012	0.033	0.046	0.95			β_2	0.010	0.033	0.046	0.90	1143.411 (46.61)
		δ	0.005	0.019	-0.028	0.95			δ	-0.003	0.017	-0.024	0.94	
	$\lambda = 4$	β_1	0.012	0.024	-0.034	0.93		β_1	0.013	0.023	-0.033	0.95		
		β_2	0.012	0.025	0.035	0.95		β_2	0.013	0.026	0.036	0.93	1099.089 (63.05)	
		δ	0.004	0.022	0.016	0.93		δ	-0.011	0.022	0.016	0.85		
GR-P	$\lambda = 0,5$	β_1	0.005	0.023	-0.033	0.93	GR-RP	$\lambda = 0,5$	β_1	-0.0003	0.027	-0.038	0.88	
		β_2	0.004	0.023	0.033	0.96			β_2	-0.001	0.024	0.034	0.93	1466.199 (112.45)
		δ	0.006	0.019	-0.027	0.97			δ	0.002	0.022	-0.032	0.93	
	$\lambda = 4$	β_1	0.005	0.012	-0.017	0.94		β_1	-0.006	0.031	-0.045	0.81		
		β_2	0.003	0.012	0.017	0.93		β_2	-0.005	0.025	0.037	0.82	1389.66 (120.63)	
		δ	0.004	0.017	0.012	0.91		δ	0	0.034	0.024	0.81		
L-P	$\lambda = 0,5$	β_1	0.006	0.026	-0.037	0.95	L-RP	$\lambda = 0,5$	β_1	0.0005	0.025	-0.036	0.97	
		β_2	0.007	0.027	0.039	0.91			β_2	0.007	0.027	0.039	0.92	1543.623 (231.80)
		δ	0	0.018	-0.026	0.88			δ	0.003	0.017	-0.024	0.94	
	$\lambda = 4$	β_1	0.003	0.020	-0.028	0.94		β_1	0.005	0.019	-0.028	0.97		
		β_2	0.003	0.019	0.028	0.95		β_2	0.003	0.019	0.028	0.98	1351.497 (2.29)	
		δ	0.001	0.016	0.011	0.95		δ	0.002	0.017	0.013	0.91		
N-P	$\lambda = 0,5$	β_1	0.007	0.019	-0.027	0.94	N-RP	$\lambda = 0,5$	β_1	0.003	0.017	-0.024	0.95	
		β_2	0.004	0.018	0.025	0.93			β_2	0.004	0.017	0.025	0.94	1153.177 (192.16)
		δ	0.005	0.02	-0.028	0.88			δ	0.007	0.019	-0.028	0.91	
	$\lambda = 4$	β_1	0.004	0.013	-0.019	0.91		β_1	0.001	0.015	-0.022	0.92		
		β_2	0.005	0.013	0.019	0.94		β_2	0.002	0.016	0.022	0.91	935.488 (8.01)	
		δ	0.005	0.017	0.012	0.97		δ	0.002	0.022	0.016	0.91		
LA-P	$\lambda = 0,5$	β_1	-0.008	0.024	-0.034	0.89	LA-RP	$\lambda = 0,5$	β_1	-0.013	0.025	-0.036	0.91	
		β_2	-0.012	0.024	0.035	0.88			β_2	-0.016	0.025	0.036	0.91	1211.98 (213.77)
		δ	-0.007	0.017	-0.025	0.95			δ	-0.005	0.018	-0.026	0.94	
	$\lambda = 4$	β_1	-0.03	0.027	-0.04	0.75		β_1	-0.027	0.027	-0.04	0.78		
		β_2	-0.027	0.025	0.037	0.79		β_2	-0.023	0.024	0.035	0.84	1347.61 (307.2)	
		δ	-0.022	0.035	0.026	0.59		δ	-0.02	0.033	0.024	0.68		

Fonte: Elaborada pelo autor.

Tabela 17 – Comparação de modelos para diferentes conjuntos de dados de tamanho 100, gerados a partir de um modelo Logito.

Ligação	DIC	EAIC	EBIC	WAIC	Lm	LPML
Logito	5964.46	7053.71	7058.92	7051.60	2414.93	-3525.80
C-P	5976.60	7073.28	7081.10	7070.15	2414.57	-3535.08
C-RP	5973.29	7073.24	7081.06	7070.13	2414.16	-3535.07
G-RP	6120.64	7267.11	7274.92	7265.38	2345.68	-3632.69
GR-RP	6117.88	7267.30	7275.11	7265.13	2345.17	-3632.56
L-P	5960.82	7055.62	7063.43	7052.45	2414.66	-3526.22
L-RP	5966.44	7055.57	7063.38	7052.41	2414.35	-3526.20
N-P	6000.49	7154.95	7162.77	7152.23	2358.85	-3576.11
N-RP	6006.85	7155.48	7163.30	7152.75	2358.21	-3576.37

Fonte: Elaborada pelo autor.

4.4 Aplicação

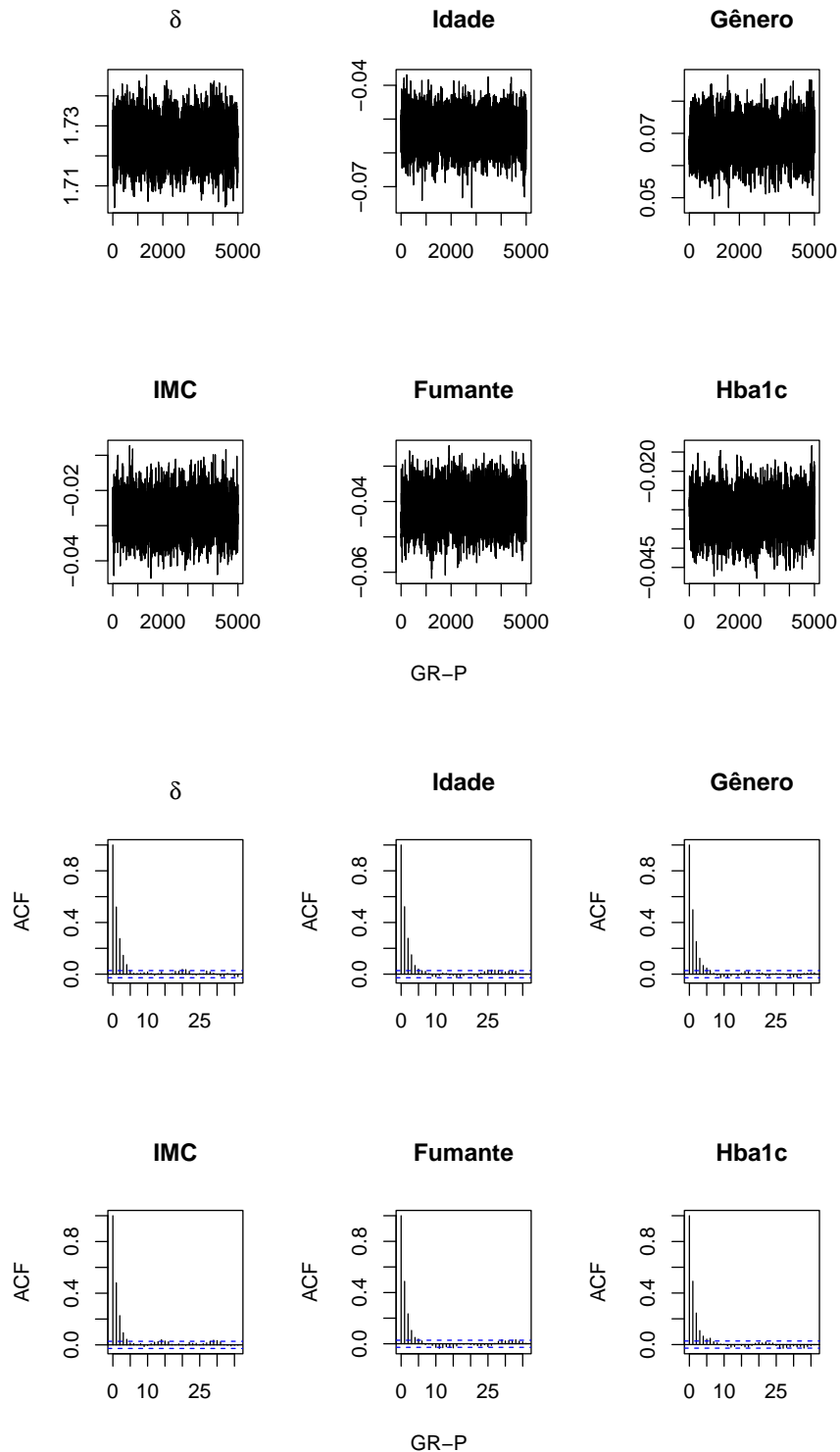
Nesse momento, ilustramos o uso do modelo proposto, sob funções de ligação assimétrica, com uma aplicação aos dados periodontais descritos na Seção 4.1. A resposta é uma variável binária para a presença de doença periodontal com base em medições do nível de inserção clínica (CAL), $Y = 1$ se $CAL > 5\text{mm}$ ou $Y = 0$, caso contrário. As medições de CAL estão disponíveis para $m = 168$ locais de $r = 288$ indivíduos.

As covariáveis foram padronizadas subtraindo a média dos valores da covariável e dividindo pelo desvio padrão dos valores. Os resultados mostrados nesta Seção são baseados em 10.000 iterações após descartar as primeiras 5.000 como *burn-in*. A avaliação da convergência das cadeias de Markov foi baseada na inspeção visual de gráficos de traços e gráficos de autocorrelação. Todos eles indicaram que as cadeias atingiram a estacionariedade relativamente rápido e com baixas autocorrelações. Para ilustrar, nas Figuras 17 e 18 apresentamos os gráficos de traços e autocorrelação dos modelos GR-P e GR-RP; os outros modelos não foram incluídos para economizar espaço. No entanto, cabe ressaltar que estes também mostraram comportamento semelhante ao exibido nas Figura 17 e 18. As estimativas dos coeficientes dos modelos estão na Tabela 18. Percebe-se que todas as covariáveis apresentaram significância estatística para explicar a ocorrência do evento de interesse, pois em nenhum dos intervalos de credibilidade dos coeficientes há a presença do zero.

Comparamos os modelos espaciais com diferentes funções de ligação em termos das métricas de bondade de ajuste apresentadas no Capítulo 2, e das métricas apresentadas na Seção 3.4. Os resultados aparecem na Tabela 19. A partir da Tabela, notamos que os modelos com uma função de ligação assimétrica têm uma clara vantagem sobre os modelos Logito, Probit e Cloglog, e são claramente preferíveis em termos dos critérios acima mencionados.

Para exemplificar o uso da divergência de Bregman (Seção 2.6) para detecção de *outliers*, usamos novamente a perturbação de caso excluído dada em 2.36 para todas as funções de ligação apresentadas na Tabela 2. Calculamos as probabilidades de ser um *outlier*, conforme dado em 2.43, com uma função

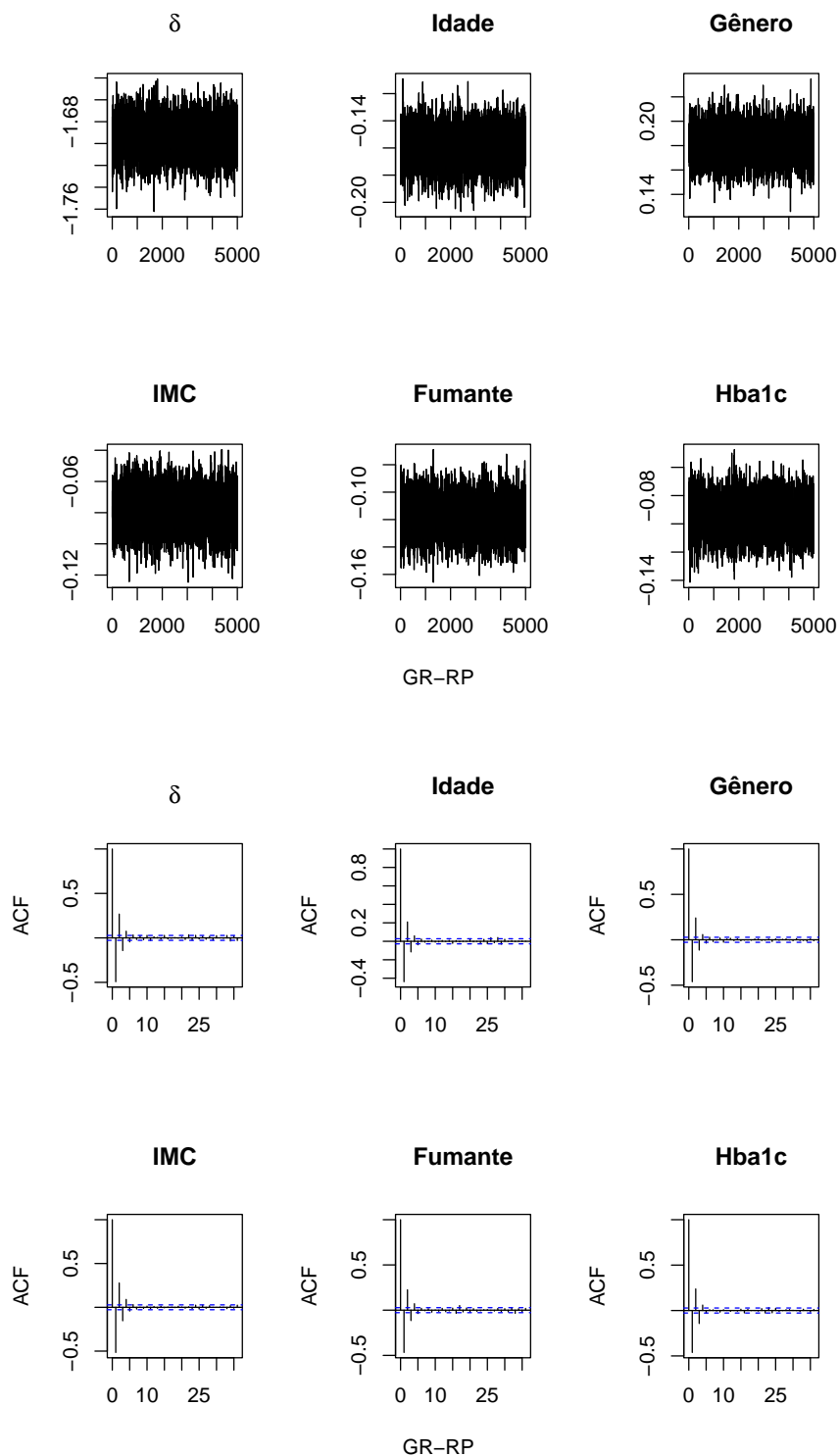
Figura 17 – Trace-plots e gráficos de autocorrelação para o modelos GR-P.



Fonte: Elaborada pelo autor.

convexa ψ_1 (divergência de Kullback-Leibler) e as divergências de Bregman normalizadas em 2.44 para todas as observações, como tínhamos feito para o modelo apresentado no Capítulo 3.

Figura 18 – Trace-plots e gráficos de autocorrelação para os modelos GR-RP.



Fonte: Elaborada pelo autor.

As probabilidades estimadas de *outliers* (gráfico superior) e os valores da divergência de Bregman normalizada (gráfico inferior) são apresentados na Figura 19, para cada observação e função de ligação. Pode-se ver que modelos com ligação assimétrica levam a um número maior de *outliers* possíveis

Tabela 18 – Estimativas dos parâmetros usando diferentes funções de ligação: C-P, C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, Probit, Logito e Cloglog. As estimativas são a média a posteriori, desvio padrão (em parênteses) e intervalo de credibilidade de 95%.

	C-P	C-RP	N-P	N-RP	L-P
Idade	-0.080 (0.007) (-0.094, -0.066)	-0.233 (0.020) (-0.275, -0.192)	-0.061 (0.005) (-0.072, -0.051)	-0.165 (0.014) (-0.193, -0.137)	-0.098 (0.008) (-0.114, -0.081)
Sexo	0.097 (0.007) (0.083, 0.112)	0.266 (0.020) (0.224, 0.308)	0.076 (0.005) (0.065, 0.087)	0.199 (0.014) (0.169, 0.229)	0.121 (0.009) (0.102, 0.139)
IMC	-0.039 (0.006) (-0.052, -0.025)	-0.101 (0.018) (-0.137, -0.064)	-0.030 (0.005) (-0.040, -0.020)	-0.082 (0.013) (-0.108, -0.055)	-0.048 (0.008) (-0.064, -0.032)
Fumante	-0.060 (0.006) (-0.074, -0.046)	-0.169 (0.019) (-0.207, -0.130)	-0.047 (0.005) (-0.058, -0.036)	-0.125 (0.014) (-0.153, -0.096)	-0.075 (0.008) (-0.092, -0.057)
Hba1c	-0.046 (0.006) (-0.059, -0.034)	-0.137 (0.019) (-0.175, -0.100)	-0.036 (0.005) (-0.046, -0.026)	-0.097 (0.013) (-0.124, -0.070)	-0.057 (0.008) (-0.073, -0.041)
δ	1.300 (0.006) (1.288, 1.312)	-2.119 (0.015) (-2.149, -2.08)	1.309 (0.006) (1.297, 1.321)	-2.120 (0.015) (-2.150, -2.090)	1.298 (0.006) (1.287, 1.310)
	L-RP	GR-P	GR-RP	Probit	Logito
Idade	-0.267 (0.022) (-0.313, -0.222)	-0.053 (0.005) (-0.063, -0.043)	-0.159 (0.013) (-0.186, -0.133)	-0.036 (0.005) (-0.047, -0.025)	-0.057 (0.009) (-0.075, -0.039)
Sexo	0.320 (0.024) (0.271, 0.367)	0.067 (0.005) (0.056, 0.079)	0.181 (0.013) (0.154, 0.208)	0.041 (0.005) (0.029, 0.053)	0.064 (0.009) (0.046, 0.083)
IMC	-0.130 (0.022) (-0.173, -0.087)	-0.026 (0.005) (-0.037, -0.016)	-0.079 (0.012) (-0.102, -0.054)	-0.017 (0.005) (-0.028, -0.005)	-0.027 (0.009) (-0.046, -0.009)
Fumante	-0.201 (0.023) (-0.246, -0.155)	-0.041 (0.005) (-0.052, -0.031)	-0.117 (0.013) (-0.142, -0.092)	-0.025 (0.005) (-0.036, -0.014)	-0.041 (0.009) (-0.059, -0.022)
Hba1c	-0.157 (0.022) (-0.201, -0.114)	-0.031 (0.004) (-0.041, -0.021)	-0.095 (0.012) (-0.119, -0.070)	-0.020 (0.005) (-0.031, -0.009)	-0.032 (0.008) (-0.050, -0.015)
δ	-2.116 (0.015) (-2.146, -2.086)	1.730 (0.006) (1.717, 1.742)	-1.695 (0.016) (-1.728, -1.663)	-	-
	Cloglog	LA-P	LA-RP		
Idade	-0.023 (0.003) (-0.031, -0.016)	-0.056 (0.004) (-0.064, -0.048)	-0.161 (0.014) (-0.191, -0.133)		
Sexo	0.027 (0.004) (0.019, 0.035)	0.068 (0.004) (0.059, 0.076)	0.187 (0.015) (0.158, 0.217)		
IMC	-0.011 (0.004) (-0.020, -0.003)	-0.026 (0.004) (-0.035, -0.018)	-0.069 (0.013) (-0.095, -0.043)		
Fumante	-0.017 (0.004) (-0.025, -0.008)	-0.042 (0.003) (-0.049, -0.035)	-0.119 (0.013) (-0.146, -0.092)		
Hba1c	-0.013 (0.004) (-0.021, -0.005)	-0.032 (0.004) (-0.042, -0.023)	-0.096 (0.013) (-0.123, -0.070)		
δ	-	1.310 (0.006) (1.298, 1.322)	-2.122 (0.015) (-2.153, -2.090)		

Fonte: Elaborada pelo autor.

(gráfico superior), como ocorreu no Capítulo 3. Como antes, acreditamos que a flexibilidade dos modelos assimétricos em se ajustar melhor a dados com a presença de desbalanceamento, justifique a sensibilidade desses modelos a observações discrepantes.

Em geral, para esses modelos, não há uma única observação que tenha uma probabilidade de ser um *outlier* maior que 0,40; a observação que apresentou a maior probabilidade foi verificada para o modelo sob ligação LA-P para a observação 12, com um valor de 0.3535. Para os modelos Logito, Probit e Cloglog, os comportamentos foram semelhantes, nenhuma observação com probabilidade de ser um *outlier* maior que 0.1, exceto para a ligação Probit que apresentou uma observação com um valor de probabilidade de 0,102.

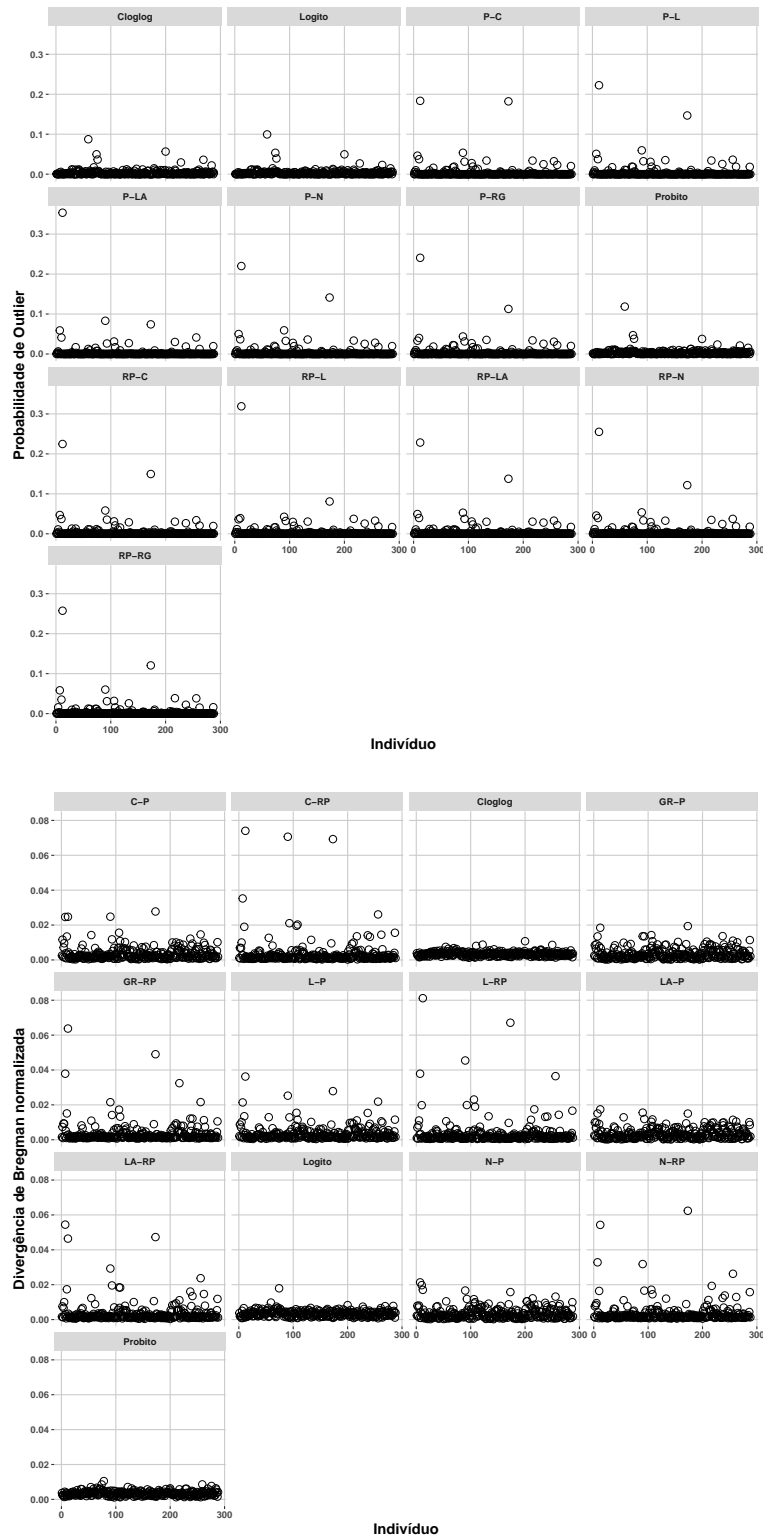
Tabela 19 – Comparação de modelos usando diferentes funções de ligação: C-P, C-RP, N-P, N-RP, L-P, L-RP, LA-P, LA-RP, Logito, Probit, Cloglog, GR-P e GR-RP.

	C-P	C-RP	N-P	N-RP	L-P
WAIC	27445.88	27384.140	27478.810	27386.250	27417.1900
DIC	27440.420	27309.080	27523.140	27323.800	27390.9900
EAIC	27148.880	27277.750	27054.520	27260.380	27205.8000
EBIC	27170.850	27299.720	27076.500	27282.360	27227.7800
CPO	-47.650	-47.540	-47.700	-47.550	-47.6000
Lm	7322.677	7302.794	7331.593	7302.478	7316.67
	L-RP	LA-P	LA-RP	GR-P	GR-RP
WAIC	27376.600	27435.670	27390.920	27519.590	27371.210
DIC	27303.570	27527.090	27326.710	27587.180	27310.820
EAIC	27283.060	26913.380	27254.790	27003.630	27235.240
EBIC	27305.040	26935.350	27276.770	27025.610	27257.220
CPO	-47.530	-47.610	-47.550	-47.770	-47.520
Lm	7303.029	7328.952	7301.284	7338.451	7299.044
	Logito	Probit	Cloglog		
WAIC	67621.91	67762.21	93195.27		
DIC	67143.74	67378.91	92993.06		
EAIC	66978.26	66961.86	92378.50		
EBIC	66996.58	66980.18	92396.82		
CPO	-117.43	-117.65	-161.83		
Lm	24112.68	24072.70	29413.23		

Fonte: Elaborada pelo autor.

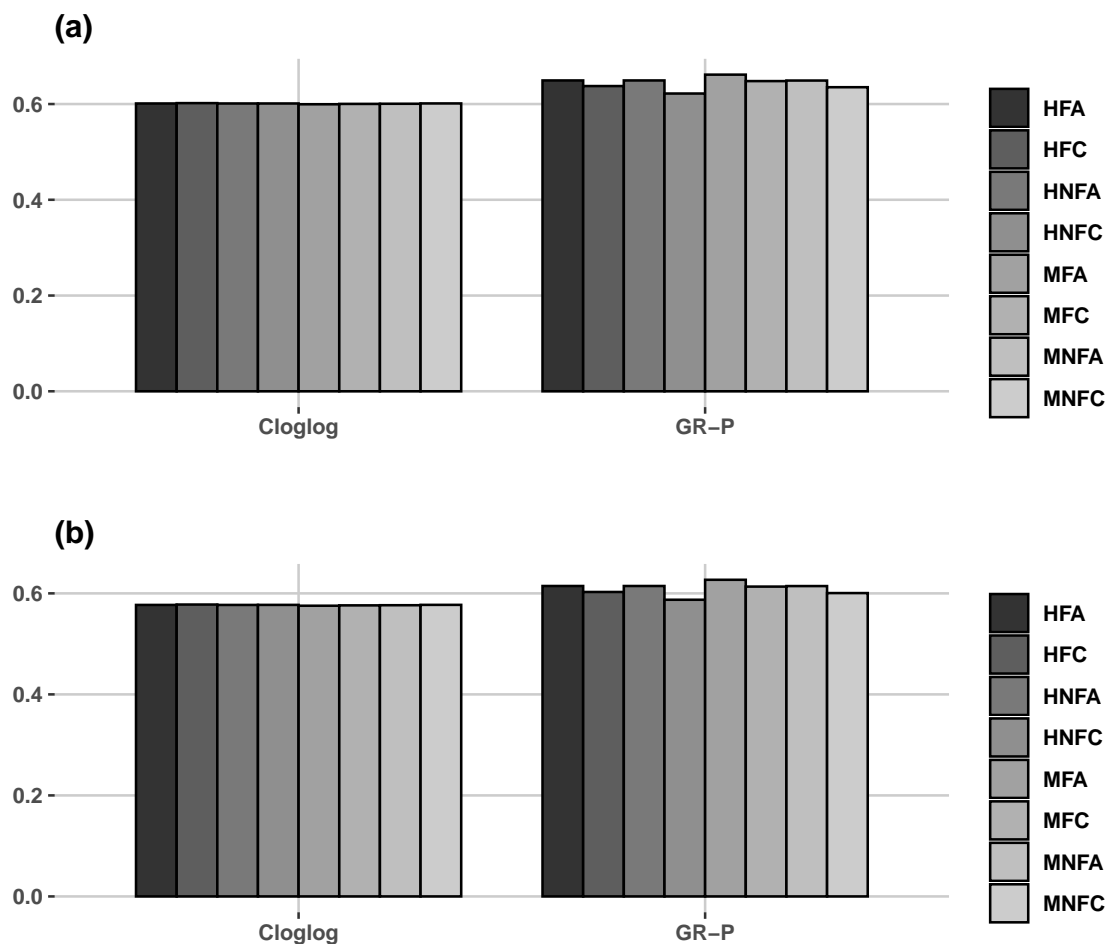
Quando avaliamos as divergências de Bregman normalizadas na Figura 19, percebemos comportamento similar ao apresentado no Capítulo 3. Os modelos com função de ligação flexível apresentam valores maiores da divergência de Bregman normalizada para cada observação. Percebe-se que o modelo L-RP apresenta muitas observações com valores altos da divergência de Bregman normalizada, e de fato, sob este modelo, temos o maior valor da divergência de Bregman normalizada, a saber, 0.081 para a observação 12. Esta observação, no banco de dados, refere-se a um indivíduo do sexo feminino, com idade de 42 anos, fumante e com diabetes alta.

Figura 19 – Probabilidades a posteriori estimadas de serem *outliers* com função convexa ψ_1 , e divergência de Bregman normalizada com função convexa ψ_1 , contra observações deletadas para os dados periodontais.



Fonte: Elaborada pelo autor.

Figura 20 – Gráfico (a): probabilidades preditivas a posteriori de doença periodontal para um sujeito aleatório com idade de 25 anos, IMC 25, e combinando todos os valores de gênero, hábitos de fumar e níveis de HbA1c. Gráfico (b): probabilidades preditivas posteriores de doença periodontal para um sujeito aleatório com idade de 50 anos, IMC 40, e combinando todos os valores de gênero, hábitos de fumar e níveis de HbA1c. HNFC: Homem - Não fuma - Diabetes controlado; HNFA: Homem - Não fuma - Diabetes alto; HFC: Homem - Fumo - Diabetes controlado; MFA: Homem - Fumo - Diabetes alto; MNFC: Mulher - Não fuma - Diabetes controlado; MNFA: Mulher - Não fuma - Diabetes alto; MFC: Mulher - Fumo - Diabetes controlado; MFA: Mulher - Fumo - Diabetes alto;



Fonte: Elaborada pelo autor.

Prosseguindo em nossas análises, comparamos as estimativas pontuais dos efeitos de covariáveis em termos de probabilidades preditivas (veja a expressão 4.7) a posteriori de doença periodontal utilizando o modelo flexível GR-P e Cloglog. Isso foi realizado calculando essa probabilidade para um sujeito aleatório com idade de 25 anos, IMC 25, e combinando todos os valores de gênero, hábitos de fumar e níveis de HbA1c. Em seguida, repetimos esse exercício para um sujeito aleatório com idade de 50 anos, IMC 40 e a mesma combinação de gênero, hábitos de fumar e níveis de HbA1c. Os resultados aparecem na Figura 20. Pelos resultados, percebe-se que o modelo com ligação assimétrica Cloglog apresenta altas

probabilidades preditivas para os dois cenários (indivíduos com idade de 25 e 50 anos) considerando todas as combinações. Entretanto, as probabilidades preditivas do modelo Gumbel Reversa Potência (GR-P) são as mais altas quando analisamos os dois cenários considerados. Vale destacar que a ligação Cloglog é um caso particular do modelo GR-P quando $\lambda = 1$ ($\delta = 0$).

Na Tabela 20, estamos comparando o ajuste dos modelos, utilizando na parte espacial:

1. usando o amostrador exato para a distribuição a posteriori de Ω (GW-E), a saber, $\Omega|\beta, \delta, \phi, \mathbf{y}, \mathbf{X} \sim \text{G-Wishart}_{\mathbf{W}}\left(\kappa + r, \mathbf{S} + \sum_{s=1}^r \phi_s \phi_s^{\top}\right)$;
2. atribuindo priori paramétrica CAR para Ω .

Percebe-se, pela Tabela, que os melhores resultados de ajuste foram para os modelos sob esquema de amostragem GW-E, exceto pela medida L , em que os ajustes realizados sob modelo CAR apresentaram resultados melhores. Em relação ao tempo de execução dos modelos, os ajustes realizados com priori CAR são mais rápidos. Também cabe destacar que, sob distribuição a priori CAR, para alguns modelos, a métrica CPO não pôde ser calculada (modelos C-P, N-P, L-P, LA-P e GR-RP).

Na Tabela 21, temos os resultados, para cada modelo, das métricas de predição apresentadas na Seção 2.7. Vale mencionar que as estimativas de probabilidade $P(Y_{si} = 1)$ foram realizadas por meio de estimativas da média preditiva em 4.7. Pelos resultados, vemos que os modelos potência e reversa de potência apresentam resultados similares com destaque para o modelo GR-P, que apresenta resultados tão bons quanto qualquer outro modelo em quase todas as métricas, exceto em duas, ACC e SN. Cabe destacar também o desempenho do modelo Cloglog, que apresentou bons resultados em algumas métricas. Também é interessante observar que os pontos de corte estimados pelos modelos potência e reversa de potência, segundo o critério apresentado na Seção 2.7, não ultrapassaram 0.10, o que é coerente com a proporção de CAL presente nos dados, como podemos constatar pela Figura 16.

Na Tabela 22, comparamos os resultados das métricas de qualidade da predição sob os esquemas de amostragem GW-E e atribuindo distribuição a priori CAR para Ω . Percebe-se pela Tabela que os ajustes realizados usando priori CAR proporcionaram, para algumas métricas, os melhores valores, com destaque para ACC, SN e ES (modelos GR-P, L-RP e GR-P, respectivamente) com valores bem relevantes.

Tabela 20 – Comparação de modelos usando diferentes funções de ligação: C-P, C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato de [Mohammadi e Wit \(2015\)](#) (GW-E); ii) e atribuindo priori CAR para a estrutura espacial (CAR).

Métrica	Método	C-P	C-RP	N-P	N-RP	L-P	L-RP	LA-P	LA-RP	GR-P	GR-RP
WAIC	GW-E	27445.880	27384.140	27478.810	27386.250	27417.1900	27376.600	27435.670	27390.920	27519.590	27371.210
	CAR	893719.900	68696.630	11158198.700	164698.750	887945.250	67778.970	2726004.480	113802.190	30318917.040	435813.980
DIC	GW-E	27440.420	27309.080	27523.140	27323.800	27390.9900	27303.570	27527.090	27326.710	27587.180	27310.820
	CAR	45765.090	29501.290	97284.070	32702.780	45319.620	29380.050	65143.810	31278.580	157734.600	36349.050
EAIC	GW-E	27148.880	27277.750	27054.520	27260.380	27205.8000	27283.060	26913.380	27254.790	27003.630	27235.240
	CAR	40622.080	28317.820	81970.530	30073.110	39929.190	28246.600	59260.600	29279.680	141460.840	31837.610
EBIC	GW-E	27170.850	27299.720	27076.500	27282.360	27227.7800	27305.040	26935.350	27276.770	27025.610	27257.220
	CAR	40644.060	28339.800	81992.510	30095.080	39951.170	28268.580	59282.580	29301.660	141482.820	31859.590
CPO	GW-E	-47.650	-47.540	-47.700	-47.550	-47.6000	-47.530	-47.610	-47.550	-47.770	-47.520
	CAR	-	-69.110	-	-134.170	-	-80.090	-	-92.430	-	-
Lm	GW-E	7322.677	7302.794	7331.593	7302.478	7316.67	7303.029	7328.952	7301.284	7338.451	7299.044
	CAR	5892.948	7043.488	5139.051	6707.136	7043.488	7026.219	5380.435	6865.051	4872.259	6751.631
Tempo (em seg.)	GW-E	9394.304	9549.379	8714.653	8906.014	11764.95	12049.09	10412.21	10347.38	10917.25	11274.28
	CAR	8215.497	7020.785	6519.593	6588.059	9217.21	9520.668	7519.685	7526.479	8482.446	8030.21

Fonte: Elaborada pelo autor.

Tabela 21 – Comparação de modelos, segundo métricas de qualidade da predição, usando diferentes funções de ligação C-P, C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, Logito, Probit e Cloglog. P_cmax - estimativa do ponto de corte.

	C-P	C-RP	N-P	N-RP	L-P	L-RP	LA-P	LA-RP
P _c max	0.080	0.080	0.100	0.080	0.080	0.080	0.100	0.080
ACC	0.530	0.499	0.781	0.503	0.517	0.498	0.777	0.505
SN	0.709	0.676	0.376	0.683	0.700	0.673	0.432	0.679
ES	0.514	0.483	0.818	0.486	0.500	0.482	0.808	0.489
PRC	0.117	0.106	0.157	0.107	0.112	0.105	0.169	0.107
FPR	0.486	0.517	0.182	0.514	0.500	0.518	0.192	0.511
CSI	0.111	0.101	0.125	0.102	0.107	0.100	0.138	0.102
SSI	0.059	0.053	0.066	0.054	0.057	0.053	0.074	0.054
FAITH	0.294	0.278	0.406	0.280	0.287	0.277	0.406	0.280
KAPPA	0.067	0.046	0.118	0.049	0.059	0.045	0.140	0.049
MCC	0.123	0.088	0.134	0.094	0.111	0.086	0.162	0.093
AUC	0.119	0.102	0.127	0.105	0.114	0.102	0.135	0.103
F1	0.200	0.183	0.222	0.186	0.194	0.182	0.243	0.185
	GR-P	GR-RP	Logito	Probit	Cloglog			
P _c max	0.100	0.080	0.500	0.520	0.640			
ACC	0.803	0.500	0.518	0.799	0.738			
SN	0.381	0.692	0.692	0.335	0.508			
ES	0.841	0.482	0.502	0.841	0.758			
PRC	0.178	0.108	0.112	0.160	0.160			
FPR	0.159	0.518	0.498	0.159	0.242			
CSI	0.138	0.103	0.106	0.121	0.138			
SSI	0.074	0.054	0.056	0.065	0.074			
FAITH	0.417	0.279	0.288	0.413	0.390			
KAPPA	0.146	0.050	0.058	0.117	0.134			
MCC	0.160	0.097	0.107	0.128	0.167			
AUC	0.133	0.105	0.110	0.125	0.068			
F1	0.242	0.187	0.192	0.216	0.243			

Fonte: Elaborada pelo autor.

Tabela 22 – Comparação de modelos, segundo métricas de qualidade da predição, usando diferentes funções de ligação: C-P,C-RP, N-P, N-RP, L-P, L-RP, GR-P, GR-RP, LA-P e LA-RP, sob os esquemas de amostragem: i) mostrando da distribuição G-Wishart pelo método exato de [Mohammadi e Wit \(2015\)](#) (GW-E); ii) e atribuindo priori CAR para a estrutura espacial (CAR).

Métrica	C-P		C-RP		N-P		N-RP		L-P		L-RP		LA-P		LA-RP		GR-P		GR-RP	
	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR	GW-E	CAR
Pcmax	0.08	0.03	0.08	0.11	0.10	0.01	0.08	0.08	0.08	0.02	0.08	0.06	0.10	0.01	0.08	0.08	0.10	0.01	0.08	0.09
ACC	0.53	0.86	0.50	0.88	0.78	0.88	0.50	0.86	0.52	0.83	0.50	0.42	0.78	0.83	0.51	0.79	0.80	0.91	0.50	0.86
SN	0.71	0.24	0.68	0.17	0.38	0.16	0.68	0.22	0.70	0.29	0.67	0.78	0.43	0.30	0.68	0.30	0.38	0.03	0.69	0.21
ES	0.51	0.92	0.48	0.94	0.82	0.95	0.49	0.92	0.50	0.88	0.48	0.39	0.81	0.88	0.49	0.84	0.84	0.99	0.48	0.92
PRC	0.12	0.22	0.11	0.21	0.16	0.22	0.11	0.20	0.11	0.18	0.11	0.10	0.17	0.19	0.11	0.14	0.18	0.20	0.11	0.19
FPR	0.49	0.08	0.52	0.06	0.18	0.05	0.51	0.08	0.50	0.12	0.52	0.62	0.19	0.12	0.51	0.16	0.16	0.01	0.52	0.08
CSI	0.11	0.13	0.10	0.11	0.13	0.10	0.10	0.12	0.11	0.13	0.10	0.10	0.14	0.13	0.10	0.11	0.14	0.03	0.10	0.11
SSI	0.06	0.07	0.05	0.06	0.07	0.05	0.05	0.06	0.06	0.07	0.05	0.05	0.07	0.07	0.05	0.06	0.07	0.02	0.05	0.06
FAITH	0.29	0.44	0.28	0.45	0.41	0.45	0.28	0.44	0.29	0.43	0.28	0.24	0.41	0.43	0.28	0.41	0.42	0.46	0.28	0.44
KAPPA	0.07	0.15	0.05	0.12	0.12	0.13	0.05	0.13	0.06	0.14	0.05	0.04	0.14	0.14	0.05	0.09	0.15	0.04	0.05	0.13
MCC	0.12	0.15	0.09	0.12	0.13	0.13	0.09	0.13	0.11	0.14	0.09	0.09	0.16	0.15	0.09	0.10	0.16	0.05	0.10	0.13
AUC	0.12	0.05	0.10	0.11	0.13	0.01	0.11	0.12	0.11	0.07	0.10	0.12	0.14	0.03	0.10	0.12	0.13	0.00	0.11	0.12
F1	0.20	0.23	0.18	0.19	0.22	0.19	0.19	0.21	0.19	0.23	0.18	0.18	0.24	0.23	0.19	0.19	0.24	0.06	0.19	0.20

Fonte: Elaborada pelo autor.

4.5 Discussões Finais

Neste Capítulo, apresentamos um modelo de regressão binária espacial sob função de ligação flexível, que pode ser visto como uma generalização do modelo espacial apresentado no Capítulo 3. Os métodos de amostragem desenvolvidos para os parâmetros do modelo são idênticos aos desenvolvidos para o modelo anterior e, portanto, são eficientes em termos computacionais. Implementações R/Rcpp do modelo estão disponíveis no repositório `GitHub`: <<https://github.com/alan-assuncao/SpatialBinReg>>.

Através do estudo de simulação, obtemos resultados semelhantes aos que tínhamos obtido para o modelo apresentado no Capítulo 3, ou seja, mostramos que o modelo proposto, sob as funções de ligação flexíveis, e os métodos de amostragem das distribuições a posteriori apresentaram bom desempenho em termos de recuperação do verdadeiro valor dos parâmetros nas diversas combinações de tamanho de amostra e valores de λ considerados. Resultados similares também foram obtidos quando se avaliou o desempenho desses modelos em um cenário de má especificação, considerando critérios de bondade de ajuste e qualidade da predição.

No tocante à parte da aplicação aos dados reais, alguns resultados podem ser destacados. Os modelos com funções de ligação assimétrica apresentaram resultados melhores em termos de qualidade de ajuste quando comparados aos modelos com funções de ligação Logito, Probit e Cloglog. Na parte do diagnóstico de *outliers*, os modelos assimétricos foram os que identificaram um maior número de possíveis observações influentes. Como antes, acreditamos que isso advenha do fato de que esses modelos com assimetria flexível são mais sensíveis a observações discrepantes.

Também comparamos os modelos sob função de ligação assimétrica ajustados sob distribuição a priori G-Wishart e distribuição a priori CAR. Os resultados, em termos de qualidade de ajuste, mostraram que os ajustes realizados sob distribuição a priori G-Wishart apresentam melhor desempenho em todas as métricas consideradas, exceto na medida L, no qual os modelos ajustados com priori CAR apresentaram os melhores resultados. Quando avaliamos sob critérios de qualidade da predição, os resultados apontam que os ajustes com distribuição a priori G-Wishart são melhores para a maioria das métricas consideradas. Também realizamos a mesma comparação, utilizando métricas de predição para os modelos Logito, Probit e Cloglog. Para a metade das métricas avaliadas, o modelo Cloglog sob distribuição a priori G-Wishart é o mais preferível. Esses resultados encontram-se no Apêndice E (Seção E.1).

REGRESSÃO BINÁRIA ESPACIAL COM MODELOS DE GRAFOS ACÍCLICOS DIRECIONADOS AUTOREGRESSIVOS (DAGAR)

Apresentamos, nos Capítulos anteriores, modelos espaciais ajustados sob distribuição a priori G-Wishart, e buscamos avaliar seu desempenho por meio de estudos de simulação e aplicações em dados reais. Neste Capítulo, buscamos desenvolver um estudo do modelo visto no Capítulo 4 sob modelagem espacial DAGAR para a matriz de precisão Ω .

As análises aqui desenvolvidas são preliminares, uma vez que os resultados que obtivemos com o modelo DAGAR foram bem sucedidos em cenários específicos, como descreveremos logo mais na Seção 5.2. Porém, quando buscamos aplicar os modelos sob priori DAGAR ao conjunto de dados reais de periodontite (Seção 4.1), verificou-se problemas de convergência nas cadeias de Markov dos parâmetros, e por isso, decidimos não incluir essa análise neste momento, e buscaremos apresentá-la em um trabalho futuro. No estudo de simulação proposto, fixamos o parâmetro ρ do modelo DAGAR em 0.9, como fizemos para o ρ da priori CAR, o qual analisamos nos Capítulos anteriores, e que também utilizamos para ser a matriz escala S da distribuição G-Wishart (veja Seção 3.2). O modelo DAGAR é descrito na Seção 5.1; um estudo de simulação do modelo sob priori DAGAR é apresentado na Seção 5.2, e finalizamos esse Capítulo com uma breve discussão dos resultados na Seção 5.3.

5.1 O modelo DAGAR

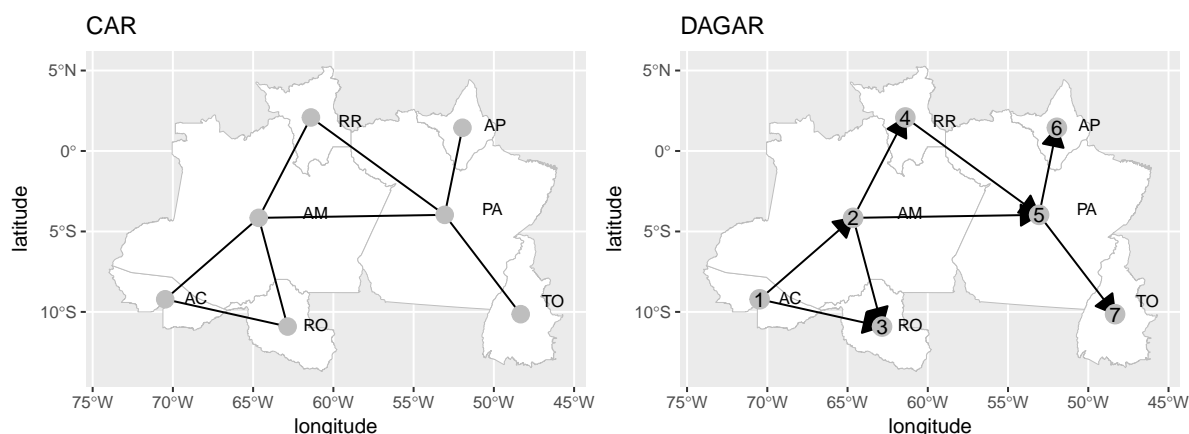
O modelo DAGAR (*directed acyclic graph autoregressive*) é um modelo paramétrico, assim como o CAR e o ICAR, autoregressivo baseado em grafos acíclicos direcionados. O modelo DAGAR foi introduzido por [Datta et al. \(2019\)](#), no qual a matriz de precisão Ω é construída usando um grafo acíclico direcionado (DAG) a partir do grafo não direcionado original. As vantagens desse modelo em relação aos modelos tradicionais CAR e ICAR, são, entre outras, a possibilidade do parâmetro ρ possuir uma interpretação clara como um parâmetro de autocorrelação espacial, e a escalabilidade para analisar

conjuntos de dados de área muito grandes (DATTA *et al.*, 2019). Perceba que o modelo CAR em 3.3 também possui o parâmetro ρ , mas, como argumentam Datta *et al.* (2019), este não pode ser interpretado totalmente como um parâmetro de autocorrelação espacial.

Parafraseando os próprios autores do modelo DAGAR, ao invés de modelarmos a matriz de precisão diretamente, modelamos os fatores de Cholesky dos efeitos aleatórios espaciais que, para qualquer distribuição Gaussiana multivariada, são determinados pelas distribuições condicionais dos ϕ_i 's (DATTA *et al.*, 2019). Porém, em decorrência do uso dos fatores de Cholesky, o modelo fica dependente da ordem das observações.

Na Figura 21, temos a diferença entre o grafo não direcionado utilizado para o modelo CAR e o grafo direcionado, considerando a direção do oeste para o leste, sob o modelo DAGAR, para o conjunto de estados da região norte brasileira. No grafo do lado esquerdo da Figura 21, não há ordenação entre os estados, logo a definição do conjunto de vizinhos de cada estado dependeria simplesmente dos estados que possuem fronteira em comum, por exemplo, a modelagem do efeito aleatório espacial atrelado ao estado do Amazonas ficaria restrita às contribuições dos estados vizinhos Acre, Rondônia, Roraima e Pará. No grafo do lado direito, temos uma ordenação dos estados brasileiros no sentido Oeste \rightarrow Leste calculando o centroide de cada estado. Nessa ordenação, o primeiro estado é o Acre, o segundo é o Amazonas, Rondônia em terceiro e assim sucessivamente. Com essa estrutura, considera-se como vizinhos da sub-região i todas as sub-regiões $j < i$, tal que $j \sim i$. Sob essa ótica, o Acre não teria vizinhos, o Amazonas teria o estado do Acre como único vizinho, Rondônia teria os estados do Acre e do Amazonas como vizinhos e assim até chegar no último estado, Tocantins, que teria o estado do Pará como único vizinho.

Figura 21 – Diferença entre o grafo sob o modelo CAR e o grafo acíclico direcionado sob o modelo DAGAR, considerando a direção Oeste \rightarrow Leste, para os estados que compõem o norte brasileiro.



Fonte: Elaborada pelo autor

A partir desse momento, seguiremos o passo a passo apresentado por Datta *et al.* (2019) para

construir o modelo DAGAR. Seremos sucintos em nossa apresentação, tentando destacar os elementos essenciais para a construção do modelo. Os leitores interessados podem consultar o artigo original e bibliografia relacionada para obter mais detalhes.

Como primeiro passo, apresentamos a modelagem de matrizes de covariância baseadas em fatores de Cholesky esparso. Para modelar os fatores de Cholesky, especificamos as distribuições dos efeitos aleatórios espaciais como

$$\phi_1 = \epsilon_1, \quad \phi_2 = b_{21}\phi_1 + \epsilon_2, \dots, \quad \phi_m = b_{m1}\phi_1 \dots + b_{m,m-1}\phi_{m-1} + \epsilon_m, \quad (5.1)$$

em que ϵ_i 's são erros aleatórios independentes $N(0, \tau_i)$. Em 5.1, a matriz $\mathbf{B} = (b_{ij})$ é estritamente triangular inferior. Assumindo-se a matriz diagonal $\mathbf{F} = \text{diag}(\tau_1, \dots, \tau_m)$, temos que $\boldsymbol{\phi} \sim N_m(\mathbf{0}, \mathbf{L}^\top \mathbf{F} \mathbf{L})$, em que $\mathbf{L} = \mathbf{I}_m - \mathbf{B}$, e \mathbf{I}_m é uma matriz identidade $m \times m$. O Teorema 1 a seguir apresentado em [Datta et al. \(2019\)](#), e cuja prova consta no material suplementar do referido artigo, garante que a especificação da matriz de covariância de $\boldsymbol{\phi}$, da forma supracitada, não é restritiva e nos permite apresentá-la dessa forma.

Teorema 1. Seja $\boldsymbol{\phi} \sim N_m(\mathbf{0}, \boldsymbol{\Omega})$, em que $\boldsymbol{\Omega}$ é a matriz de precisão (possivelmente singular). Então existe uma matriz de permutação \mathbf{P} , uma matriz estritamente triangular inferior \mathbf{B} , e uma matriz diagonal \mathbf{F} com entradas não negativas tal que $\mathbf{P}\boldsymbol{\Omega}\mathbf{P}^\top = (\mathbf{I} - \mathbf{B})^\top \mathbf{F} (\mathbf{I} - \mathbf{B})$.

Para apresentarmos o modelo DAGAR completamente, temos de especificar a forma de \mathbf{F} e de \mathbf{B} . Para isso, a seguinte estratégia é considerada: seja $N(i)$ o conjunto de vizinhos tal que $b_{ij} = 0$ para todo $j \notin N(i)$. A definição do conjunto de vizinhos pode ser naturalmente inferida da estrutura de vizinhança subjacente do grafo (veja a Figura 21). Para $i > 1$ definimos $N(i) = \{j < i, j \sim i\}$, em que a restrição $i > j$ é necessária para garantir que \mathbf{B} seja triangular inferior como em 5.1. Assim, as especificações em 5.1 resumem-se a

$$\phi_1 = \epsilon_1, \quad \phi_i = \sum_{j \in N(i)} b_{ij}\phi_j + \epsilon_i, \quad (i = 2, \dots, m). \quad (5.2)$$

Para finalizarmos a definição em 5.2, precisamos especificar b_{ij} e τ_i . Para isso, primeiro note que seria bem intuitivo, quando se trata de dados de área, pensar em atribuir o mesmo peso a todos os vizinhos pertencentes à $N(i)$, ou seja, fazer $b_{ij} = b_i$ para todo $j \in N(i)$. Segundo, seria coerente especificar τ_i de forma proporcional ao tamanho do conjunto de vizinhos da sub-região i , tal que quanto maior o conjunto, maior a precisão condicional. Dessa forma, para $i > 1$ poderíamos pensar em definir $b_i = 1/n_{<i}$, em que $n_{<i} = |N(i)|$ denota a cardinalidade do conjunto de vizinhos da unidade de área i , e $\tau_i \propto n_{<i}$; porém, para $i = 1$ haveria a necessidade de definir b_1 e τ_1 de outra maneira, já que $n_{<1} = 0$.

A solução apresentada por [Datta et al. \(2019\)](#) foi definir uma regra geral para b_{ij} e τ_i por meio da criação de árvores¹ de abrangências locais (*local spanning tree*) do grafo original dos dados, uma vez que grafos de dados de área dificilmente são acíclicos (veja o grafo do lado esquerdo da Figura 21), e assim modelar os efeitos aleatórios espaciais $\boldsymbol{\phi}$ por meio de um modelo similar ao autoregressivo de ordem um (AR(1)). As árvores de abrangência local são baseadas nos grafos que podem ser formados pela unidade de área i e o seu conjunto de vizinhos, para $i = 1, \dots, m$, e permitem que \mathbf{B} seja definida como em 5.1. Por exemplo, do grafo direito na Figura 21, teríamos os seguintes subgrafos (conjunto

¹ árvores, na teoria dos grafos, são grafos que não possuem ciclos

de estados) formados: para o estado do Amazonas: <Acre>; Rondônia: <Acre, Amazonas>; Roraima: <Amazonas>; Pará: <Amazonas,Roraima>; Amapá: <Pará>; e Tocantins: <Pará>. Esses subgrafos dariam origem às árvores necessárias para a modelagem, e como há a restrição $j < i$ para todo $j \in N(i)$, para $i = 2, \dots, m$, então \mathbf{B} permanece sendo triangular inferior.

Partindo desses *insights*, definimos

$$b_{ij} = \frac{\rho}{1 + (n_{<i} - 1)\rho^2} \quad (i = 2, \dots, m; j \in N(i)), \quad \tau_i = \frac{1 + (n_{<i} - 1)\rho^2}{1 - \rho^2} \quad (i = 1, \dots, m), \quad (5.3)$$

em que $0 < \rho < 1$. Pelo fato dos τ_i 's serem positivos, então a distribuição de $\boldsymbol{\phi} \sim N_m(\mathbf{0}, \mathbf{L}^\top \mathbf{F} \mathbf{L})$ será própria. No caso limite $\rho = 1$, temos que $b_i = 1/n_{<i}$ e $\tau_i \propto n_{<i}$ (DATTA *et al.*, 2019). Por fim, considerando um direcionamento arbitrário π , a generalização de 5.3, em termos de distribuições condicionais, pode ser dada da seguinte forma:

$$\phi_i | \phi_{<i, \pi} \sim N \left(\frac{\rho}{1 + (n_{\pi(i)} - 1)\rho^2} \sum_{j \in N(i)} \phi_j, \frac{1 + (n_{\pi(i)} - 1)\rho^2}{1 - \rho^2} \right), \quad (5.4)$$

e a distribuição conjunta dos efeitos aleatórios espaciais resultante,

$$\boldsymbol{\phi} \sim N_m \left(\mathbf{0}, (\mathbf{I}_m - \mathbf{B})_\pi^\top \mathbf{F}_\pi (\mathbf{I}_m - \mathbf{B})_\pi \right). \quad (5.5)$$

Feita a apresentação do modelo DAGAR, a partir de 5.5, podemos especificar o modelo de regressão binária espacial generalizado da seguinte forma:

$$\begin{aligned} p_{si} &= F_\delta(\mathbf{x}_s^\top \boldsymbol{\beta} + \phi_{si}), \quad i = 1, \dots, m, \quad s = 1, \dots, r \\ \boldsymbol{\beta} &\sim p(\boldsymbol{\beta}), \quad \delta \sim p(\delta) \\ \boldsymbol{\phi}_s &\sim N_m(\mathbf{0}, (\mathbf{I} - \mathbf{B})_\pi^\top \mathbf{F}_\pi (\mathbf{I} - \mathbf{B})_\pi) \quad s = 1, \dots, r, \end{aligned} \quad (5.6)$$

em que $\mathbf{B} = (b_{ij})$, para $i, j = 1, \dots, m$, e $\mathbf{F} = \text{diag}(\tau_1, \dots, \tau_m)$, os quais são especificados como em 5.3. As distribuições a posteriori dos parâmetros são idênticas às apresentadas na Seção 4.2, com excessão dos efeitos aleatórios espaciais $\boldsymbol{\phi}_s$, para $s = 1, \dots, r$, cujas distribuições a posteriori sofrem uma leve modificação:

$$\begin{aligned} p(\boldsymbol{\phi}_s | \boldsymbol{\beta}, \delta, \Omega, \mathbf{y}_s, \mathbf{x}_s) &\propto p(\mathbf{y}_s | \mathbf{x}_s, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_s) p(\boldsymbol{\phi}_s | (\mathbf{I} - \mathbf{B})_\pi^\top \mathbf{F}_\pi (\mathbf{I} - \mathbf{B})_\pi) \\ &\propto \prod_{i=1}^m [F_\delta(\mathbf{x}_s^\top \boldsymbol{\beta} + \phi_{si})]^{y_{si}} \prod_{i=1}^m [1 - F_\delta(\mathbf{x}_s^\top \boldsymbol{\beta} + \phi_{si})]^{1-y_{si}} \times \\ &\quad \exp \left\{ -\frac{1}{2} \boldsymbol{\phi}_s^\top [(\mathbf{I} - \mathbf{B})_\pi^\top \mathbf{F}_\pi (\mathbf{I} - \mathbf{B})_\pi] \boldsymbol{\phi}_s \right\}. \end{aligned} \quad (5.7)$$

5.2 Estudo de Simulação

O estudo de simulação desta Seção busca avaliar o desempenho do modelo apresentado em 5.6. Nesse estudo, estaremos utilizando este modelo sob função de ligação C-P, especificado sob distribuição a priori G-Wishart, CAR e DAGAR, totalizando três modelos espaciais distintos. Nosso intuito é avaliar a robustez desses modelos de acordo com a distribuição a priori espacial utilizada.

A primeira parte do estudo consiste em avaliar a capacidade do modelo em recuperar o verdadeiro valor dos parâmetros. Realizamos as simulações para $r = 100$, $m = 60$ unidades de área, duas covariáveis geradas independentemente de uma distribuição Normal padrão, considerando $\lambda = 0,5$ e $\lambda = 4$, e $\beta_1 = -0.7$ e $\beta_2 = 0.7$. Para cada combinação de valor de λ , realizamos 100 replicações de conjuntos de dados.

No segundo cenário, simulamos conjuntos de dados multivariados dos três modelos candidatos. Consideramos tamanhos de amostras $r = 100$ e $r = 200$, $m = 60$ unidades de área, e valores do parâmetro de forma como sendo $\lambda = 0,5$ e $\lambda = 4$. Duas covariáveis, geradas independentemente da distribuição Normal padrão, foram consideradas para os três modelos, e os valores dos coeficientes de regressão foram fixados em $\beta_1 = 0.7$ e $\beta_2 = -0.7$. Para gerar a estrutura espacial, utilizamos uma submatriz 60×60 da matriz de adjacência dos dados de periodontite. No caso em que $r = 100$, de cada modelo, para cada combinação de valor de λ , geramos 100 replicações de conjuntos de dados; quando $r = 200$, geramos 50 replicações. Os três modelos candidatos foram então ajustados para cada réplica simulada. No caso do modelo DAGAR, como comentado antes, torna-se dependente da ordem das observações. Em nosso estudo de simulação, consideramos a ordem natural das observações geradas. Para cada um dos cenários de simulação apresentados, extraímos 5000 amostras MCMC descartando as primeiras 2500 como *burn-in*. Aqui, novamente utilizaremos uma submatriz de dimensão $m \times m$ extraída da matriz de adjacência dos dados motivacionais (Seção 4.1) para gerar a estrutura de vizinhança das sub-regiões.

Tabela 23 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para o modelo C-P, baseado em 100 replicações para o tamanho de amostras $r = 100$. 5000 iterações, descartando 2500 como *burn-in*.

		$r = 100$					
Ligação	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	
C-P	$\lambda = 0,5$	β_1	0.039	0.074	-0.102	0.87	
		β_2	0.043	0.071	0.098	0.85	934 (425.07)
		δ	0.02	0.036	-0.051	0.87	
	$\lambda = 4$	β_1	0.023	0.065	-0.09	0.74	
		β_2	0.035	0.069	0.095	0.72	1191.94 (548.03)
		δ	0.004	0.04	0.029	0.88	

Fonte: Elaborada pelo autor.

De acordo com as simulações iniciais para calibrar os hiperparâmetros do HMC para o modelo DAGAR, adotamos a seguinte configuração: o valor do tamanho do passo para $\epsilon = 0.1$ e quantidade de passos $L = 22$, para o bloco de parâmetros (β, δ) ; $\epsilon = 0.1$ e $L = 22$ para ϕ_s , para $s = 1, \dots, r$. O modelo sob priori CAR: $\epsilon = 0.1$ e $L = 22$ para o bloco de parâmetros (β, δ) ; $\epsilon = 0.05$ e $L = 22$ para ϕ_s , para $s = 1, \dots, r$. No tocante ao modelo sob distribuição a priori G-Wishart, as configurações são as mesmas que as adotadas na Seção 4.3.

Na Tabela 23, temos o resultado do estudo de simulação do primeiro cenário. A Tabela mostra bom desempenho do modelo C-P sob distribuição a priori DAGAR, em termos de viés relativo, RMSE,

NRMSE e probabilidade de cobertura. No entanto, cabe destacar que para o caso em que $\lambda = 4$, os valores de probabilidade de cobertura dos parâmetros β_1 e β_2 são menores quando comparados ao caso em que $\lambda = 0, 5$.

Na segunda parte do estudo de simulação, os resultados são mostrados nas Tabelas 24 e 25 . Em ambos os cenários de tamanhos amostrais $r = 100$ e 200 , percebe-se que os modelos sob priori paramétrica CAR e DAGAR apresentam os melhores resultados, tanto em termos de qualidade de ajuste quanto em termos de qualidade de predição para qualquer valor de λ considerado, seja para $\lambda = 0, 5$ ou $\lambda = 4$, com destaque para o modelo sob distribuição a priori CAR, que apresentou melhor desempenho em quase todos os cenários considerados, exceto pela medida AUC, que foi praticamente igual para todos os modelos considerados.

5.3 Discussões Finais

Neste Capítulo, buscamos apresentar um estudo de simulação do modelo proposto no Capítulo 4 sob diferentes especificações espaciais, incluindo o modelo DAGAR, utilizando apenas a função de ligação C-P com o intuito de avaliar a robustez do modelo proposto sob especificações espaciais distintas.

Pelos resultados, constata-se que o modelo sob distribuição a priori DAGAR consegue recuperar bem os verdadeiros valores dos parâmetros, considerando amostras de tamanho $r = 100$, apresentando bons desempenhos das métricas consideradas, principalmente para o cenário em que $\lambda = 0, 5$. No entanto, quando $\lambda = 4$, o desempenho da probabilidade de cobertura para β_1 e β_2 diminui. Na segunda parte do estudo de simulação, constata-se que o modelo sob priori CAR obteve os melhores resultados em praticamente todas as métricas de comparação adotadas, em todos os cenários considerados.

Isso mostra o poder e utilidade do modelo CAR em termos de aplicação, apesar de ser o modelo mais antigo dos três modelos espaciais. No entanto, estudos de simulação mais abrangentes necessitam ser desenvolvidos. Uma ideia seria aumentar o número de sub-regiões na estrutura espacial considerada, uma vez que constatamos, para o conjunto de dados motivacionais de periodontite cuja matriz de adjacência possuía 168 localizações, que os modelos sob distribuição a priori G-Wishart apresentaram desempenho bem melhor que o modelo CAR. Uma outra ideia seria tentar ajustar os modelos sob priori DAGAR, considerando ordenações diferentes das observações para poder avaliar o impacto nos resultados finais.

Tabela 24 – Comparação do modelo C-P sob os modelos espaciais G-Wishart, CAR e DAGAR para diferentes conjuntos de dados simulados de tamanho $r = 100$.

modelo		$\lambda = 0,5$			$\lambda = 4$		
verdadeiro	ajustado	G-Wishart	CAR	DAGAR	G-Wishart	CAR	DAGAR
	G-Wishart	DIC	5967.06	5484.00	5561.03	1936.77	886.75
WAIC		6679.18	6019.54	6302.71	3476.49	2702.93	3050.11
ml		2264.31	2079.27	2110.95	1102.09	929.50	1009.00
LPML		-3339.59	-3009.77	-3151.35	-1738.24	-1351.47	-1525.05
ACC		0.76	0.81	0.78	0.76	0.81	0.78
AUC		0.50	0.50	0.49	0.50	0.50	0.49
CSI		0.51	0.59	0.54	0.51	0.59	0.54
F1		0.66	0.73	0.69	0.66	0.73	0.69
FAITH		0.57	0.59	0.58	0.57	0.59	0.58
KAPPA		0.25	0.46	0.34	0.25	0.46	0.33
MCC		0.29	0.48	0.37	0.29	0.48	0.37
SSI		0.37	0.43	0.39	0.36	0.43	0.38
CAR		DIC	6075.07	5679.03	5812.67	2763.68	2132.51
	WAIC	6631.11	6051.69	6366.18	3777.31	2935.43	3420.61
	ml	2266.27	2127.38	2160.84	1217.74	1021.05	1131.92
	LPML	-3315.55	-3025.84	-3183.09	-1888.65	-1467.71	-1710.30
	ACC	0.75	0.82	0.77	0.75	0.82	0.78
	AUC	0.52	0.53	0.51	0.53	0.53	0.52
	CSI	0.52	0.62	0.55	0.52	0.63	0.55
	F1	0.67	0.76	0.70	0.67	0.77	0.70
	FAITH	0.57	0.60	0.57	0.57	0.60	0.58
	KAPPA	0.28	0.50	0.35	0.28	0.50	0.35
	MCC	0.32	0.52	0.38	0.32	0.52	0.38
	SSI	0.37	0.46	0.39	0.37	0.46	0.39
	DAGAR	DIC	5950.67	5659.09	5746.35	2675.89	2274.52
WAIC		6517.41	6008.70	6272.69	3775.64	3020.98	3411.36
ml		2221.68	2121.45	2144.84	1232.72	1059.99	1147.21
LPML		-3258.70	-3004.35	-3136.35	-1887.82	-1510.49	-1705.68
ACC		0.76	0.82	0.78	0.76	0.83	0.78
AUC		0.53	0.55	0.53	0.53	0.54	0.52
CSI		0.54	0.64	0.57	0.54	0.64	0.57
F1		0.69	0.78	0.72	0.69	0.78	0.71
FAITH		0.57	0.60	0.58	0.57	0.60	0.58
KAPPA		0.31	0.51	0.38	0.31	0.53	0.39
MCC		0.34	0.53	0.41	0.35	0.54	0.42
SSI		0.38	0.47	0.41	0.38	0.48	0.41

Fonte: Elaborada pelo autor.

Tabela 25 – Comparação do modelo C-P sob os modelos espaciais G-Wishart, CAR e DAGAR para diferentes conjuntos de dados simulados de tamanho $r = 200$.

modelo		$\lambda = 0,5$			$\lambda = 4$		
verdadeiro	ajustado	G-Wishart	CAR	DAGAR	G-Wishart	CAR	DAGAR
	G-Wishart	DIC	11840.20	10907.91	11069.53	3716.26	1723.22
WAIC		13321.33	12007.60	12582.42	6849.34	5332.80	6012.02
ml		4517.52	4149.40	4214.38	2172.28	1834.78	1990.39
LPML		-6660.66	-6003.80	-6291.21	-3424.67	-2666.40	-3006.01
ACC		0.76	0.81	0.78	0.76	0.81	0.78
AUC		0.50	0.50	0.49	0.51	0.50	0.49
CSI		0.51	0.58	0.54	0.51	0.59	0.54
F1		0.66	0.73	0.68	0.66	0.73	0.69
FAITH		0.57	0.59	0.57	0.57	0.59	0.58
KAPPA		0.25	0.45	0.34	0.26	0.46	0.34
MCC		0.29	0.47	0.37	0.30	0.48	0.37
SSI		0.36	0.42	0.38	0.37	0.43	0.39
CAR	DIC	11870.98	11272.28	11551.67	5008.25	4297.18	4473.05
	WAIC	13018.59	12017.05	12648.35	7310.81	5902.64	6858.71
	ml	4431.14	4220.23	4288.14	2382.45	2059.51	2279.59
	LPML	-6509.29	-6008.52	-6324.17	-3655.40	-2951.32	-3429.35
	ACC	0.76	0.82	0.77	0.76	0.82	0.77
	AUC	0.52	0.54	0.52	0.52	0.53	0.52
	CSI	0.54	0.62	0.55	0.54	0.62	0.55
	F1	0.69	0.76	0.70	0.68	0.76	0.70
	FAITH	0.57	0.60	0.58	0.57	0.60	0.58
	KAPPA	0.30	0.49	0.36	0.31	0.50	0.35
	MCC	0.34	0.51	0.39	0.34	0.52	0.38
	SSI	0.38	0.46	0.39	0.38	0.46	0.40
DAGAR	DIC	11694.05	11237.49	11429.12	4868.75	4418.33	4502.46
	WAIC	12864.09	11955.55	12480.54	7208.25	5910.59	6686.84
	ml	4370.88	4217.75	4263.95	2366.79	2072.06	2245.33
	LPML	-6432.05	-5977.77	-6240.27	-3604.12	-2955.29	-3343.42
	ACC	0.77	0.83	0.79	0.76	0.82	0.78
	AUC	0.52	0.54	0.52	0.52	0.54	0.52
	CSI	0.54	0.64	0.57	0.54	0.63	0.57
	F1	0.69	0.77	0.71	0.69	0.77	0.71
	FAITH	0.57	0.60	0.58	0.57	0.60	0.58
	KAPPA	0.33	0.52	0.39	0.33	0.51	0.39
	MCC	0.36	0.53	0.41	0.36	0.53	0.42
	SSI	0.39	0.47	0.41	0.39	0.47	0.41

Fonte: Elaborada pelo autor.

CONCLUSÕES E TRABALHOS FUTUROS

Nesta tese, foram estudados modelos de regressão binária espacial para dados desbalanceados. Os modelos apresentados possuem algumas distinções: o primeiro deles pode ser visto como sendo o modelo de regressão binária espacial usual presente na literatura de dados de área (que abreviaremos por RBES - Regressão binária espacial simples), e o segundo foi motivado por um conjunto de dados de periodontite (Seção 4.1), e que também pode ser considerado uma generalização do primeiro modelo (que abreviaremos de RBEG - Regressão binária espacial generalizada). Para ambos os modelos, utilizamos uma classe de funções de ligação flexível, as chamadas funções de ligação potência e reversa de potência. Estas possuem um parâmetro de forma que permitem que os modelos se ajustem de forma conveniente à assimetria presente nos dados, e são excelentes alternativas às funções de ligação Logito e Probit, que são consideradas simétricas, e ao modelo Cloglog, que apesar de ser uma função de ligação assimétrica, é uma assimetria fixa. No tocante à estrutura espacial, avaliamos o desempenho desses modelos sob uma distribuição a priori diferente, a saber, a distribuição G-Wishart. Comparamos seu desempenho com o do modelo CAR em aplicações voltadas a conjuntos de dados reais. Ainda oferecemos um estudo de simulação preliminar para o modelo espacial DAGAR.

Em relação à metodologia inferencial aplicada aos modelos, seguimos a linha bayesiana, e desenvolvemos métodos de amostragem para os parâmetros dos modelos que fossem eficientes em termos computacionais, garantindo rapidez no processo de estimação. Produzimos algoritmos baseados nos métodos MCMC, e que levassem em consideração a esparsidade da matriz de precisão Ω dos efeitos aleatórios espaciais para garantir escalabilidade em grandes conjuntos de dados. Os algoritmos foram escritos em C++ por meio do pacote Rcpp, utilizando o software R como interface. Para o modelo generalizado (Capítulo 4), disponibilizamos implementações R/Rcpp no repositório GitHub: <<https://github.com/alan-assuncao/SpatialBinReg>>.

Pelos estudos de simulação, os resultados apontaram que para ambos os modelos (RBES e RBEG), sob as funções de ligação flexíveis adotadas, os métodos de amostragem propostos apresentaram bom desempenho em termos de recuperação do verdadeiro valor dos parâmetros e cenário de má especificação em todas as situações consideradas. No tocante à parte das aplicações aos dados reais, para o modelo RBES, os ajustes realizados com funções de ligação flexíveis apresentaram resultados melhores em termos

de qualidade de ajuste quando comparados aos modelos com funções de ligação Logito, Probit e Cloglog. Quando comparamos os modelos com funções de ligação flexíveis ajustados sob distribuição a priori G-Wishart e distribuição a priori CAR, os resultados mostraram que os ajustes realizados sob distribuição a priori CAR apresentam melhor desempenho, considerando a qualidade de ajuste e qualidade da predição. Para o modelo RBEG, os ajustes com funções de ligação flexíveis apresentaram resultados melhores em termos de qualidade de ajuste quando comparados aos modelos tradicionais (funções de ligação Logito, Probit e Cloglog). Ao compararmos os desempenhos dos modelos sob distribuições a priori G-Wishart e CAR, os resultados em termos de qualidade de ajuste mostraram que os modelos sob distribuição a priori G-Wishart apresentam melhor desempenho em todas as métricas consideradas, exceto na medida L, na qual os modelos ajustados com priori CAR apresentaram os melhores resultados. Quando avaliamos sob critérios de qualidade da predição, os resultados apontam que os ajustes com distribuição a priori G-Wishart são melhores para a maioria das métricas consideradas.

Neste trabalho, desenvolvemos ferramentas de diagnóstico bayesianas para detecção de *outliers*, usando a divergência de Bregman. Quando a implementamos para os modelos RBES e RBEG, os resultados apontam que os modelos com funções de ligação flexíveis foram os que identificaram um maior número de possíveis *outliers*. Indiretamente, apesar de não serem métricas de comparação entre modelos, os resultados dessas análises de diagnóstico de observações influentes evidenciam aquilo que a priori era apenas uma hipótese, a saber, que assumindo que modelos com assimetria flexível se ajustam melhor a dados com a presença de desbalanceamento, é de se esperar que eles sejam mais sensíveis a observações discrepantes.

O modelo espacial DAGAR é recente na literatura, sendo uma forma alternativa de especificação da matriz de precisão Ω dos efeitos aleatórios espaciais. Por meio do estudo de simulação preliminar realizado com foco nesse modelo, constata-se que o modelo sob distribuição a priori DAGAR consegue recuperar bem os verdadeiros valores dos parâmetros, considerando amostras de tamanho $r = 100$. No entanto, quando $\lambda = 4$, o desempenho da probabilidade de cobertura diminui para os coeficientes de regressão considerados. Na segunda parte do estudo de simulação, em que avaliamos o modelo RBEG sob distribuições a priori G-Wishart, CAR e DAGAR, constatou-se que o modelo sob priori CAR obteve os melhores resultados em praticamente todas as métricas de comparação adotadas, e em todos os cenários considerados.

Retomando os objetivos iniciais desta tese, constatamos que estes foram de fato atingidos, e as contribuições a priori mencionadas também foram efetivadas. No entanto, para esse campo de estudo de modelos de regressão binária com funções de ligação assimétrica e modelagem espacial, propostas de trabalho e desenvolvimentos futuros são perfeitamente possíveis.

6.1 Produções

6.1.1 Trabalhos Apresentados em Eventos

- Alan S. Assunção, Ricardo S. Ehlers. *Modelos de regressão binária espacial bayesiano para dados desbalanceados*. In: SEMANA DA ESTATÍSTICA DA UFPI, 10, 2024, Teresina: UFPI. Palestra.

- Alan S. Assunção, Ricardo S. Ehlers. *Modelos de regressão binária espacial bayesiana para dados desbalanceados*. In: 10th Workshop on Probabilistic and Statistical Methods, 10, 2024, São Carlos: UFSCar/USP. Pôster.

6.2 Propostas de Trabalhos Futuros

Com relação aos resultados obtidos nesta tese, parte das análises apresentadas no Capítulo 4 foi colocada em um artigo e está prestes a ser submetida a um periódico. O artigo referido é

- Alan S. Assunção, Ricardo S. Ehlers, Indranil Sahoo, Dipankar Bandyopadhyay (2025). Bayesian inference under sparsity for spatial binary regression models.

Dito isso, antes de finalizar este Capítulo, algumas sugestões de trabalhos futuros podem ser mencionadas:

1. primeiro, poderíamos tentar propor um esquema de amostragem para o κ da distribuição G-Wishart, e tentar estimá-lo a partir dos dados, e posteriormente verificar o desempenho dos modelos RBES e RBEG com essa nova configuração;
2. uma outra ideia seria propor uma distribuição a priori complexa penalizada para λ , o parâmetro de forma da classe de funções de ligação potência e reversa de potência, conforme proposto em [Ordoñez et al. \(2024\)](#). Nesse sentido, já temos iniciado a construção de um artigo tentando propor uma distribuição a priori complexa penalizada para o parâmetro de forma $\lambda \in (-1, 1)$ da classe de funções de ligação assimétricas transmutadas ([SHAW; BUCKLEY, 2009](#)), exploradas no Apêndice D (Seção D.2). Para essa classe de funções de ligação, já dispomos de um estudo de simulação (Apêndice D, Seção D.2), e uma aplicação aos dados de periodontite (Apêndice E, Seção E.2). Em ambas as situações, os modelos assimétricos transmutados desempenharam bem, sendo melhores que os modelos Logito, Probit e Cloglog na parte das aplicações em dados reais, indicando que essa classe de funções de ligação pode ser bem atrativa para desenvolvimentos futuros;
3. poderíamos tentar usar outras funções de ligação flexíveis para avaliar o desempenho dos modelos, como a função de ligação GEV ([WANG; DEY, 2011](#));
4. no tocante ao modelo DAGAR, pretendemos apresentar futuramente um estudo de simulação mais abrangente e uma análise de aplicação a dados reais, bem como propor um esquema de amostragem para os parâmetros do modelo DAGAR, a saber, os parâmetros ρ e τ (que são inseridos pelos autores do modelo DAGAR [Datta et al. \(2019\)](#)), baseada no método de Monte Carlo Hamiltoniano. Para isso, já contamos com os gradientes das log-posterioris desses parâmetros no Apêndice F.
5. Por fim, poderíamos ainda tentar comparar os desempenhos dos modelos flexíveis aqui apresentados, com algum outro modelo de classificação que possa haver no contexto espacial de dados de área.

REFERÊNCIAS

- ALLAIRE, J.; EDDERBUETTEL, D.; FRANÇOIS, R. Rcpp Attributes. **Vignette Included in R Package Rcpp**, URL <http://CRAN.R-Project.org/package=Rcpp>, 2018. Citado na página 132.
- ALVARES, C. A.; STAPE, J.; SENTELHAS, P.; GONÇALVES, J.; SPAROVEK, G. Köppen's Climate Classification Map for Brazil. **Meteorologische Zeitschrift**, v. 22, 12 2013. Citado nas páginas 34 e 80.
- 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, 2023. Citado nas páginas 36, 48, 49, 50, 73, 144 e 153.
- ANSELIN, L. Under the Hood Issues in the Specification and Interpretation of Spatial Regression Models. **Agricultural Economics**, Wiley Online Library, v. 27, n. 3, p. 247–267, 2002. Citado nas páginas 69, 71, 151 e 152.
- ARANDA-ORDAZ, F. J. On two Families of Transformations to Additivity for Binary Response Data. **Biometrika**, Oxford University Press, v. 68, n. 2, p. 357–363, 1981. Citado nas páginas 35 e 36.
- ASL, M. G.; RASHIDI, M. M.; ABAD, S. A. H. E. Emerging digital economy companies and leading cryptocurrencies: insights from blockchain-based technology companies. **Journal of Enterprise Information Management**, v. 34, n. 5, p. 1506–1550, 09 2021. ISSN 1741-0398. Disponível em: <<https://doi.org/10.1108/JEIM-08-2020-0348>>. Citado na página 77.
- ASSUNÇÃO, G. O.; IZBICKI, R.; PRATES, M. O. Is augmentation effective in improving prediction in imbalanced datasets? **Journal of Data Science**, School of Statistics, Renmin University of China, p. 1–16, 2024. ISSN 1680-743X. Citado nas páginas 64, 86 e 156.
- BANDYOPADHYAY, D.; CANALE, A. Non-Parametric Spatial Models for Clustered Ordered Periodontal Data. **Journal of the Royal Statistical Society Series C: Applied Statistics**, Oxford University Press, v. 65, n. 4, p. 619–640, 2016. Citado nas páginas 72, 73, 93, 94 e 95.
- BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. **Hierarchical Modeling and Analysis for Spatial Data**. [S.l.]: Chapman and Hall/CRC, 2003. Citado nas páginas 34, 67, 68, 71, 72 e 151.
- 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 36, 38 e 73.
- BAZÁN, J. L.; BOLFARINE, H.; BRANCO, M. D. A Framework for Skew-Probit Links in Binary Regression. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 39, n. 4, p. 678–697, 2010. Citado nas páginas 35, 36 e 53.
- BAZAN, J. L.; ROMEO, J. S.; RODRIGUES, J. *et al.* Bayesian Skew-Probit Regression for Binary Response Data. **Brazilian Journal of Probability and Statistics**, v. 28, n. 4, p. 467, 2014. Citado na página 139.
- BEAM, A. L.; GHOSH, S. K.; DOYLE, J. Fast Hamiltonian Monte Carlo using Gpu Computing. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 25, n. 2, p. 536–548, 2016. Citado na página 47.

- BESAG, J. Spatial Interaction and the Statistical Analysis of Lattice Systems. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 36, n. 2, p. 192–225, 1974. Citado nas páginas 70 e 71.
- BETANCOURT, M. A Conceptual Introduction to Hamiltonian Monte Carlo. **ArXiv Preprint ArXiv:1701.02434**, 2017. Citado nas páginas 36 e 45.
- BROWN, L. J.; OLIVER, R.; LOE, H. Evaluating Periodontal Status of US Employed Adults. **Journal of the American Dental Association (1939)**, v. 121, n. 2, p. 226–232, 1990. Citado na página 93.
- 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 35 e 36.
- CHOI, S.-S.; CHA, S.-H.; TAPPERT, C. C. *et al.* A Survey of Binary Similarity and Distance Measures. **Journal of Systemics, Cybernetics and Informatics**, v. 8, n. 1, p. 43–48, 2010. Citado na página 64.
- COLLETT, D. **Modelling Binary Data**. [S.l.]: CRC press, 2002. Citado nas páginas 33, 36 e 53.
- CRESSIE, N. **Statistics for Spatial Data**. [S.l.]: John Wiley & Sons, 1993. Citado nas páginas 37, 67, 68, 69, 70 e 152.
- CZADO, C.; SANTNER, T. J. The Effect of Link Misspecification on Binary Regression Inference. **Journal of Statistical Planning and Inference**, Elsevier, v. 33, n. 2, p. 213–231, 1992. Citado na página 35.
- DANILEVICZ, I. M.; EHLERS, R. S. Bayesian Influence Diagnostics Using Normalized Functional Bregman Divergence. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 51, n. 6, p. 1637–1652, 2022. Citado nas páginas 26, 39, 62 e 84.
- DATTA, A.; BANERJEE, S.; HODGES, J. S.; GAO, L. Spatial Disease Mapping using Directed Acyclic Graph Auto-Regressive (DAGAR) Models. **Bayesian Analysis**, NIH Public Access, v. 14, n. 4, p. 1221, 2019. Citado nas páginas 38, 113, 114, 115, 116, 123 e 167.
- DUA, D.; GRAFF, C. **UCI Machine Learning Repository**. 2017. Disponível em: <<http://archive.ics.uci.edu/ml>>. Citado na página 55.
- DUANE, S.; KENNEDY, A. D.; PENDLETON, B. J.; ROWETH, D. Hybrid Monte Carlo. **Physics Letters B**, Elsevier, v. 195, n. 2, p. 216–222, 1987. Citado nas páginas 36 e 45.
- EDDELBUETTEL, D.; BALAMUTA, J. J. Extending R with C++: a brief introduction to Rcpp. **The American Statistician**, Taylor & Francis, v. 72, n. 1, p. 28–36, 2018. Citado na página 132.
- EDDELBUETTEL, D.; FRANÇOIS, R. Rcpp: Seamless R and C++ Integration. **Journal of Statistical Software**, v. 40, p. 1–18, 2011. Citado na página 131.
- _____. Rcpp Syntactic Sugar. **Vignette Included in R Package Rcpp**, Citeseer, 2017. Citado na página 133.
- GAMERMAN, D.; LOPES, H. F. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. [S.l.]: CRC press, 2006. Citado nas páginas 42, 43 e 45.
- GEISSER, S.; EDDY, W. F. A predictive approach to model selection. **Journal of the American Statistical Association**, Taylor & Francis, v. 74, n. 365, p. 153–160, 1979. Citado na página 77.
- GELFAND, A. E.; GHOSH, S. K. Model Choice: a Minimum Posterior Predictive Loss Approach. **Biometrika**, Oxford University Press, v. 85, n. 1, p. 1–11, 1998. Citado na página 59.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian Data Analysis. Texts in Statistical Science**. [S.l.]: Chapman & Hall London, 2004. Citado na página 42.

GIROLAMI, M.; CALDERHEAD, B. Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. **Journal of The Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 73, n. 2, p. 123–214, 2011. Citado na página 47.

GOH, G.; DEY, D. K. Bayesian Model Diagnostics Using Functional Bregman Divergence. **Journal of Multivariate Analysis**, Elsevier, v. 124, p. 371–383, 2014. Citado nas páginas 39, 60 e 61.

GUERRERO, V. M.; JOHNSON, R. A. Use of the Box-Cox Transformation with Binary Response Models. **Biometrika**, Oxford University Press, v. 69, n. 2, p. 309–314, 1982. Citado nas páginas 35 e 36.

HARTMANN, M.; EHLERS, R. S. Bayesian Inference for Generalized Extreme Value Distributions via Hamiltonian Monte Carlo. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 46, n. 7, p. 5285–5302, 2017. Citado na página 48.

HOFFMAN, M. D.; GELMAN, A. *et al.* The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. **Journal of Machine Learning Research**, v. 15, n. 1, p. 1593–1623, 2014. Citado na página 36.

HOSSIN, M.; SULAIMAN, M. N. A Review on Evaluation Metrics for Data Classification Evaluations. **International journal of Data Mining & Knowledge Management Process**, Academy & Industry Research Collaboration Center (AIRCC), v. 5, n. 2, p. 1, 2015. Citado na página 63.

HUAYANAY, A. de la C.; BAZÁN, 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 35, 36, 48 e 55.

HUAYANAY, A. de la C.; BAZÁN, J. L.; RUSSO, C. M. Performance of Evaluation Metrics for Classification in Imbalanced Data. **Computational Statistics**, Springer, v. 40, n. 3, p. 1447–1473, 2025. Citado nas páginas 48, 49, 50, 52, 64 e 156.

HUGOSON, A.; THORSTENSSON, H.; FALTT, H.; KUYLENSTIERNA, J. Periodontal Conditions in Insulin-Dependent Diabetics. **Journal of Clinical Periodontology**, Wiley Online Library, v. 16, n. 4, p. 215–223, 1989. Citado na página 93.

JABER, A.; JUHÁSZ, J.; CSONKA, B. An Analysis of Factors Affecting the Severity of Cycling Crashes Using Binary Regression Model. **Sustainability**, MDPI, v. 13, n. 12, p. 6945, 2021. Citado na página 33.

KIM, S.; CHEN, M.-H.; DEY, D. K. Flexible Generalized t-link Models for Binary Response Data. **Biometrika**, Oxford University Press, v. 95, n. 1, p. 93–106, 2008. Citado na página 36.

KING, G.; ZENG, L. Logistic Regression in Rare Events Data. **Political Analysis**, Cambridge University Press, v. 9, n. 2, p. 137–163, 2001. Citado na página 55.

LEMONTE, A. J.; BAZÁN, J. L. New Links for Binary Regression: an Application to Coca Cultivation in Peru. **Test**, Springer, v. 27, p. 597–617, 2018. Citado nas páginas 36, 38 e 49.

LI, D.; WANG, X.; LIN, L.; DEY, D. K. Flexible Link Functions in Nonparametric Binary Regression with Gaussian Process Priors. **Biometrics**, Wiley Online Library, v. 72, n. 3, p. 707–719, 2016. Citado nas páginas 33 e 36.

MA, Y.; ZHU, X.; HSU, J. Data Poisoning Against Differentially-Private Learners: Attacks and Defenses. **ArXiv Preprint ArXiv:1903.09860**, 2019. Citado na página 55.

MOHAMMADI, A.; WIT, E. Bayesian Structure Learning in Sparse Gaussian Graphical Models. **Bayesian Analysis**, v. 10, n. 1, p. 109–138, 2015. Citado nas páginas 21, 22, 23, 74, 87, 90, 97, 109 e 111.

- MOHAMMADI, R.; WIT, E. C. BDgraph: An R Package for Bayesian Structure Learning in Graphical Models. **Journal of Statistical Software**, v. 89, n. 3, p. 1–30, 2019. Citado na página 74.
- MORGAN, B. J. Observations on Quantitative Analysis. **Biometrics**, JSTOR, p. 879–886, 1983. Citado na página 35.
- MORRIS, M.; WHEELER-MARTIN, K.; SIMPSON, D.; MOONEY, S. J.; GELMAN, A.; DIMAGGIO, C. Bayesian Hierarchical Spatial Models: Implementing the Besag York Mollié model in stan. **Spatial and Spatio-Temporal Epidemiology**, Elsevier, v. 31, p. 100301, 2019. Citado nas páginas 37 e 71.
- NEAL, R. M. MCMC Using Hamiltonian Dynamics. **Handbook of Markov Chain Monte Carlo**, Chapman and Hall/CRC, v. 2, n. 11, p. 2, 2010. Citado nas páginas 45 e 46.
- NEGRATO, C. A.; TARZIA, O.; JOVANOVIČ, L.; CHINELLATO, L. E. M. Periodontal Disease and Diabetes Mellitus. **Journal of Applied Oral Science**, SciELO Brasil, v. 21, p. 1–12, 2013. Citado na página 93.
- ORDOÑEZ, J. A.; PRATES, M. O.; BAZÁN, J. L.; LACHOS, V. H. Penalized Complexity Priors for the Skewness Parameter of Power Links. **Canadian Journal of Statistics**, Wiley Online Library, v. 52, n. 1, p. 98–117, 2024. Citado na página 123.
- PRENTICE, R. L. A Generalization of the Probit and Logit Methods for Dose Response Curves. **Biometrics**, JSTOR, p. 761–768, 1976. Citado na página 35.
- REIS, E. A.; GAMERMAN, D.; PAEZ, M. S.; MARTINS, T. G. Bayesian Dynamic Models for Space–Time Point Processes. **Computational Statistics & Data Analysis**, Elsevier, v. 60, p. 146–156, 2013. Citado nas páginas 67 e 68.
- ROSSONI, D. F. *et al.* Autoregressive Analysis of Variance for Experiments with Spatial Dependence Between Plots: A Simulation Study. **Brazilian Journal of Biometrics**, v. 37, n. 2, p. 244–257, 2019. Citado na página 152.
- ROVERATO, A. Hyper Inverse Wishart Distribution for Non-Decomposable Graphs and its Application to Bayesian Inference for Gaussian Graphical Models. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 29, n. 3, p. 391–411, 2002. Citado nas páginas 37, 38 e 72.
- SANTOS, B.; BOLFARINE, H. On Bayesian Quantile Regression and Outliers. **ArXiv Preprint ArXiv:1601.07344**, 2016. Citado nas páginas 61 e 62.
- SHAW, W. T.; BUCKLEY, I. R. The Alchemy of Probability Distributions: Beyond Gram-Charlier Expansions, and a Skew-Kurtotic-Normal Distribution from a Rank Transmutation Map. **ArXiv Preprint ArXiv:0901.0434**, 2009. Citado nas páginas 123 e 153.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian Measures of Model Complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002. Citado na página 58.
- STUKEL, T. A. Generalized Logistic Models. **Journal of the American Statistical Association**, Taylor & Francis, v. 83, n. 402, p. 426–431, 1988. Citado na página 35.
- TEAM, S. D. *et al.* **Stan modeling language user’s guide and reference manual, version 2.17.0**. [S.l.]: Stan Development Team, 2018. Citado na página 74.
- VEHTARI, A.; GELMAN, A.; GABRY, J. Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC. **Statistics and Computing**, Springer, v. 27, p. 1413–1432, 2017. Citado na página 57.

VUJOVIĆ, Ž. *et al.* Classification Model Evaluation Metrics. **International Journal of Advanced Computer Science and Applications**, v. 12, n. 6, p. 599–606, 2021. Citado na página [63](#).

WANG, X.; DEY, D. K. Generalized Extreme Value Regression for Binary Response Data: An Application to B2B Electronic Payments System Adoption. **arXiv e-prints**, p. arXiv-1101, 2011. Citado nas páginas [35](#), [36](#), [55](#) e [123](#).

WATANABE, S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. **Journal of Machine Learning Research**, v. 11, n. 12, 2010. Citado na página [58](#).

ZOU, Q.; XIE, S.; LIN, Z.; WU, M.; JU, Y. Finding the Best Classification Threshold in Imbalanced Classification. **Big Data Research**, Elsevier, v. 5, p. 2–8, 2016. Citado nas páginas [64](#) e [66](#).

A

RCPP: UMA BREVE INTRODUÇÃO

O Rcpp (EDDELBUETTEL; FRANÇOIS, 2011) é um pacote (*package*) R que permite a extensão da linguagem R através do C++. Em outras palavras, o Rcpp facilita a transição de objetos entre códigos R e C++, permitindo, dessa forma, que se possa programar diretamente em C++, e posteriormente, tais códigos disponíveis para se utilizar como funções R. O Rcpp também possui a vantagem de ser utilizado como uma interface para outros softwares fazerem parte do R. Um exemplo é o uso da biblioteca de álgebra matricial Armadillo (através do pacote RcppArmadillo) do C++.

O C++ é uma linguagem de programação estaticamente tipada, ou seja, necessitamos declarar o tipo de armazenamento que será usado por uma variável. Pode-se argumentar que isso seja uma desvantagem frente a outras linguagens de programação mais dinâmicas, no entanto, essa característica do C++ permite alguns benefícios. O primeiro benefício seria a facilidade de correção, isto é, o compilador do C++ apontará um erro no código toda vez que uma correspondência incorreta for atribuída a uma variável. O segundo seria a performance, pois o compilador do C++ pode otimizar o código com base nas características de armazenamento do CPU.

Para operar, o Rcpp precisa não apenas do R, mas também de um compilador. Esse compilador pode ser obtido facilmente pelos usuários do sistema Windows, ao instalar a caixa de ferramentas (um software) Rtools. Então, basicamente, para poder utilizar as funcionalidades do Rcpp no R, o usuário terá de fazer:

1. a instalação do Rtools (se for usuário do Windows)¹;
2. a instalação do pacote Rcpp no R. A forma mais simples é fazer `install.packages('Rcpp')`.

Para verificar se o Rcpp foi instalado corretamente, podemos usar a função `evalCpp`, que é uma das funções do pacote Rcpp, para avaliar a expressão abaixo entre aspas; se a função `compilar` normalmente, significa que o pacote foi instalado corretamente.

¹ Para os usuários do Linux ou MacOS, há também as suas respectivas caixas de ferramentas, que podem ser encontradas após fazer uma rápida busca na internet

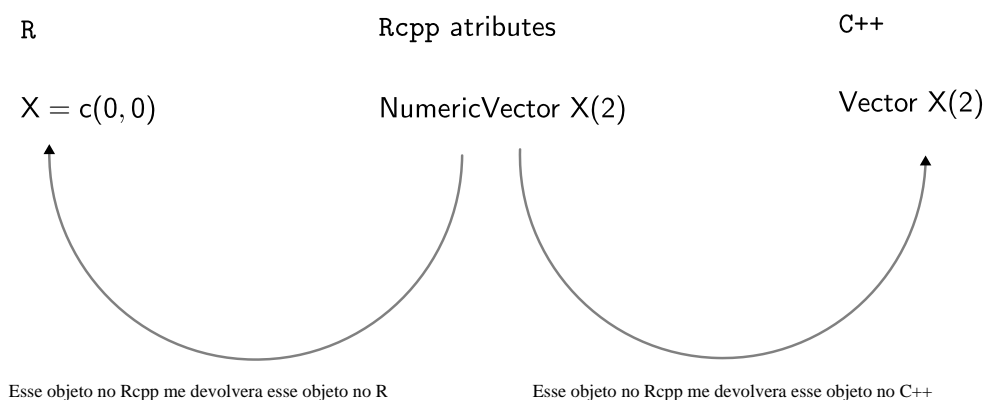
```
> library ('Rcpp')
> evalCpp("2+2")
[1] 4
```

Primeiros passos com o Rcpp

A grande vantagem do Rcpp é que ele permite a escrita de códigos C++ de forma fácil, direta e compacta. No R, dá para se escrever códigos C++ sem o Rcpp, no entanto é necessário possuir habilidades de manuseio adicionais para o preparo do código. No artigo introdutório de utilização do Rcpp (EDDELBUETTEL; BALAMUTA, 2018), os autores mostram como uma simples implementação pode se tornar muito complexa em C++ pelas vias “normais” do R.

Criamos códigos em C++ com os atributos do Rcpp (*Rcpp attributes*, (ALLAIRE; EDELBUETTEL; FRANÇOIS, 2018)), que são basicamente funções que permitem o intercâmbio entre objetos do R e do C++. Na Figura 22, temos a ilustração de como um vetor numérico de duas posições, definido com o uso do Rcpp, serve de transição (como uma ponte) entre um vetor numérico no R e um vetor numérico no C++. Conjecturando um exemplo, imagine que você necessite escrever uma função usando o Rcpp, cujo retorno é um vetor numérico de dimensão 2, usando a função `NumericVector X(2)`; quando compilarmos o código (usando uma das funções do Rcpp que veremos mais adiante), o compilador entenderá que você deseja criar um vetor **X** de duas posições no C++, e quando finalizar a compilação e o código for importado ao R para ser utilizado, essa função, quando estiver em uso, retornará um objeto pertencente a uma das classe de vetores no R.

Figura 22 – Ilustração de como um objeto no Rcpp serve de transição entre objetos do R e do C++.



O R tem a vantagem de possuir funções vetorizadas, o que não ocorre com o C++. No entanto, desenvolveu-se o *Rcpp sugar* (EDDELBUETTEL; FRANÇOIS, 2017), que são funções do Rcpp que permitem a obtenção de objetos vetorizados de forma similar ao R, mantendo a alta performance do C++. Exemplos dessas funções *sugar* do Rcpp incluem distribuições estatísticas, geração de números aleatórios etc. O exemplo abaixo ilustra a diferença entre gerar valores de uma distribuição normal padrão usando uma função escalar C++ e uma função *sugar* do Rcpp.

```
> evalCpp('R::rnorm(0,1)') # esta é escalar
[1] -0.6284838
> evalCpp('Rcpp::rnorm(3)') # esta é vetorizada
[1] 1.15138992 -0.08848312 -3.27725928
```

No exemplo abaixo, foi implementada a função de distribuição acumulada da distribuição GEV.

```
NumericVector Fcpp (NumericVector x, double xi) {

    int n = x.size();
    NumericVector pp(n);
    NumericVector z(n);

    if(xi==0)
    {
        for(int i=0; i<n; i++)
        {
            pp[i] = exp(-exp(-x[i]));
        }
    }
    else{
        z=1+xi*x;
        for(int i=0; i<n; i++)
        {
            if(z[i]<=0)
            {
                z[i] = 0;
            }
            pp[i] = exp(-pow(z[i],(-1/xi)));
        }
    }
    return pp;
}
```

No R, o código acima poderia ser escrito de forma muito mais concisa, com vetorização, como mostrado a seguir.

```
F <- function(x, xi) {
  z = 1 + xi*x
  z= ifelse(z > 0, z, 0)
  return(exp(-z^(-1/xi)))
}
```

No entanto, graças às funções *sugar* do Rcpp, podemos obter um código bem mais enxuto (código abaixo) e com uma sintaxe bem próxima ao do R. Perceba que poucas modificações foram feitas de um código para o outro.

```
NumericVector Fcpp (NumericVector x, double xi) {
  NumericVector z = 1 + xi*x;
  z= ifelse(z > 0, z, 0);
  return exp(-pow(z,(-1/xi)));
}
```

Criando uma função com o Rcpp

Como exemplo, vamos implementar a função de distribuição acumulada (f.d.a) do modelo probabilístico Cauchy Potência, e depois “traduzir” o código para o Rcpp. A seguir, temos a f.d.a do modelo Cauchy Potência no R.

```
> F <- function(x,delta) {
+   z = (0.5+(1/pi)*atan(x))^(exp(delta))
+   z
+ }
>
> F(-0.5,delta=4)
[1] 1.862518e-25
```

Agora, “traduzindo” a função acima para o Rcpp, temos o resultado a seguir:

```
1 > cppFunction('
2 + NumericVector Fcpp (NumericVector x, double delta) {
3 +   double pi = 3.141593;
4 +   NumericVector z;
5 +
6 +   z = pow((0.5+(1/pi)*atan(x)),exp(delta));
7 +   return z;
8 + }')
```

```
9 >
10 > Fcpp(-0.5,delta=4)
11 [1] 1.862522e-25
```

Do resultado acima, alguns detalhes e instruções devem ser pontuados. No C++, como mencionado antes, para todas as variáveis, é necessário especificar o tipo de armazenamento a ser usado pela variável; isso também vale para o tipo de valor (ou valores) que uma função irá retornar, e necessita-se especificar também o tipo de cada variável/objeto que será recebida(o) como argumento na função. Por isso, na linha 2 do código acima, há 'NumericVector' antes do nome da função, o que significa que a função retornará um vetor numérico. Também constatamos que essa função recebe dois argumentos, um vetor numérico (`NumericVector x`) e um valor real (`double delta`).

Outro detalhe é que toda linha de comando dentro do corpo da função (das linhas 3-7) precisa ser finalizada com um ponto e vírgula (;). Por fim, a função `cppFunction` do pacote `Rcpp` compila a expressão do código entre aspas, e então importa-o para o R sob o nome da função especificada (em nosso caso, `Fcpp()`).

Formas de compilar um código Rcpp

Existem alguns caminhos para se prototipar códigos com `Rcpp` e disponibilizá-los no R. Citaremos três e os descreveremos a seguir.

1. `evalCpp`: permite uma avaliação simples de uma expressão C++. Isso pode ser útil para pequenos testes ou para checar se o `Rcpp` foi instalado corretamente.

```
> evalCpp('Rcpp::rnorm(3)') # exemplo 1
[1]  1.15138992 -0.08848312 -3.27725928
>
> evalCpp('2+2') # exemplo 2
[1] 4
>
```

2. `cppFunctions`: pode ser usado para criar funções C++ para o R em tempo real. Ela analisa o texto fornecido, extrai os nomes das funções desejadas, cria o “cabeçalho” necessário (isto é, coloca o `#include <Rcpp.h>` e `using namespace Rcpp` de forma automática), vincula e carrega o código fornecido e o disponibiliza sob o nome especificado.

```
cppFunction('
NumericVector Fcpp (NumericVector x, double xi) {
    NumericVector z = 1 + xi*x;
    z= ifelse(z > 0, z, 0);
    return exp(-pow(z,(-1/xi)));
}')

> Fcpp(c(-5,0,5),0.5)
[1] 0.00000000 0.3678794 0.9216104
```

3. `sourceCpp`: pode integrar arquivos inteiros de uma vez para definir múltiplas funções. Ela ler o arquivo e compila, vincula e carrega o(s) código(s) nele, e os disponibiliza sob o(s) nome(s)

especificado(s). Vejamos um exemplo a seguir no qual implementamos o logaritmo da distribuição a posteriori dos coeficientes de regressão de um modelo Logito; a expressão se encontra em 2.22.

```

1 // [[Rcpp::depends(RcppArmadillo)]]
2 #include <RcppArmadillo.h>
3 using namespace Rcpp;
4
5 // [[Rcpp::export]]
6
7 arma::vec Fcpp (arma::vec x) {
8     arma::vec z;
9     z = exp(x)/(1+exp(x));
10    return z;
11 }
12
13 // [[Rcpp::export]]
14
15 double lpostbetaCpp(arma::vec theta, arma::vec y, arma::mat X){
16     int n= X.n_rows ;
17     int p = X.n_cols ;
18     arma::colvec bet(p);
19     bet = theta(arma::span(0,(p-1)));
20
21     arma::mat sigmabeta = (1/100)*arma::diagmat(arma::ones(p,p));
22     // variancia a priori
23     arma::vec log_lik(n);
24     arma::vec zi(n);
25
26     zi = X*bet;
27
28     log_lik = sum(y*log(Fcpp(zi))) +
29                sum((1-y)%log(1-Fcpp(zi)));
30
31     double lpostbetaxi =
32     arma::sum(log_lik)-0.5*as_scalar(bet.t()*sigmabeta*bet);
33
34     arma::vec lp = {lpostbetaxi,1};
35     if (lp.has_inf())
36     {
37         return -10000000000000000;
38     }
39     if (lp.has_nan())
40     {

```

```

39     return -10000000000000000;
40
41     }
42     else{
43         return lpostbetaxi;
44     }
45 }

```

Na expressão acima, estamos usando a biblioteca matricial `Armadillo` por meio do pacote `RcppArmadillo`. O comando `// [[Rcpp::depends(RcppArmadillo)]]` na linha 1 informará ao compilador que configure o ambiente de compilação para compilar corretamente e vincular ao pacote `RcppArmadillo`. Os caracteres `//` seguidos de string (como aparece na linha 21) são a forma de comentar no C++. Na linha 2, temos algo semelhante ao comando `'library()'` do R, ou seja, é como se estivéssemos “chamando” a biblioteca `RcppArmadillo` para poder usar as suas funções. Na linha 3, especificamos o namespace², e os comandos `// [[Rcpp::export]]` que aparecem nas linhas 5 e 13 são para, exportar as funções ao R e deixá-las disponíveis para uso, sob os nomes especificados, após terem sido compiladas.

Por fim, podemos salvar essas funções em um arquivo com o nome, por exemplo, `lpostbeta.cpp` no diretório de trabalho, ou em um diretório temporário. Então, dentro da sessão R, use `Rcpp::sourceCpp("lpostbeta.cpp")` (possivelmente usando um caminho, bem como o nome do arquivo).

```

library ('Rcpp')
library ('RcppArmadillo')
sourceCpp('/home/alan/Documentos/lpostbeta.cpp')
> y=rbinom(10,1,prob = 0.5)
> X=as.matrix(cbind(rep(1,10),rnorm(10)))
> beta=c(0.5,-0.5)
> lpostbetaCpp(beta,y,X)
[1] -8.524416
>

```

² Um namespace permite a definição de estruturas, classes, funções, constantes, etc, que estarão vinculadas a ele. Isso evita duplicidade como, por exemplo, outras implementações com nomes semelhantes.

EXPRESSÕES MATEMÁTICAS ÚTEIS

Neste Apêndice, apresentamos as expressões dos gradientes e matrizes de informação de Fisher, necessárias para as implementações de nosso algoritmo para cada tipo de função de ligação utilizada. As expressões serão apresentadas para o modelo presente no Capítulo 4, que representa um caso mais geral. As expressões para o modelo do Capítulo 3, e os modelos assimétricos transmutados vistos no Apêndice D, podem ser induzidas de forma conveniente ao observar que, para Y_{si} com $i = 1, \dots, m$ e $s = 1, \dots, r$, no modelo teórico apresentado no Capítulo 4, os modelos apresentados no Capítulo 3 e Apêndice D surgem quando $r = 1$.

B.1 Funções de Ligação Potência

Vimos que se H é uma função de distribuição acumulada *baseline*, então a nova função de distribuição acumulada H potência é obtida da seguinte forma:

$$H_P(y|\lambda) = H(y)^\lambda, \quad (\text{B.1})$$

em que $\lambda > 0$. Nesta tese, usamos a reparametrização apresentada em [Bazan et al. \(2014\)](#), isto é, $\delta = \log \lambda$, tal que $\delta \in \mathbb{R}$. No contexto do nosso modelo de regressão binária espacial, teremos então

$$p_{si} = H_P(z_{iz}|\delta) = H(z_{si})^{e^\delta}, \quad (\text{B.2})$$

em que $z_{si} = \mathbf{x}^\top \boldsymbol{\beta} + \phi_{si}$, e $H_P()$ será uma das distribuições potência constantes na Tabela 2. No que segue $\mathcal{L}(\delta, \boldsymbol{\beta}) = \sum_{i=1}^m \sum_{s=1}^r \log p(\mathbf{y}_{si}|\mathbf{x}_i, \boldsymbol{\beta}, \delta, \boldsymbol{\phi})$ (a log-verossimilhança), e esta possui a seguinte expressão:

$$\begin{aligned} \mathcal{L}(\delta, \boldsymbol{\beta}) &= \sum_{i=1}^m \sum_{s=1}^r y_{si} \log p_{si} + (1 - y_{si}) \log (1 - p_{si}) \\ &= \sum_{i=1}^m \sum_{s=1}^r y_{si} \log (H(z_{si})^{e^\delta}) + (1 - y_{si}) \log (1 - H(z_{si})^{e^\delta}). \end{aligned} \quad (\text{B.3})$$

Derivando a log-verossimilhança com relação a δ , tem-se

$$\begin{aligned}
\frac{\partial L(\delta, \boldsymbol{\beta})}{\partial \delta} &= \sum_{i=1}^m \sum_{s=1}^r \frac{y_{si}}{H(z_{si})e^\delta} H(z_{si})e^\delta \log(H(z_{si}))e^\delta - \frac{(1-y_{si})}{1-H(z_{si})e^\delta} H(z_{si})e^\delta \log(H(z_{si}))e^\delta \\
&= \sum_{i=1}^m \sum_{s=1}^r y_{si} \log(H(z_{si}))e^\delta - \frac{(1-y_{si})H(z_{si})e^\delta \log(H(z_{si}))e^\delta}{1-H(z_{si})e^\delta} \\
&= \sum_{i=1}^m \sum_{s=1}^r \log(H(z_{si}))e^\delta \left[y_{si} - \frac{(1-y_{si})H(z_{si})e^\delta}{1-H(z_{si})e^\delta} \right] \\
&= \sum_{i=1}^m \sum_{s=1}^r \log(H(z_{si}))e^\delta \left[\frac{y_{si} - y_{si}H(z_{si})e^\delta - H(z_{si})e^\delta + y_{si}H(z_{si})e^\delta}{1-H(z_{si})e^\delta} \right] \\
&= \sum_{i=1}^m \sum_{s=1}^r \log(H(z_{si}))e^\delta \left[\frac{y_{si} - H(z_{si})e^\delta}{1-H(z_{si})e^\delta} \right] \\
&= \sum_{i=1}^m \sum_{s=1}^r \nabla_{si} \left[\frac{y_{si} - p_{si}}{1 - p_{si}} \right].
\end{aligned}$$

Para obter a Informação de Fisher, definimos primeiro os vetores $r \times 1 \nabla_i = (\nabla_{1i}, \dots, \nabla_{mi})^\top$ e \mathbf{w}_i^* com elementos $w_i^* = (y_{si} - p_{si})/(1 - p_{si})$. Então,

$$E(\nabla_i^\top \mathbf{w}_i^* \mathbf{w}_i^{\top*} \nabla) = \nabla_i^\top E(\mathbf{w}_i^* \mathbf{w}_i^{\top*}) \nabla_i = \nabla_i^\top \Lambda^* \nabla_i, \quad (\text{B.4})$$

em que Λ^* é uma matriz diagonal $r \times r$ com elementos:

$$\lambda_i^* = \frac{1}{(1-p_{si})^2} E(y_{si} - p_{si})^2 = \frac{p_{si}(1-p_{si})}{(1-p_{si})^2} = \frac{p_{si}}{(1-p_{si})} \quad s = 1, \dots, r. \quad (\text{B.5})$$

A informação de Fisher esperada é então dada por

$$\sum_{i=1}^m \nabla_i^\top \Lambda^* \nabla_i. \quad (\text{B.6})$$

Antes de obter a derivada parcial da log-verossimilhança com relação a $\boldsymbol{\beta}$, note primeiro que

$$\frac{\partial z_{si}}{\partial \boldsymbol{\beta}} = \mathbf{x}_i. \quad (\text{B.7})$$

Então

$$\begin{aligned}
\frac{\partial L(\delta, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^m \sum_{s=1}^r \frac{y_{si}}{H(z_{si})e^\delta} H(z_{si})e^{\delta-1} h(z_{si})e^\delta \mathbf{x}_i - \frac{(1-y_{si})}{1-H(z_{si})e^\delta} H(z_{si})e^{\delta-1} h(z_{si})e^\delta \mathbf{x}_i \\
&= \sum_{i=1}^m \sum_{s=1}^r \frac{y_{si}}{H(z_{si})} e^\delta h(z_{si}) \mathbf{x}_i - \frac{(1-y_{si})e^\delta H(z_{si})e^{\delta-1} h(z_{si}) \mathbf{x}_i}{1-H(z_{si})e^\delta} \\
&= \sum_{i=1}^m \sum_{s=1}^r e^\delta h(z_{si}) \mathbf{x}_i [H(z_{si})]^{-1} \left[y_{si} - \frac{(1-y_{si})H(z_{si})e^\delta}{1-H(z_{si})e^\delta} \right] \\
&= \sum_{i=1}^m \sum_{s=1}^r e^\delta h(z_{si}) \mathbf{x}_i [H(z_{si})]^{-1} \left[\frac{y_{si} - y_{si}H(z_{si})e^\delta - H(z_{si})e^\delta + y_{si}H(z_{si})e^\delta}{1-H(z_{si})e^\delta} \right] \\
&= \sum_{i=1}^m \sum_{s=1}^r e^\delta h(z_{si}) \mathbf{x}_i [H(z_{si})]^{-1} \left[\frac{y_{si} - p_{si}}{1 - p_{si}} \right] \\
&= \sum_{i=1}^m \mathbf{X}^\top \mathbf{w}_i, \quad (\text{B.8})
\end{aligned}$$

em que \mathbf{w}_i é um vetor $r \times 1$ com

$$\mathbf{w}_i = \frac{(y_{si} - p_{si}) e^{\delta} h(z_{si})}{(1 - p_{si}) H(z_{si})} \quad s = 1, \dots, r, \quad (\text{B.9})$$

e $h(z_{si}) = \frac{\partial}{\partial z_{si}} H(z_{si})$, a função densidade de probabilidade do modelo *baseline*.

A matriz de informação esperada é então dada por

$$\sum_{i=1}^m E(\mathbf{X}^\top \mathbf{w}_i \mathbf{w}_i^\top \mathbf{X}) = \sum_{i=1}^m \mathbf{X}^\top E(\mathbf{w}_i \mathbf{w}_i^\top) \mathbf{X} = \sum_{i=1}^m \mathbf{X}^\top \Lambda \mathbf{X}, \quad (\text{B.10})$$

em que Λ é uma matriz diagonal $r \times r$ com elementos:

$$\lambda_i = \left(\frac{e^{\delta} h(z_{si})}{H(z_{si})} \right)^2 \frac{1}{(1 - p_{si})^2} E(y_{si} - p_{si})^2 = \left(\frac{e^{\delta} h(z_{si})}{H(z_{si})} \right)^2 \frac{p_{si}}{(1 - p_{si})} \quad s = 1, \dots, r. \quad (\text{B.11})$$

Para completar a matriz de informação, temos

$$\sum_{i=1}^m E(\mathbf{X}^\top \mathbf{w}_i \mathbf{w}_i^\top \nabla_i) = \sum_{i=1}^m \mathbf{X}^\top E(\mathbf{w}_i \mathbf{w}_i^\top) \nabla_i = \sum_{i=1}^m \mathbf{X}^\top \Lambda^{**} \nabla_i, \quad (\text{B.12})$$

em que Λ^{**} é uma matriz diagonal $r \times r$ com elementos:

$$\lambda_i^{**} = \left(\frac{e^{\delta} h(z_{si})}{H(z_{si})} \right) \frac{1}{(1 - p_{si})^2} E(y_{si} - p_{si})^2 = \left(\frac{e^{\delta} h(z_{si})}{H(z_{si})} \right) \frac{p_{si}}{(1 - p_{si})} \quad s = 1, \dots, r. \quad (\text{B.13})$$

B.1.1 Amostrando os Efeitos Aleatórios Espaciais

O logaritmo da distribuição condicional completa de cada ϕ_s é dado por

$$\begin{aligned} L &= \log p(\phi_s | \beta, \delta, \Omega, \mathbf{y}_s, \mathbf{x}_s) \\ &= \sum_{i=1}^m y_{si} \log(H(z_{si}) e^{\delta}) + (1 - y_{si}) \log(1 - H(z_{si}) e^{\delta}) - \frac{1}{2} \phi_s^\top \Omega \phi_s \\ &= L_z - \frac{1}{2} \phi_s^\top \Omega \phi_s. \end{aligned} \quad (\text{B.14})$$

Agora, derivando L_z com relação a cada elemento de ϕ_s , obtemos o seguinte resultado:

$$\begin{aligned} \frac{\partial L_z}{\partial \phi_{si}} &= \frac{y_{si}}{H(z_{si}) e^{\delta}} e^{\delta} [H(z_{si})]^{e^{\delta}-1} h(z_{si}) - \frac{(1 - y_{si})}{1 - H(z_{si}) e^{\delta}} e^{\delta} H(z_{si})^{e^{\delta}-1} h(z_{si}) \\ &= \frac{y_{si}}{H(z_{si})} e^{\delta} h(z_{si}) - \frac{(1 - y_{si})}{1 - H(z_{si}) e^{\delta}} e^{\delta} H(z_{si})^{e^{\delta}-1} h(z_{si}) \\ &= e^{\delta} [H(z_{si})]^{-1} h(z_{si}) \left[y_{si} - \frac{(1 - y_{si}) H(z_{si}) e^{\delta}}{1 - H(z_{si}) e^{\delta}} \right] \\ &= e^{\delta} [H(z_{si})]^{-1} g(z_{si}) \left[\frac{(y_{si} - p_{si})}{1 - p_{si}} \right] \end{aligned} \quad (\text{B.15})$$

No que segue, o vetor gradiente requerido é dado por

$$\nabla_{\phi_s} L = \nabla_{\phi_s} L_z - \Omega \phi_s, \quad (\text{B.16})$$

em que cada elemento de $\nabla_{\phi_s} L_z$ é dado por B.15.

B.2 Funções de Ligação Reversa de Potência

Seja H uma função de distribuição acumulada *baseline*, então a nova função de distribuição acumulada H reversa de potência é obtida da seguinte forma:

$$H_{RP}(y|\lambda) = 1 - H(-y)^\lambda, \quad (\text{B.17})$$

em que $\lambda > 0$. Novamente, faremos a seguinte reparametrização $\delta = \log \lambda$ tal que $\delta \in \mathbb{R}$. No contexto do nosso modelo de regressão binária espacial, teremos então que

$$p_{si} = H_{RP}(z_{si}|\delta) = 1 - H(-z_{si})^{e^\delta}, \quad (\text{B.18})$$

em que $z_{si} = \mathbf{x}^\top \boldsymbol{\beta} + \phi_{si}$, e $H_{RP}(\cdot)$ é uma das distribuições reversa de potência constantes na Tabela 2. No que segue, $\mathcal{L}(\delta, \boldsymbol{\beta}) = \sum_{i=1}^m \sum_{s=1}^r \log p(\mathbf{y}_{si}|\mathbf{x}_i, \boldsymbol{\beta}, \delta, \boldsymbol{\phi})$ (a log-verossimilhança), e esta possui a seguinte expressão:

$$\begin{aligned} \mathcal{L}(\delta, \boldsymbol{\beta}) &= \sum_{i=1}^m \sum_{s=1}^r y_{si} \log p_{si} + (1 - y_{si}) \log (1 - p_{si}) \\ &= \sum_{i=1}^m \sum_{s=1}^r y_{si} \log (1 - H(-z_{si})^{e^\delta}) + (1 - y_{si}) \log (H(-z_{si})^{e^\delta}). \end{aligned} \quad (\text{B.19})$$

Derivando a log-verossimilhança com relação a δ , tem-se

$$\begin{aligned} \frac{\partial \mathcal{L}(\delta, \boldsymbol{\beta})}{\partial \delta} &= \sum_{i=1}^m \sum_{s=1}^r \frac{y_{si}}{1 - H(-z_{si})^{e^\delta}} (-H(-z_{si})^{e^\delta} \log (H(-z_{si})^{e^\delta})) + \frac{(1 - y_{si})}{H(-z_{si})^{e^\delta}} H(-z_{si})^{e^\delta} \log (H(-z_{si})^{e^\delta}) \\ &= \sum_{i=1}^m \sum_{s=1}^r (1 - y_{si}) \log (H(-z_{si})^{e^\delta}) - \frac{y_{si} H(-z_{si})^{e^\delta} \log (H(-z_{si})^{e^\delta})}{1 - H(-z_{si})^{e^\delta}} \\ &= \sum_{i=1}^m \sum_{s=1}^r \log (H(-z_{si})^{e^\delta}) \left[(1 - y_{si}) - \frac{y_{si} H(-z_{si})^{e^\delta}}{1 - H(-z_{si})^{e^\delta}} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r \log (H(-z_{si})^{e^\delta}) \left[\frac{1 - H(-z_{si})^{e^\delta} - y_{si} (1 - H(-z_{si})^{e^\delta}) - y_{si} H(-z_{si})^{e^\delta}}{1 - H(-z_{si})^{e^\delta}} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r \log (H(-z_{si})^{e^\delta}) \left[\frac{1 - H(-z_{si})^{e^\delta} - y_{si} + y_{si} H(-z_{si})^{e^\delta} - y_{si} H(-z_{si})^{e^\delta}}{1 - H(-z_{si})^{e^\delta}} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r -\log (H(-z_{si})^{e^\delta}) \left[\frac{y_{si} - (1 - H(-z_{si})^{e^\delta})}{1 - H(-z_{si})^{e^\delta}} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r \nabla_{si} \left[\frac{y_{si} - p_{si}}{p_{si}} \right]. \end{aligned}$$

Para obter a informação de Fisher, definimos primeiro os vetores $r \times 1 \nabla_i = (\nabla_{1i}, \dots, \nabla_{mi})^\top$ e \mathbf{w}_i^* com elementos $w_i^* = (y_{si} - p_{si})/p_{si}$. Então,

$$E(\nabla_i^\top \mathbf{w}_i^* \mathbf{w}_i^{\top*} \nabla) = \nabla_i^\top E(\mathbf{w}_i^* \mathbf{w}_i^{\top*}) \nabla_i = \nabla_i^\top \Lambda^* \nabla_i, \quad (\text{B.20})$$

em que Λ^* é uma matriz diagonal $r \times r$ com elementos:

$$\lambda_i^* = \frac{1}{p_{si}^2} E(y_{si} - p_{si})^2 = \frac{p_{si}(1 - p_{si})}{p_{si}^2} = \frac{1 - p_{si}}{p_{si}} \quad s = 1, \dots, r. \quad (\text{B.21})$$

A informação de Fisher esperada é então dada por

$$\sum_{i=1}^m \nabla_i^\top \Lambda^* \nabla_i. \quad (\text{B.22})$$

Antes de obter a derivada parcial da log-verossimilhança com relação a β , note primeiro que

$$\frac{\partial z_{si}}{\partial \beta} = \mathbf{x}_i. \quad (\text{B.23})$$

Então

$$\begin{aligned} \frac{\partial L(\delta, \beta)}{\partial \beta} &= \sum_{i=1}^m \sum_{s=1}^r \frac{y_{si}}{1 - H(-z_{si})e^{\delta}} (-H(-z_{si})e^{\delta-1} h(-z_{si})e^{\delta} (-\mathbf{x}_i)) - \frac{(1 - y_{si})}{H(-z_{si})e^{\delta}} H(-z_{si})e^{\delta-1} h(-z_{si})e^{\delta} \mathbf{x}_i \\ &= \sum_{i=1}^m \sum_{s=1}^r -(1 - y_{si})h(-z_{si})e^{\delta} \mathbf{x}_i + \frac{y_{si}e^{\delta} H(-z_{si})e^{\delta-1} h(-z_{si})\mathbf{x}_i}{1 - H(-z_{si})e^{\delta}} \\ &= \sum_{i=1}^m \sum_{s=1}^r e^{\delta} h(-z_{si})\mathbf{x}_i [H(z_{si})]^{-1} \left[-(1 - y_{si}) + \frac{y_{si}H(-z_{si})e^{\delta}}{1 - H(-z_{si})e^{\delta}} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r e^{\delta} h(-z_{si})\mathbf{x}_i [H(-z_{si})]^{-1} \left[\frac{-(1 - H(-z_{si})e^{\delta}) + y_{si} - y_{si}H(-z_{si})e^{\delta} + y_{si}H(-z_{si})e^{\delta}}{1 - H(-z_{si})e^{\delta}} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r e^{\delta} h(-z_{si})\mathbf{x}_i [H(-z_{si})]^{-1} \left[\frac{y_{si} - p_{si}}{p_{si}} \right] \\ &= \sum_{i=1}^m \mathbf{X}^\top \mathbf{w}_i, \end{aligned} \quad (\text{B.24})$$

em que \mathbf{w}_i é um vetor $r \times 1$ com

$$w_i = \frac{(y_{si} - p_{si}) e^{\delta} h(-z_{si})}{p_{si} H(-z_{si})} \quad s = 1, \dots, r, \quad (\text{B.25})$$

e $h(-z_{si}) = -\frac{\partial}{\partial z_{si}} H(-z_{si})$, a função densidade de probabilidade do modelo *baseline*.

A matriz de informação esperada é então dada por

$$\sum_{i=1}^m E(\mathbf{X}^\top \mathbf{w}_i \mathbf{w}_i^\top \mathbf{X}) = \sum_{i=1}^m \mathbf{X}^\top E(\mathbf{w}_i \mathbf{w}_i^\top) \mathbf{X} = \sum_{i=1}^m \mathbf{X}^\top \Lambda \mathbf{X}, \quad (\text{B.26})$$

em que Λ é uma matriz diagonal $r \times r$ com elementos:

$$\lambda_i = \left(\frac{e^{\delta} h(-z_{si})}{H(-z_{si})} \right)^2 \frac{1}{p_{si}^2} E(y_{si} - p_{si})^2 = \left(\frac{e^{\delta} h(-z_{si})}{H(-z_{si})} \right)^2 \frac{1 - p_{si}}{p_{si}} \quad s = 1, \dots, r. \quad (\text{B.27})$$

Para completar a matriz de informação, temos

$$\sum_{i=1}^m E(\mathbf{X}^\top \mathbf{w}_i \mathbf{w}_i^\top \nabla_i) = \sum_{i=1}^m \mathbf{X}^\top E(\mathbf{w}_i \mathbf{w}_i^\top) \nabla_i = \sum_{i=1}^m \mathbf{X}^\top \Lambda^{**} \nabla_i, \quad (\text{B.28})$$

em que Λ^{**} é uma matriz diagonal $r \times r$ com elementos:

$$\lambda_i^{**} = \left(\frac{e^{\delta} h(-z_{si})}{H(-z_{si})} \right) \frac{1}{p_{si}^2} E(y_{si} - p_{si})^2 = \left(\frac{e^{\delta} h(-z_{si})}{H(-z_{si})} \right) \frac{1 - p_{si}}{p_{si}} \quad s = 1, \dots, r. \quad (\text{B.29})$$

B.2.1 Amostrando os Efeitos Aleatórios Espaciais

O logaritmo da distribuição condicional completa de cada ϕ_s é dado por

$$\begin{aligned}
 L &= \log p(\phi_s | \beta, \delta, \Omega, \mathbf{y}_s, \mathbf{x}_s) \\
 &= \sum_{i=1}^m y_{si} \log(1 - H(-z_{si})e^\delta) + (1 - y_{si}) \log(H(-z_{si})e^\delta) - \frac{1}{2} \phi_s^\top \Omega \phi_s \\
 &= L_z - \frac{1}{2} \phi_s^\top \Omega \phi_s.
 \end{aligned} \tag{B.30}$$

Agora, derivando L_z com relação a cada elemento de ϕ_s , obtemos o seguinte resultado:

$$\begin{aligned}
 \frac{\partial L_z}{\partial \phi_{si}} &= \frac{-y_{si}}{1 - H(-z_{si})e^\delta} (-H(-z_{si})e^{\delta-1} h(-z_{si})e^\delta (-1)) - \frac{(1 - y_{si})}{H(-z_{si})e^\delta} H(-z_{si})e^{\delta-1} h(-z_{si})e^\delta \\
 &= -(1 - y_{si})h(-z_{si})e^\delta + \frac{y_{si}e^\delta H(-z_{si})e^{\delta-1} h(-z_{si})}{1 - H(-z_{si})e^\delta} \\
 &= e^\delta h(-z_{si}) [H(z_{si})]^{-1} \left[-(1 - y_{si}) + \frac{y_{si}H(-z_{si})e^\delta}{1 - H(-z_{si})e^\delta} \right] \\
 &= e^\delta h(-z_{si}) [H(-z_{si})]^{-1} \left[\frac{-(1 - H(-z_{si})e^\delta) + y_{si} - y_{si}H(-z_{si})e^\delta + y_{si}H(-z_{si})e^\delta}{1 - H(-z_{si})e^\delta} \right] \\
 &= e^\delta h(-z_{si}) [H(-z_{si})]^{-1} \left[\frac{y_{si} - p_{si}}{p_{si}} \right]
 \end{aligned} \tag{B.31}$$

No que segue, o vetor gradiente requerido é dado por

$$\nabla_{\phi_s} L = \nabla_{\phi_s} L_z - \Omega \phi_s, \tag{B.32}$$

em que cada elemento de $\nabla_{\phi_s} L_z$ é dado por B.31.

B.3 Funções de Ligação Assimétrica Transmutada

Seja H uma função de distribuição acumulada *baseline*, então a nova função de distribuição acumulada H assimétrica transmutada é obtida da seguinte forma:

$$H_{AT}(y|\lambda) = H(y)[1 + \lambda\{1 - H(y)\}], \tag{B.33}$$

em que $|\lambda| < 1$. Seguindo Alves, Bazán e Arellano-Valle (2023), adotamos a seguinte reparametrização $\delta = \log[-(\lambda + 1)/(\lambda - 1)]$ tal que $\delta \in \mathbb{R}$. Em um contexto de modelo de regressão binária espacial, teremos então que

$$p_{si} = H_{AT}(z_{iz}|\lambda) = H(z_{si})[1 + \lambda\{1 - H(z_{si})\}], \tag{B.34}$$

em que $z_{si} = \mathbf{x}^\top \beta + \phi_{si}$. No que segue, $\mathcal{L}(\delta, \beta) = \sum_{i=1}^m \sum_{s=1}^r \log p(\mathbf{y}_{si} | \mathbf{x}_i, \beta, \delta, \phi)$ (a log-verossimilhança), e esta possui a seguinte expressão:

$$\begin{aligned}
 \mathcal{L}(\delta, \beta) &= \sum_{i=1}^m \sum_{s=1}^r y_{si} \log p_{si} + (1 - y_{si}) \log(1 - p_{si}) \\
 &= \sum_{i=1}^m \sum_{s=1}^r y_{si} \log(H(z_{si})[1 + \delta\{1 - H(z_{si})\}]) + (1 - y_{si}) \log(1 - H(z_{si})[1 + \delta\{1 - H(z_{si})\}])
 \end{aligned} \tag{B.35}$$

Antes de obtermos a derivada com relação a δ , note primeiro que

$$\lambda = \frac{e^\delta - 1}{e^\delta + 1} \text{ e } \frac{\partial \lambda}{\partial \delta} = \frac{e^\delta(e^\delta + 1) - (e^\delta - 1)e^\delta}{(e^\delta + 1)^2} = \frac{2e^\delta}{(e^\delta + 1)^2} \quad (\text{B.36})$$

Agora, derivando a log-verossimilhança com relação a δ , tem-se

$$\begin{aligned} \frac{\partial L(\lambda, \beta)}{\partial \delta} &= \sum_{i=1}^m \sum_{s=1}^r \frac{y_{si}}{H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} H(z_{si})\{1 - H(z_{si})\} \frac{\partial \lambda}{\partial \delta} + \frac{(1 - y_{si})}{1 - H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} \times \\ &\quad (-H(z_{si})\{1 - H(z_{si})\}) \frac{\partial \lambda}{\partial \delta} \\ &= \sum_{i=1}^m \sum_{s=1}^r \frac{y_{si}}{H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} H(z_{si})\{1 - H(z_{si})\} \frac{2e^\delta}{(e^\delta + 1)^2} + \frac{(1 - y_{si})}{1 - H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} \times \\ &\quad (-H(z_{si})\{1 - H(z_{si})\}) \frac{2e^\delta}{(e^\delta + 1)^2} \\ &= \sum_{i=1}^m \sum_{s=1}^r \frac{2e^\delta}{(e^\delta + 1)^2} (1 - H(z_{si})) \left[\frac{y_{si}}{[1 + \lambda\{1 - H(z_{si})\}]} - \frac{(1 - y_{si})H(z_{si})}{1 - H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r \frac{2e^\delta}{(e^\delta + 1)^2} (1 - H(z_{si})) \left[\frac{y_{si}(1 - p_{si}) - H(z_{si})[1 + \lambda\{1 - H(z_{si})\}](1 - y_{si})}{(1 - p_{si})[1 + \lambda\{1 - H(z_{si})\}]} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r \frac{2e^\delta}{(e^\delta + 1)^2} \frac{(1 - H(z_{si}))}{[1 + \lambda\{1 - H(z_{si})\}]} \left[\frac{y_{si}(1 - p_{si}) - p_{si}(1 - y_{si})}{(1 - p_{si})} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r \frac{2e^\delta}{(e^\delta + 1)^2} \frac{(1 - H(z_{si}))}{[1 + \lambda\{1 - H(z_{si})\}]} \left[\frac{y_{si} - y_{si}p_{si} - p_{si} + y_{si}p_{si}}{(1 - p_{si})} \right] \\ &= \sum_{i=1}^m \sum_{s=1}^r \nabla_{si} \left[\frac{y_{si} - p_{si}}{1 - p_{si}} \right]. \end{aligned}$$

Para obter a informação de Fisher, definimos primeiro os vetores $r \times 1 \nabla_i = (\nabla_{1i}, \dots, \nabla_{mi})^\top$ e \mathbf{w}_i^* com elementos $w_i^* = (y_{si} - p_{si})/(1 - p_{si})$. Então,

$$E(\nabla_i^\top \mathbf{w}_i^* \mathbf{w}_i^{\top*} \nabla) = \nabla_i^\top E(\mathbf{w}_i^* \mathbf{w}_i^{\top*}) \nabla_i = \nabla_i^\top \Lambda^* \nabla_i, \quad (\text{B.37})$$

em que Λ^* é uma matriz diagonal $r \times r$ com elementos:

$$\lambda_i^* = \frac{1}{(1 - p_{si})^2} E(y_{si} - p_{si})^2 = \frac{p_{si}(1 - p_{si})}{(1 - p_{si})^2} = \frac{p_{si}}{(1 - p_{si})} \quad s = 1, \dots, r. \quad (\text{B.38})$$

A informação de Fisher esperada é então dada por

$$\sum_{i=1}^m \nabla_i^\top \Lambda^* \nabla_i. \quad (\text{B.39})$$

Para obter a derivada parcial da log-verossimilhança com relação a β , note primeiro que

$$\frac{\partial p_{si}}{\partial z_{si}} = \frac{\partial H_{AT}(z_{si})}{\partial z_{si}} = h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \text{ e } \frac{\partial z_{si}}{\partial \beta} = \mathbf{x}_i. \quad (\text{B.40})$$

Então

$$\begin{aligned}
\frac{\partial L(\lambda, \beta)}{\partial \beta} &= \sum_{i=1}^m \sum_{s=1}^r \frac{y_{si}}{H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \mathbf{x}_i + \\
&\quad \frac{(1 - y_{si})}{1 - H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} (-h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}]) \mathbf{x}_i \\
&= \sum_{i=1}^m \sum_{s=1}^r h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \mathbf{x}_i \left[\frac{y_{si}}{p_{si}} - \frac{(1 - y_{si})}{1 - p_{si}} \right] \\
&= \sum_{i=1}^m \sum_{s=1}^r h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \mathbf{x}_i \left[\frac{y_{si}(1 - p_{si}) - p_{si}(1 - y_{si})}{p_{si}(1 - p_{si})} \right] \\
&= \sum_{i=1}^m \sum_{s=1}^r h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \mathbf{x}_i \left[\frac{y_{si} - p_{si}}{p_{si}(1 - p_{si})} \right] \\
&= \sum_{i=1}^m \mathbf{X}^\top \mathbf{w}_i, \tag{B.41}
\end{aligned}$$

em que \mathbf{w}_i é um vetor $r \times 1$ com

$$w_i = \frac{(y_{si} - p_{si})}{p_{si}(1 - p_{si})} h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \quad s = 1, \dots, r. \tag{B.42}$$

A matriz de informação esperada é então dada por

$$\sum_{i=1}^m E(\mathbf{X}^\top \mathbf{w}_i \mathbf{w}_i^\top \mathbf{X}) = \sum_{i=1}^m \mathbf{X}^\top E(\mathbf{w}_i \mathbf{w}_i^\top) \mathbf{X} = \sum_{i=1}^m \mathbf{X}^\top \Lambda \mathbf{X}, \tag{B.43}$$

em que Λ é uma matriz diagonal $r \times r$ com elementos:

$$\begin{aligned}
\lambda_i &= (h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}])^2 \frac{1}{p_{si}^2(1 - p_{si})^2} E(y_{si} - p_{si})^2 \\
&= \Delta_{si}^2 \frac{1}{(1 - p_{si})}, \quad i = 1, \dots, n. \tag{B.44}
\end{aligned}$$

em que $\Delta_{si} = h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}]$ para $s = 1, \dots, r$ e $s = 1, \dots, m$.

Para completar a matriz de informação, temos

$$\sum_{i=1}^m E(\mathbf{X}^\top \mathbf{w}_i \mathbf{w}_i^{\top*} \nabla_i) = \sum_{i=1}^m \mathbf{X}^\top E(\mathbf{w}_i \mathbf{w}_i^{\top*}) \nabla_i = \sum_{i=1}^m \mathbf{X}^\top \Lambda^{**} \nabla_i, \tag{B.45}$$

em que Λ^{**} é uma matriz diagonal $r \times r$ com elementos:

$$\lambda_i^{**} = \frac{\Delta_{si} E(y_{si} - p_{si})^2}{p_{si} (1 - p_{si})^2} = \frac{\Delta_{si}}{1 - p_{si}} \quad s = 1, \dots, r. \tag{B.46}$$

B.3.1 Amostrando os Efeitos Aleatórios Espaciais

O logaritmo da distribuição condicional completa de cada ϕ_s é dado por

$$\begin{aligned}
L &= \log p(\phi_s | \beta, \delta, \Omega, \mathbf{y}_s, \mathbf{x}_s) \\
&= \sum_{i=1}^m y_{si} \log(H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]) + (1 - y_{si}) \log(1 - H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]) - \frac{1}{2} \phi_s^\top \Omega \phi_s \\
&= L_z - \frac{1}{2} \phi_s^\top \Omega \phi_s. \tag{B.47}
\end{aligned}$$

Agora, derivando L_z com relação a cada elemento de ϕ_s , obtemos o seguinte resultado:

$$\begin{aligned}
\frac{\partial L_z}{\phi_{si}} &= \frac{y_{si}}{H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] + \\
&\quad \frac{(1 - y_{si})}{1 - H(z_{si})[1 + \lambda\{1 - H(z_{si})\}]} (-h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}]) \\
&= h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \left[\frac{y_{si}}{p_{si}} - \frac{(1 - y_{si})}{1 - p_{si}} \right] \\
&= h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \left[\frac{y_{si}(1 - p_{si}) - p_{si}(1 - y_{si})}{p_{si}(1 - p_{si})} \right] \\
&= h(z_{si})[1 + \lambda\{1 - 2H(z_{si})\}] \left[\frac{y_{si} - p_{si}}{p_{si}(1 - p_{si})} \right] \tag{B.48}
\end{aligned}$$

No que segue, o vetor gradiente requerido é dado por

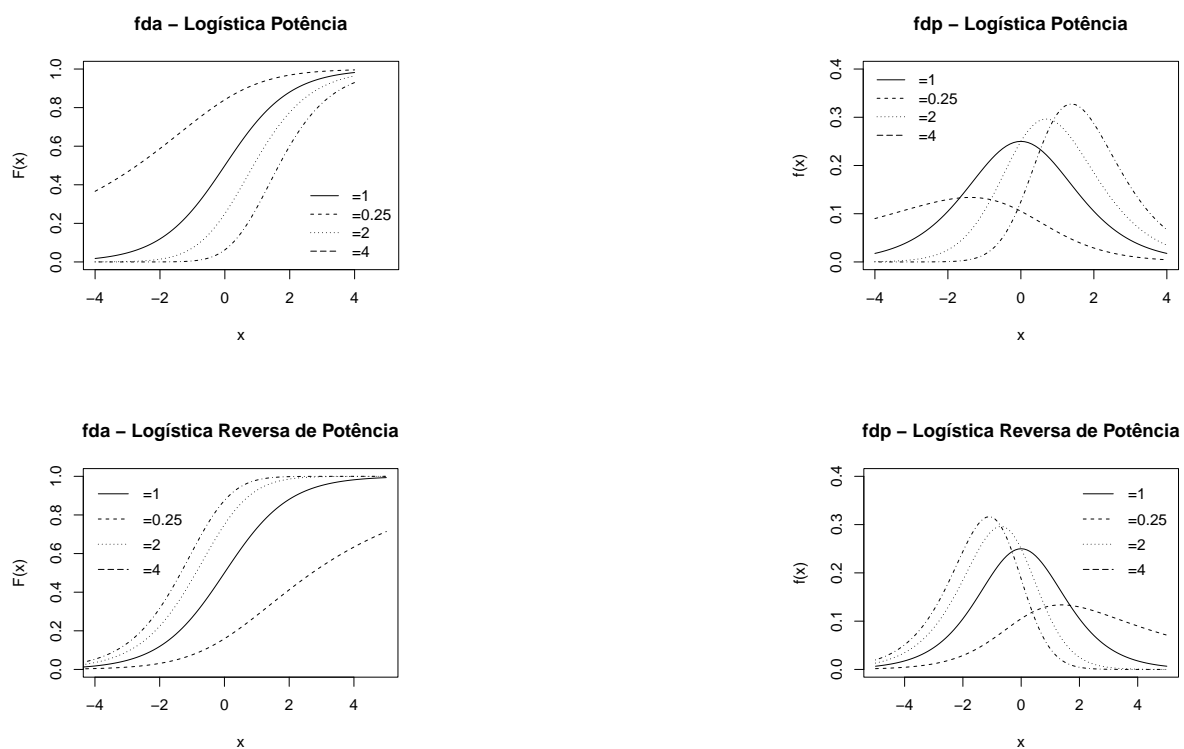
$$\nabla_{\phi_s} L = \nabla_{\phi_s} L_z - \Omega \phi_s, \tag{B.49}$$

em que cada elemento de $\nabla_{\phi_s} L_z$ é dado por B.48.

CAPÍTULO 2

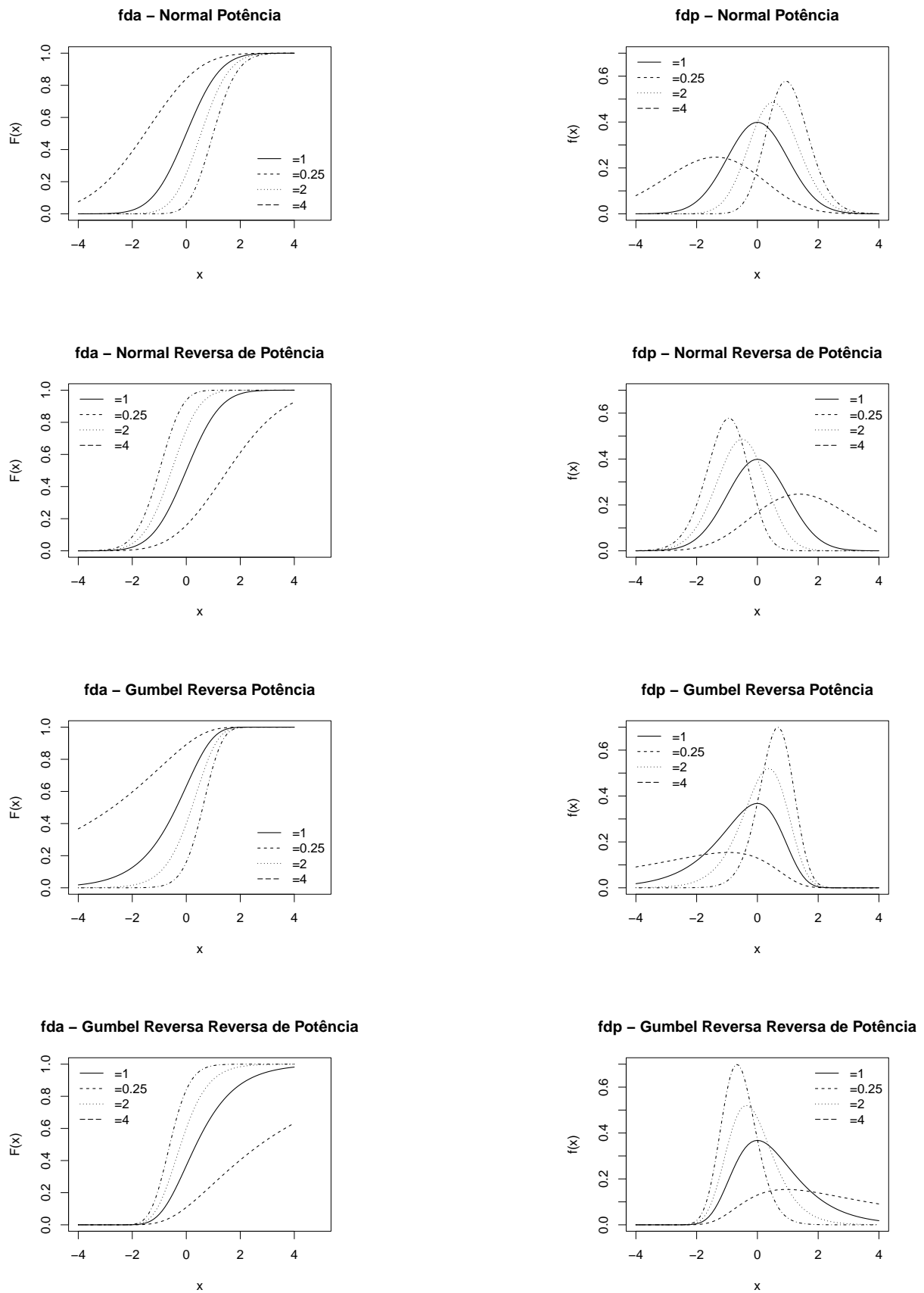
C.1 Curvas das Fda's e Fdp's dos Modelos L-P, L-RP, N-P, N-RP, GR-P e GR-RP

Figura 23 – Funções de distribuição acumulada e funções densidade de probabilidade para os modelos: L-P e L-RP, usando diferentes valores de λ .



Fonte: Elaborada pelo autor.

Figura 24 – Funções de distribuição acumulada e funções densidade de probabilidade para os modelos: N-P, N-RP, GR-P e GR-RP, usando diferentes valores de λ .



Fonte: Elaborada pelo autor.

CAPÍTULO 3

D.1 Modelos Simultaneamente Autoregressivos (SAR)

Os modelos SAR são customeiramente introduzidos em contextos de regressão. [Anselin \(2002\)](#) nomeia esses modelos, falando em um contexto de econometria espacial, como modelos de defasagem espacial (*spatial lag*), fazendo uma analogia com modelos vistos em análise de séries temporais, diferenciando-os dos chamados modelos de erro espacial (*spatial error*). A diferença básica entre esses modelos, portanto, é que os modelos de defasagem espacial lidam com a autocorrelação espacial ao nível substancial (*substantive spatial correlation*), em outras palavras, é a dependência espacial esperada para Y_i apenas por conhecer os seus vizinhos, mas sem a inserção de qualquer efeito aleatório espacial. Ao passo que, para os modelos de erro espacial, a autocorrelação espacial é tratada como ruído (*nuisance*), e o tratamento desta se dá por meio da inserção de efeitos aleatórios espaciais que modelam a estrutura de covariância do erro $\mathbb{E}(\epsilon\epsilon^\top)$ por meio, por exemplo, de um processo estocástico espacial.

Portanto, considerando um contexto de regressão, assumindo que os termos ϵ_i sejam independentes ($i = 1, \dots, m$), e que $\epsilon \sim N_m(\mathbf{0}, \tilde{\mathbf{D}})$ em que $\tilde{\mathbf{D}}$ é uma matriz diagonal com $(\tilde{\mathbf{D}})_{ii} = \sigma_i^2$; dada uma matriz $\tilde{\mathbf{B}}$ de pesos que codifica a estrutura de vizinhança espacial¹, apresentamos o modelo SAR da seguinte forma:

$$\mathbf{Y} = \tilde{\mathbf{B}}\mathbf{Y} + (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (\text{D.1})$$

em que \mathbf{X} é uma matriz $m \times k$ contendo os valores observados de k variáveis explicativas; $\boldsymbol{\beta}$ é um vetor $k \times 1$ de coeficientes de regressão.

De [D.1](#), constatamos que os dados \mathbf{Y} são modelados através de um componente que providencia um ponderamento espacial dos seus vizinhos e uma parte equivalente a uma regressão linear ordinária ([BANERJEE; CARLIN; GELFAND, 2003](#)). Neste caso, $\tilde{\mathbf{B}}$ desempenha um papel interessante no modelo, pois se $\tilde{\mathbf{B}} = \mathbf{I}$, obtemos um modelo puramente espacial, e caso $\tilde{\mathbf{B}}$ seja uma matriz de zeros, [D.1](#) torna-se

¹ Estamos assumindo uma matriz de proximidade qualquer \mathbf{B} como forma de apresentar o modelo SAR em uma apresentação mais geral. Mas, um caso particular seria, por exemplo, fazer $\mathbf{B} = \mathbf{W}$ uma matriz de adjacência.

uma regressão linear ordinária. Uma outra característica que podemos considerar é que o modelo não traz a inserção de qualquer efeito aleatório espacial, como mencionado no início desta Seção, pois os termos de erro aleatório são independentes.

Rearranjando os termos em (D.1), podemos obter a seguinte expressão:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \tilde{\mathbf{B}})^{-1}\boldsymbol{\epsilon}. \quad (\text{D.2})$$

A expressão acima ilustra como a determinação de cada Y_i , para $i = 1, \dots, m$, é resultante dos termos de erro em todas as sub-regiões do sistema considerado, e não apenas o erro em i (ANSELIN, 2002). Esse fato evidencia a “simultaneidade” atrelada à determinação de cada Y_i , diferente da abordagem condicional, na qual o valor de cada uma das sub-regiões é condicionado especificamente aos seus vizinhos (ROSSONI *et al.*, 2019).

Através de D.2, assumindo que $(\mathbf{I} - \tilde{\mathbf{B}})^{-1}$ exista, podemos induzir a distribuição para os dados \mathbf{Y} . Sendo estes uma combinação linear de $\boldsymbol{\epsilon}$, então a especificação conjunta (ou simultânea) para $\boldsymbol{\epsilon}$ induz uma distribuição conjunta para os dados. Portanto, \mathbf{Y} segue distribuição Normal multivariada tal que $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, (\mathbf{I} - \tilde{\mathbf{B}})^{-1}\tilde{\mathbf{D}}((\mathbf{I} - \tilde{\mathbf{B}})^{-1})^\top)$, pois

$$\begin{aligned} \mathbb{E}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) &= \mathbb{E}((\mathbf{I} - \tilde{\mathbf{B}})^{-1}\boldsymbol{\epsilon}) \\ \mathbb{E}(\mathbf{Y}) &= \mathbb{E}(\mathbf{X}\boldsymbol{\beta}) + (\mathbf{I} - \tilde{\mathbf{B}})^{-1}\mathbb{E}(\boldsymbol{\epsilon}) \\ \mathbb{E}(\mathbf{Y}) &= \mathbf{X}\boldsymbol{\beta} \\ &\text{e,} \\ \text{Var}(\mathbf{Y}) &\equiv \mathbb{E}[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top] \\ &= (\mathbf{I} - \tilde{\mathbf{B}})^{-1}\text{Var}(\boldsymbol{\epsilon})((\mathbf{I} - \tilde{\mathbf{B}})^{-1})^\top \\ &= (\mathbf{I} - \tilde{\mathbf{B}})^{-1}\tilde{\mathbf{D}}((\mathbf{I} - \tilde{\mathbf{B}})^{-1})^\top. \end{aligned} \quad (\text{D.3})$$

De D.3, constatamos que $\text{Cov}(\boldsymbol{\epsilon}, \mathbf{Y}) = \mathbb{E}(\boldsymbol{\epsilon}^\top \mathbf{Y}) = \tilde{\mathbf{D}}((\mathbf{I} - \tilde{\mathbf{B}})^{-1})^\top$, que não é diagonal (CRESSIE, 1993, p. 406), e assim os componentes de erro não são independentes das variáveis \mathbf{Y} , sendo essa mais uma característica que diferencia esse modelo da abordagem condicional.

Cressie (1993, p. 408-410) faz algumas comparações entre os modelos CAR e SAR para o caso gaussiano. Além das diferenças já mencionadas, no decorrer da apresentação dos modelos SAR, o autor menciona que, em decorrência de os termos de erro ϵ_i serem não correlacionados com $\{Y_j; j \neq i\}$ para os modelos CAR, estes são mais preferíveis em relação aos modelos SAR quando se trata de estimação e interpretação dos parâmetros do modelo. Referindo-se à abordagem SAR, Cressie cita em seu livro resultados que demonstram que os estimadores de mínimos quadrados para os parâmetros desse modelo são inconsistentes. Para analisar mais comparações entre esses dois modelos, o leitor pode-se dirigir à Cressie (1993) e às referências citadas por este.

D.2 Distribuições Assimétricas Transmutadas

A classe de distribuições assimétricas transmutadas (*Skew Transmuted*) também propõe-se a ser uma forma flexível de modelar dados desbalanceados. À semelhança dos modelos potência e reversa

de potência, também possui um parâmetro de forma, e o procedimento de construção das distribuições consiste, basicamente, em fazer uma transformação em uma distribuição *baseline* H . Dessa forma, a partir de uma distribuição *baseline* H , fazendo a transformação abaixo, obtemos a nova distribuição assimétrica transmutada H ,

$$H_{AT}(y|\lambda) = H(y)[1 + \lambda\{1 - H(y)\}], \quad y \in \mathbb{R}, \tag{D.4}$$

e a respectiva função densidade de probabilidade é dada por

$$h_{AT}(y|\lambda) = h(y)[1 + \lambda\{1 - 2H(y)\}], \quad y \in \mathbb{R}, \tag{D.5}$$

em que $|\lambda| > 1$. Essa formulação para obter a distribuição D.4 é uma forma de inserir assimetria a uma distribuição linha de base H , e encontra-se em Shaw e Buckley (2009); esse procedimento é chamado em inglês de *quadratic rank transmutation maps (QRTM)*.

Para exemplificar o uso das distribuições assimétricas, à semelhança do que foi adotado em Alves, Bazán e Arellano-Valle (2023), utilizaremos as distribuições Gumbel e Gumbel reversa como distribuições linha de base para produzir, respectivamente, as distribuições Gumbel assimétrica transmutada (G-AT) e Gumbel reversa assimétrica transmutada (GR-AT). Assim como propusemos para os modelos potência e reversa de potência nesta tese, adotamos a seguinte reparametrização para o modelo em D.4: $\delta = \log [-(\lambda + 1)/(\lambda - 1)]$ tal que $\delta \in \mathbb{R}$. As expressões correspondentes das fda's e fdp's dessas distribuições são apresentadas na Tabela 26.

Tabela 26 – fda e fdp das distribuições assimétricas.

Modelo	fda $F_\lambda(y)$	fdp $f_\lambda(y)$
G-AT	$\exp\{-\exp\{-y\}\} \times (1 + \delta[1 - \exp\{-\exp\{-y\}\}])$	$\exp\{-y - \exp\{-y\}\} \times (1 + \delta[1 - 2\exp\{-\exp\{-y\}\}])$
GR-AT	$[1 - \exp\{-\exp\{(y)\}\}] \times (1 + \delta\exp\{-\exp\{(y)\}\})$	$\exp\{y - \exp\{y\}\} \times (1 + \delta[1 - 2\exp\{-\exp\{y\}\}])$

Fonte: Elaborada pelo autor.

Na Figura 25, temos as curvas das fda's e fdp's das distribuições G-AT e GR-AT para valores distintos de λ . Em relação à distribuição Gumbel assimétrica transmutada, esta possui inclinação em apenas um sentido, que corresponde à mesma inclinação da distribuição linha de base para todos os valores do parâmetro de forma considerado ($-0.8, 0.25$ e 0.8). O mesmo comportamento ocorre com a distribuição Gumbel Reversa assimétrica transmutada. Em relação à função de distribuição acumulada, as curvas geradas para valores de $\lambda < 0$ ficam abaixo da curva da linha de base ($\lambda = 0$); já para valores de $\lambda > 0$, as curvas geradas ficam acima da distribuição de linha de base.

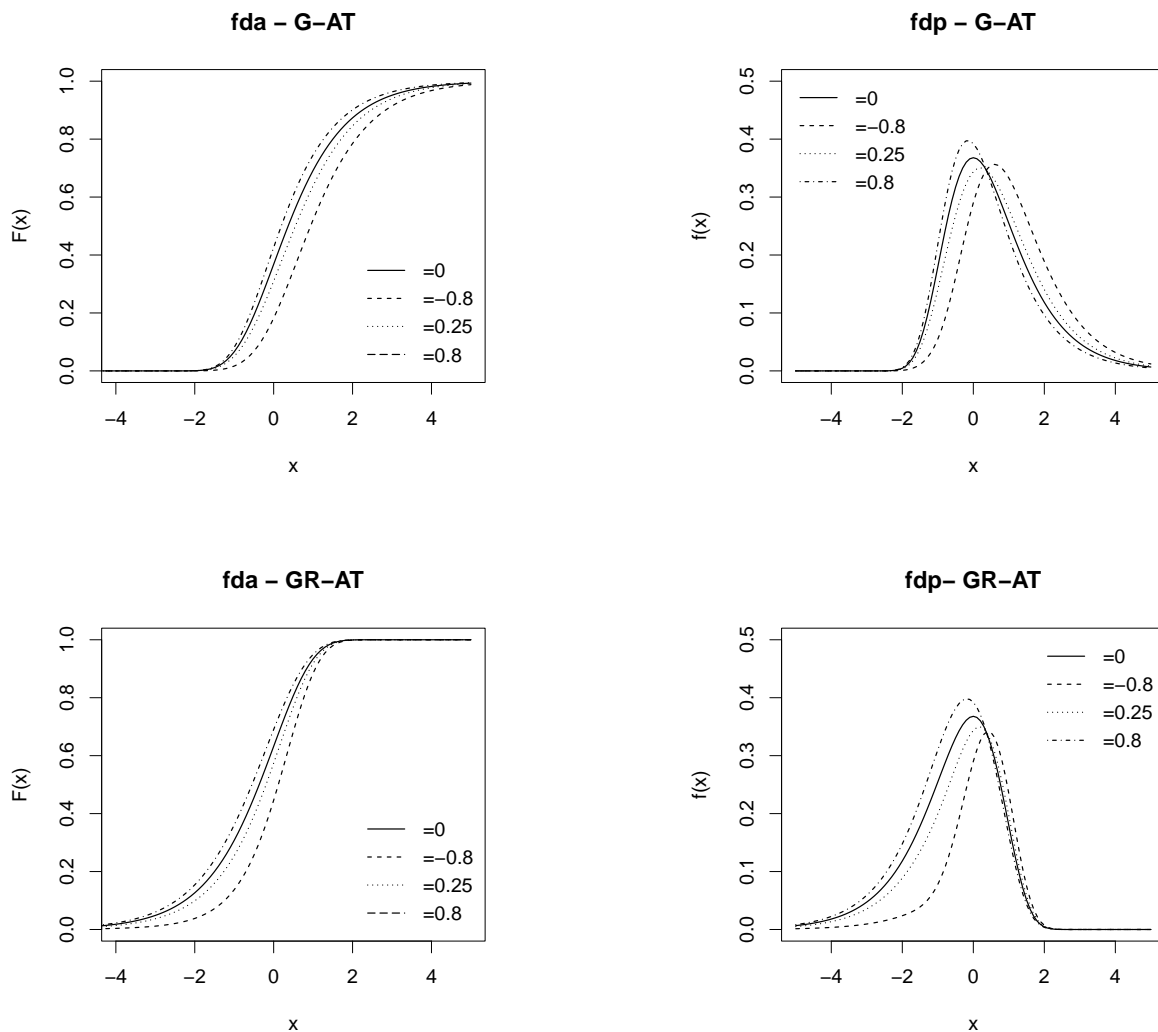
Avaliamos o desempenho dos modelos G-AT e GR-AT sob a estrutura do modelo espacial apresentado em 3.3, à semelhança do esquema de simulação proposto em duas partes na Seção 3.4: qualidade da estimativa dos parâmetros do modelo e análise de desempenho em um cenário de má

especificação, com a excessão de que, como agora trata-se de modelos diferentes, os dados serão gerados com os seguintes valores: $\lambda = -0.8$ e $\lambda = 0.8$. Novamente, usamos $\delta \sim U(-2, 2)$ como distribuição a priori para δ , e $\kappa = m$ na distribuição a priori G-Wishart.

Na Tabela 27, temos o resultado da primeira parte das simulações. Podemos constatar que os modelos G-AT e GR-AT tiveram ótimo desempenho em todos os cenários considerados, assim como ocorreu com os modelos potência e reversa de potência na Seção 3.4, com destaque para ótima recuperação do parâmetro de forma (reparametrizado) δ em todos os cenários, como podemos perceber pela probabilidade de cobertura.

Quando avaliamos esses modelos em um cenário de má especificação, também apresentam bom desempenho em todas as métricas de qualidade de ajuste, como podemos ver pela Tabela 28.

Figura 25 – Funções de distribuições acumuladas e funções densidade de probabilidade para os modelos G-AT e GR-AT, usando diferentes valores de λ .



Fonte: Elaborada pelo autor.

Tabela 27 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $n = 200$ e $n = 500$. 5000 iterações, descartando 2500 como *burn-in*. Modelos: G-AT e GR-AT.

Ligação	$n = 200$							$n = 500$						
	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio		Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	
G-AT	$\lambda = -0.8$	β_1	-0.004	0.119	-0.24	0.93		β_1	0.016	0.089	-0.176	0.97		
		β_2	0.032	0.128	0.249	0.93	497 (7.64)	β_2	0.015	0.093	0.184	0.91	3224.77 (37.75)	
		δ	0.005	0.414	-0.188	0.99		δ	0.029	0.457	-0.202	0.97		
	$\lambda = 0.8$	β_1	0.037	0.109	-0.209	0.93		β_1	0.018	0.073	-0.143	0.92		
		β_2	0.037	0.11	0.212	0.9	494.79 (10.33)	β_2	-0.002	0.071	0.143	0.93	3258.11 (51.1)	
		δ	-0.003	0.365	0.167	1		δ	0.037	0.443	0.194	0.99		
GR-AT	$\lambda = -0.8$	β_1	0.023	0.093	-0.182	0.94		β_1	0.011	0.065	-0.13	0.93		
		β_2	0.026	0.108	0.21	0.89	496 (8.78)	β_2	0.005	0.071	0.142	0.91	3243.39 (15.88)	
		δ	-0.063	0.46	-0.223	0.98		δ	-0.03	0.426	-0.2	0.97		
	$\lambda = 0.8$	β_1	0.041	0.131	-0.252	0.91		β_1	0.04	0.082	-0.158	0.96		
		β_2	0.049	0.128	0.243	0.93	490.78 (3.5)	β_2	-0.005	0.079	0.158	0.97	3239.67 (26.83)	
		δ	-0.038	0.437	0.207	0.97		δ	0.001	0.396	0.18	0.98		

Fonte: Elaborada pelo autor.

Tabela 28 – Comparação de modelos para diferentes conjuntos de dados de tamanho 100, gerados a partir de um modelo Logito. 5000 iterações, descartando as primeiras 2500 como *burn-in*.

Ligação	DIC	EAIC	EBIC	WAIC	Lm	LPML
Logito	205.9301	247.4760	254.0727	245.4157	83.62801	-122.7085
G-AT	216.0402	247.9665	254.5632	246.6572	83.79900	-123.3403
GR-AT	214.1633	249.1472	255.7438	248.0772	84.16702	-124.0559

Fonte: Elaborada pelo autor.

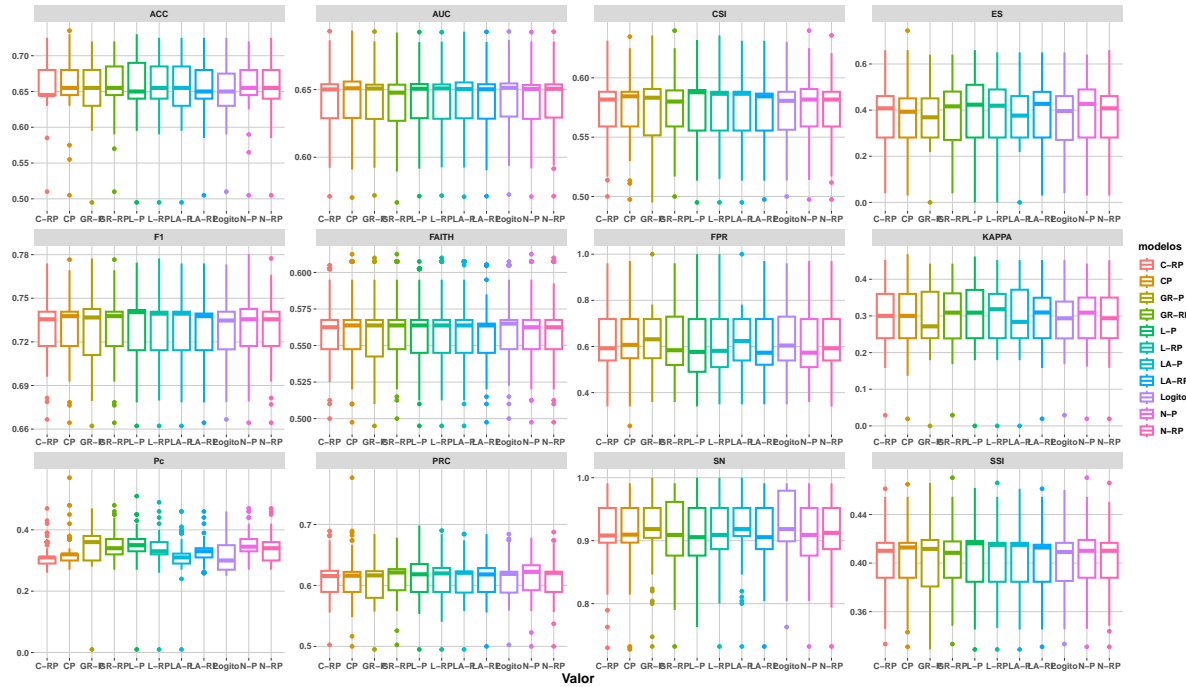
D.3 Estudo de Simulação das Métricas de Predição

Nesta seção, avaliamos o desempenho de cada modelo considerado nesta tese nas métricas de predição apresentadas em 2.7, incluindo também um cenário de má especificação da função de ligação, no qual os dados foram gerados a partir da ligação simétrica Logito. Os tamanhos amostrais, valores de λ , covariáveis e demais procedimentos de geração seguiram o descrito na Seção 3.4. Para cada combinação de tamanho amostral e valor de λ , bem como para o cenário de má especificação, realizamos 100 replicações de conjuntos de dados binários. As médias a posteriori foram utilizadas como estimadores *plug-in* da probabilidade predita. Os resultados correspondentes são apresentados na Tabela 8, e nas Figuras 26, 27 e 28.

A Figura 26 e a Tabela 29 apresenta os resultados das análises de robustez dos modelos com funções de ligação flexíveis. As médias a posteriori foram utilizadas como estimadores *plug-in* da probabilidade predita para o cálculo das métricas de predição. Os resultados indicam que os modelos potência e reversa de potência alcançam desempenho comparável ao do modelo Logito. Esse achado está em consonância com a avaliação da qualidade do ajuste em cenário de má especificação realizada anteriormente (Seção 3.4). A flexibilidade dessas funções, aliada à calibração do ponto de corte discutida em Assunção, Izbicki e Prates (2024), justifica o bom desempenho observado.

Avaliando os resultados das Figuras 27 e 28, percebe-se que os desempenhos dos modelos de acordo com as métricas não são similares. Os resultados apresentados é de que os modelos demonstram desempenhos distintos, ficando mais acentuado para algumas métricas. Quando aumentamos o tamanho da amostra, essa disparidade fica ainda mais nítida em todos os cenários considerados; nesse quesito, as métricas AUC, CSI, FAITH, F1 podem ser destacadas. Das métricas consideradas, a ACC (acurácia), talvez seja aquela na qual os modelos apresentaram mais uniformidade em todos os cenários considerados. Uma das razões que podem ter influenciado na disparidade dos desempenhos pode ter sido a proporção de 1's que cada modelo, sob os valores de λ , produz. Isso é uma hipótese que necessita ser verificada em um estudo de simulação apropriado para a nossa classe de modelos. No entanto, os leitores podem consultar Huayanay, Bazán e Russo (2025), que apresentam um estudo de simulação nesse sentido sob o contexto de regressão binária tradicional.

Figura 26 – Box-plot de cada métrica, para cada modelo, baseado em 100 conjuntos de dados gerados de um modelo Logito. pc - ponto de corte estimado



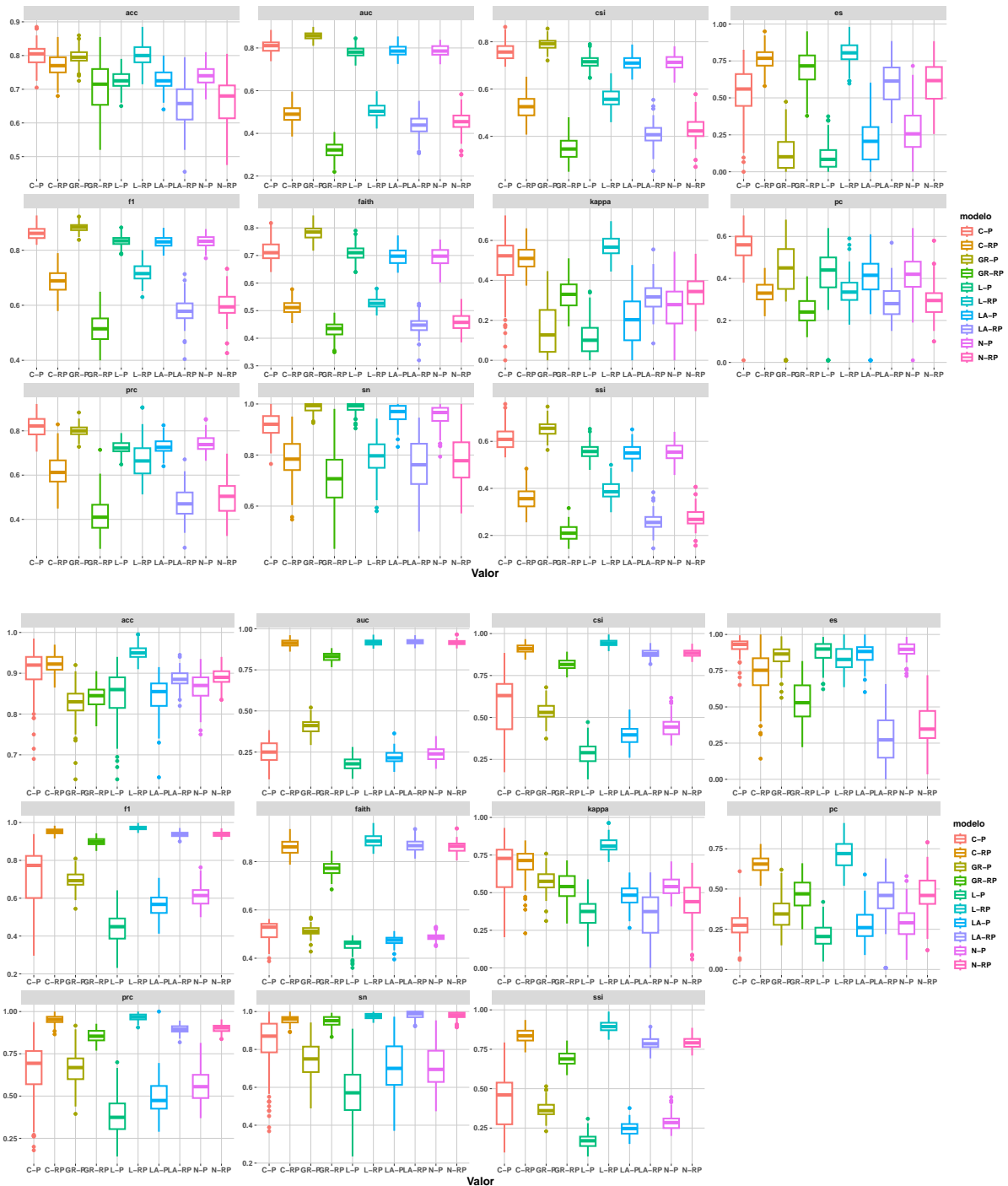
Fonte: Elaborada pelo autor.

Tabela 29 – Comparação de modelos baseada em métricas de predição para diferentes conjuntos de dados de tamanho 100, gerados a partir de um modelo Logito. pc - ponto de corte estimado.

Ligação	ACC	AUC	CSI	ES	F1	FAITH	FPR	KAPPA	pc	PRC	SN	SSI
Logito	0.65	0.64	0.58	0.36	0.73	0.56	0.64	0.29	0.31	0.61	0.93	0.41
C-P	0.66	0.64	0.58	0.38	0.73	0.56	0.62	0.30	0.32	0.61	0.92	0.41
C-RP	0.66	0.64	0.58	0.38	0.73	0.56	0.62	0.30	0.31	0.61	0.91	0.41
G-RP	0.66	0.64	0.58	0.38	0.73	0.56	0.62	0.30	0.34	0.61	0.91	0.41
GR-RP	0.66	0.64	0.58	0.39	0.73	0.56	0.61	0.30	0.35	0.61	0.91	0.41
LA-P	0.66	0.64	0.58	0.38	0.73	0.56	0.62	0.30	0.32	0.61	0.92	0.41
LA-RP	0.66	0.64	0.58	0.38	0.73	0.56	0.62	0.30	0.32	0.61	0.91	0.41
L-P	0.66	0.64	0.58	0.39	0.73	0.56	0.61	0.30	0.35	0.62	0.91	0.41
L-RP	0.66	0.64	0.58	0.39	0.73	0.56	0.61	0.30	0.33	0.61	0.91	0.41
N-P	0.66	0.64	0.58	0.39	0.73	0.56	0.61	0.30	0.35	0.61	0.91	0.41
N-RP	0.66	0.64	0.58	0.38	0.73	0.56	0.62	0.30	0.34	0.61	0.91	0.41

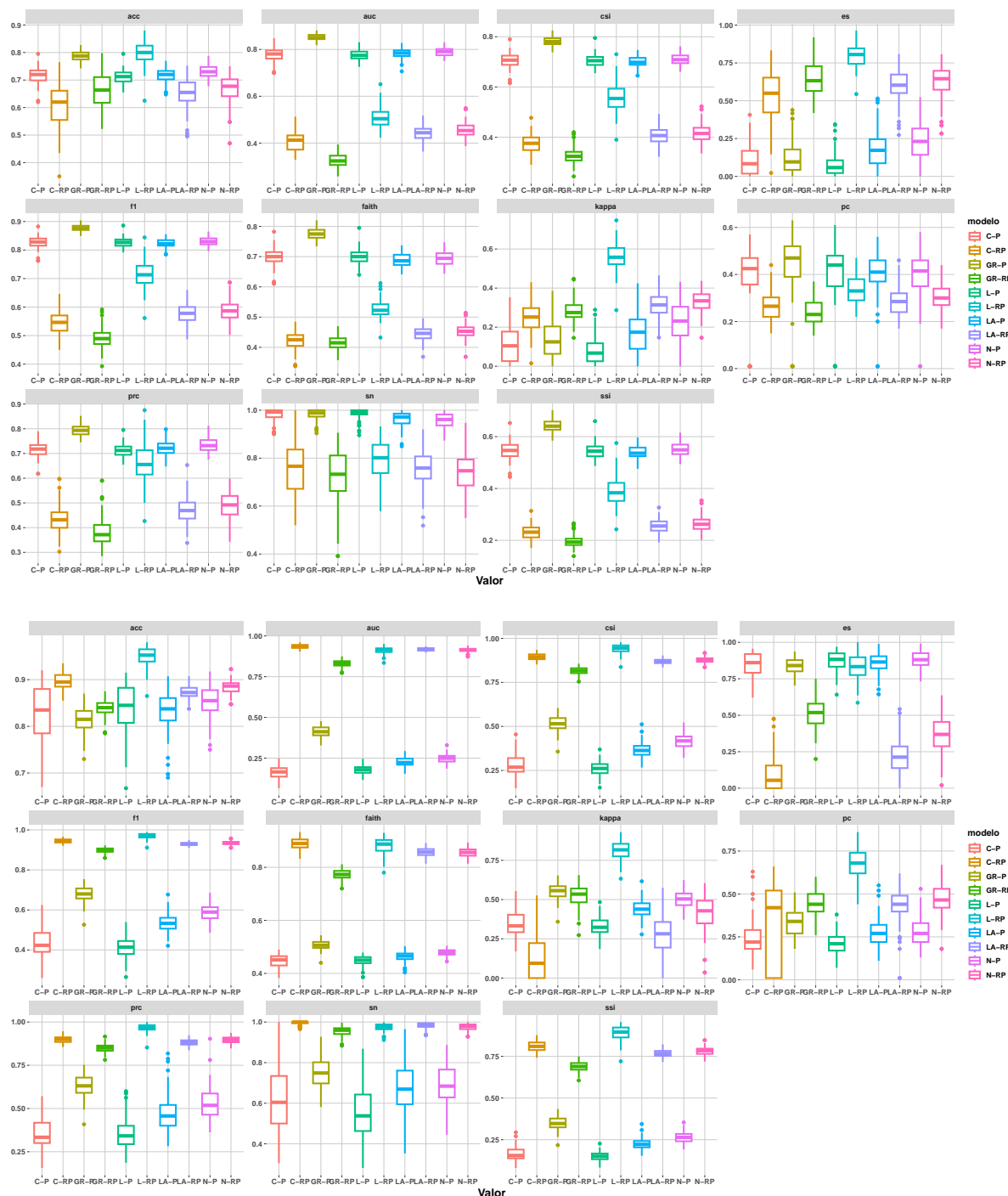
Fonte: Elaborada pelo autor.

Figura 27 – *Box-plot* de cada métrica para cada modelo. Gráfico superior : $m = 200$ e $\lambda = 0,5$; gráfico inferior: $m = 200$ e $\lambda = 4$. pc - ponto de corte estimado



Fonte: Elaborada pelo autor.

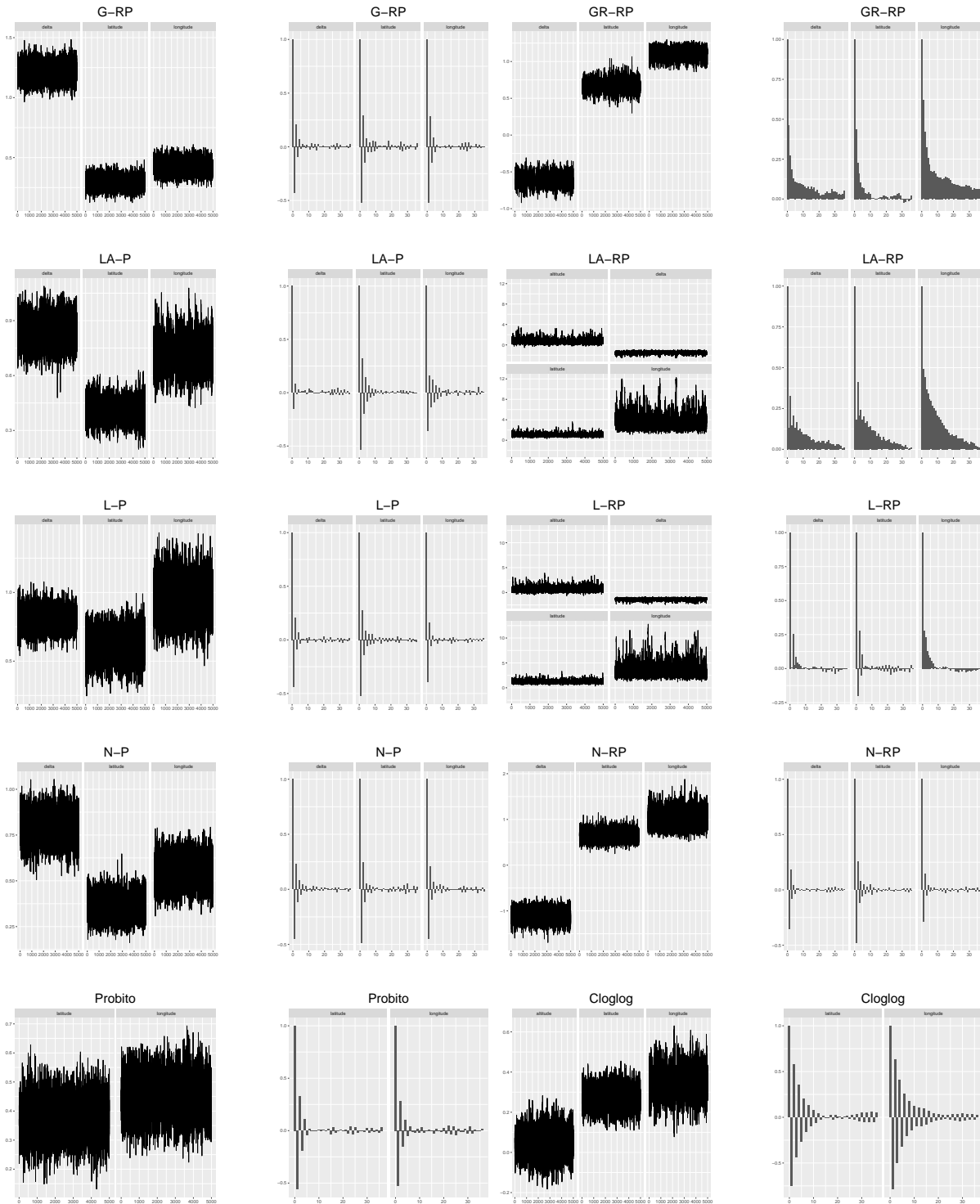
Figura 28 – Box-plot de cada métrica para cada modelo. Gráfico superior: $m = 500$ e $\lambda = 0,5$; gráfico inferior: $m = 500$ e $\lambda = 4$. pc - ponto de corte estimado



Fonte: Elaborada pelo autor.

D.4 Trace-plot e gráficos de Autocorrelação dos Modelos Ajustados

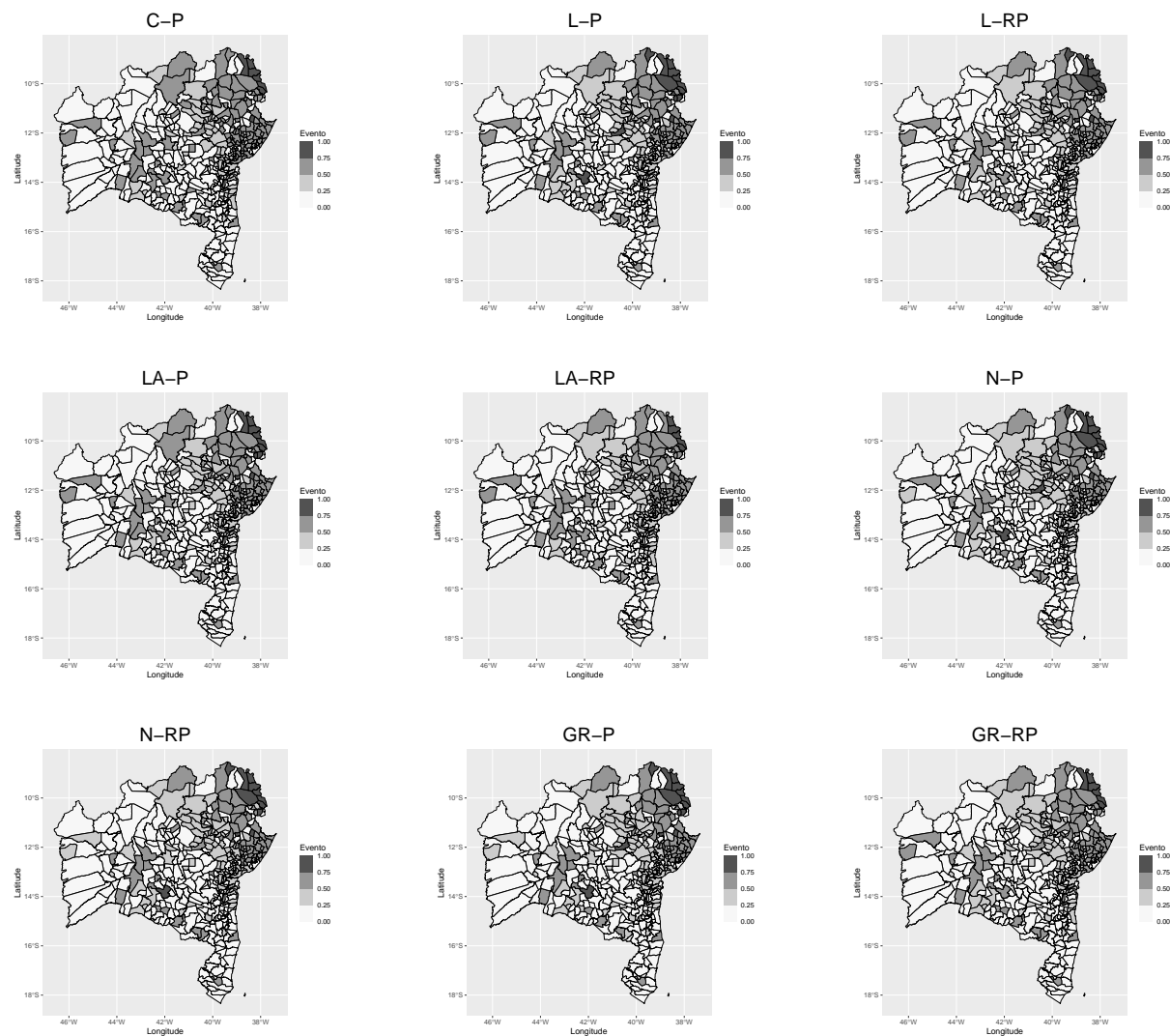
Figura 29 – Gráficos de traço e autocorrelação dos modelos ajustados.



Fonte: Elaborada pelo autor.

D.5 Probabilidades Preditivas

Figura 30 – Probabilidades preditivas para cada município da Bahia, sob os modelos C-P, N-P, N-RP, L-P, L-RP, LA-P, LA-RP e GR-RP. Modelos ajustados sob priori G-Wishart (GW-E).



Fonte: Elaborada pelo autor.

D.6 Estudo de Simulação - Modelo CAR

Abaixo segue um estudo de simulação para o modelo espacial proposto, sob as funções de ligação adotadas, ao estilo do estudo de simulação apresentado na Seção 3.4, porém sob distribuição a priori CAR.

Tabela 30 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $m = 200$ e $m = 500$. 5000 iterações, descartando 2500 como *burn-in*. Modelos: C-P, C-RP, GR-P, GR-RP e LA-P.

Ligação	Parâmetro	$n = 200$					$n = 500$						
		Rbias	RMSE	NRMSE	CP	Tempo médio	Rbias	RMSE	NRMSE	CP	Tempo médio		
C-P	$\lambda = 0,5$	β_1	0.256	0.331	-0.527	0.95	76 (8.23)	β_1	0.107	0.313	-0.565	0.92	69.49 (58.84)
		β_2	0.134	0.368	0.649	0.85	$\lambda = 0,5$	β_2	0.137	0.348	0.613	0.87	
		δ	0.083	0.187	-0.249	0.91	δ	0.062	0.176	-0.239	0.93		
	$\lambda = 4$	β_1	0.125	0.28	-0.497	0.88	71.83 (10.31)	β_1	0.068	0.274	-0.513	0.89	54.64 (8.51)
		β_2	0.116	0.26	0.466	0.92	$\lambda = 4$	β_2	0.116	0.232	0.415	0.97	
		δ	0.029	0.182	0.127	0.94	δ	0.043	0.21	0.145	0.92		
C-RP	$\lambda = 0,5$	β_1	0.102	0.291	-0.529	0.95	19 (3.08)	β_1	0.119	0.281	-0.502	0.95	17.14 (3.55)
		β_2	0.166	0.287	0.492	0.97	$\lambda = 0,5$	β_2	0.211	0.315	0.52	0.93	
		δ	0.054	0.152	-0.209	0.94	δ	0.048	0.156	-0.215	0.93		
	$\lambda = 4$	β_1	0.147	0.242	-0.423	0.93	19.03 (2.7)	β_1	0.146	0.239	-0.416	0.94	21.43 (7.77)
		β_2	0.184	0.259	0.437	0.93	$\lambda = 4$	β_2	0.115	0.23	0.412	0.94	
		δ	0.034	0.241	0.168	0.87	δ	0.03	0.19	0.133	0.93		
GR-P	$\lambda = 0,5$	β_1	-0.12	0.187	-0.424	0.88	27 (3.73)	β_1	-0.07	0.176	-0.379	0.9	26.56 (4.8)
		β_2	-0.021	0.155	0.316	0.96	$\lambda = 0,5$	β_2	-0.132	0.159	0.366	0.97	
		δ	0.034	0.212	-0.296	0.95	δ	-0.011	0.172	-0.251	0.98		
	$\delta = 4$	β_1	0.021	0.109	-0.213	0.97	26.21 (4.1)	β_1	0.014	0.118	-0.232	0.94	25.64 (4.21)
		β_2	0.023	0.108	0.211	0.95	$\lambda = 4$	β_2	-0.023	0.113	0.232	0.94	
		δ	0.062	0.382	0.259	0.86	δ	0.079	0.37	0.247	0.87		
GR-RP	$\lambda = 0,5$	β_1	-0.065	0.148	-0.318	0.98	19 (0.53)	β_1	-0.005	0.151	-0.303	0.98	19.09 (2.41)
		β_2	-0.091	0.182	0.4	0.92	$\lambda = 0,5$	β_2	-0.049	0.17	0.358	0.96	
		δ	0.02	0.225	-0.319	0.91	δ	0.122	0.205	-0.263	0.91		
	$\lambda = 4$	β_1	-0.023	0.118	-0.241	0.96	19.42 (0.42)	β_1	-0.009	0.1	-0.203	0.96	19.33 (0.32)
		β_2	0.018	0.104	0.203	0.96	$\lambda = 4$	β_2	-0.012	0.109	0.221	0.97	
		δ	-0.025	0.302	0.223	0.9	δ	-0.02	0.275	0.203	0.94		
LA-P	$\lambda = 0,5$	β_1	0.083	0.209	-0.385	0.95	19 (0.21)	β_1	0.184	0.268	-0.452	0.91	19.45 (2.36)
		β_2	0.091	0.225	0.412	0.92	$\lambda = 0,5$	β_2	0.116	0.189	0.339	0.97	
		δ	0.03	0.191	-0.268	0.94	δ	0.09	0.202	-0.267	0.9		
	$\lambda = 4$	β_1	0.046	0.156	-0.298	0.93	19.36 (0.22)	β_1	0.117	0.166	-0.298	0.95	19.25 (0.18)
		β_2	0.045	0.162	0.309	0.93	$\lambda = 4$	β_2	0.082	0.179	0.33	0.94	
		δ	0.035	0.295	0.206	0.85	δ	0.026	0.235	0.166	0.95		

Fonte: Elaborada pelo autor.

Tabela 31 – Viés relativo (Rbias); raiz do erro quadrático médio (RMSE); raiz do erro quadrático médio normalizado (NRMSE); probabilidade de cobertura (CP) e tempo computacional médio em segundos (desvio padrão) para cada modelo, baseado em 100 replicações para tamanho de amostras $m = 200$ e $m = 500$. 5000 iterações, descartando 2500 como *burn-in*. Modelos: L-P, L-RP, N-P, N-RP e LA-RP

Ligação	$n = 200$						$n = 500$						
	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	Parâmetro	Rbias	RMSE	NRMSE	CP	Tempo médio	
L-P	$\lambda = 0,5$	β_1	0.004	0.229	-0.456	0.95	21 (0.08)	β_1	0.008	0.228	-0.452	0.97	
		β_2	0.114	0.238	0.428	0.95		β_2	0.118	0.252	0.45	0.94	21.75 (2.75)
		δ	0.098	0.169	-0.223	0.95		δ	0.048	0.169	-0.232	0.97	
	$\lambda = 4$	β_1	0.055	0.19	-0.36	0.93		β_1	0.022	0.201	-0.393	0.92	
		β_2	0.001	0.203	0.405	0.9	21.45 (0.06)	β_2	0.171	0.219	0.374	0.9	21.46 (0.07)
		δ	0.023	0.2	0.141	0.94		δ	0.047	0.219	0.151	0.9	
L-RP	$\lambda = 0,5$	β_1	0.057	0.288	-0.545	0.89	22 (0.15)	β_1	0.156	0.301	-0.52	0.88	
		β_2	0.064	0.223	0.42	0.97		β_2	0.11	0.26	0.468	0.95	22.22 (2.69)
		δ	0.048	0.156	-0.215	0.96		δ	0.045	0.167	-0.231	0.97	
	$\lambda = 4$	β_1	0.057	0.205	-0.389	0.92		β_1	0.093	0.169	-0.309	0.96	
		β_2	0.066	0.183	0.344	0.94	21.92 (0.25)	β_2	0.075	0.198	0.369	0.94	21.83 (0.25)
		δ	0.054	0.221	0.151	0.91		δ	0.021	0.197	0.139	0.93	
N-P	$\lambda = 0,5$	β_1	0.041	0.168	-0.323	0.94	16 (0.22)	β_1	0.092	0.155	-0.284	0.95	
		β_2	0.09	0.151	0.277	0.98		β_2	0.042	0.154	0.295	0.96	16.28 (2.07)
		δ	0.072	0.221	-0.297	0.97		δ	0.052	0.213	-0.293	0.95	
	$\lambda = 4$	β_1	0.007	0.128	-0.254	0.96		β_1	0.047	0.142	-0.272	0.95	
		β_2	0.045	0.131	0.251	0.97	16.15 (0.27)	β_2	0.026	0.134	0.262	0.97	16.18 (0.31)
		δ	0.033	0.293	0.204	0.95		δ	0.047	0.356	0.245	0.89	
N-RP	$\lambda = 0,5$	β_1	0.059	0.181	-0.343	0.93	25 (3.82)	β_1	0.048	0.185	-0.353	0.89	
		β_2	0.074	0.154	0.287	0.97		β_2	0.061	0.179	0.338	0.93	23.87 (5.65)
		δ	0.097	0.198	-0.261	0.96		δ	0.078	0.237	-0.317	0.92	
	$\lambda = 4$	β_1	0.002	0.115	-0.23	0.98		β_1	-0.038	0.123	-0.256	0.94	
		β_2	-0.017	0.117	0.238	0.96	23.24 (3.93)	β_2	-0.017	0.111	0.227	0.98	23.88 (4)
		δ	-0.079	0.221	0.173	0.96		δ	-0.067	0.192	0.149	0.99	
LA-RP	$\lambda = 0,5$	β_1	0.085	0.202	-0.372	0.95	19.1888 (0.16)	β_1	0.081	0.209	-0.387	0.94	
		β_2	0.08	0.209	0.388	0.96		β_2	0.039	0.206	0.396	0.95	19.42 (2.42)
		δ	0.096	0.229	-0.301	0.84		δ	0.081	0.212	-0.283	0.88	
	$\lambda = 4$	β_1	0.051	0.136	-0.259	0.97		β_1	0.043	0.147	-0.282	0.96	
		β_2	0.046	0.166	0.318	0.94	19.19 (0.17)	β_2	0.039	0.149	0.286	0.96	19.09 (0.13)
		δ	0.001	0.263	0.19	0.88		δ	0.002	0.231	0.167	0.92	

Fonte: Elaborada pelo autor.

CAPÍTULO 4

E.1 Comparação de Modelos segundo Métricas de Qualidade da Predição

Os resultados da Tabela a seguir são para os modelos Logito, Probit e Cloglog.

Tabela 32 – Comparação de modelos segundo métricas de qualidade da predição, usando diferentes funções de ligação: Logito, Probit e Cloglog, sob os esquemas de amostragem: i) amostrando da distribuição G-Wishart pelo método exato (GW-E); ii) e atribuindo priori CAR para a estrutura espacial (CAR). P_{cmax} - ponto de corte estimado.

Métrica	Logito		Probit		Cloglog	
	GW-E	CAR	GW-E	CAR	GW-E	CAR
P _{cmax}	0.500	0.150	0.520	0.030	0.640	0.100
ACC	0.518	0.543	0.799	0.550	0.738	0.515
SN	0.692	0.505	0.335	0.510	0.508	0.547
ES	0.502	0.546	0.841	0.554	0.758	0.512
PRC	0.112	0.091	0.160	0.094	0.160	0.092
FPR	0.498	0.454	0.159	0.446	0.242	0.488
CSI	0.106	0.084	0.121	0.086	0.138	0.085
SSI	0.056	0.044	0.065	0.045	0.074	0.045
FAITH	0.288	0.292	0.413	0.296	0.390	0.280
KAPPA	0.058	0.017	0.117	0.021	0.134	0.018
MCC	0.107	0.028	0.128	0.036	0.167	0.033
AUC	0.110	0.086	0.125	0.081	0.068	0.088
F1	0.192	0.155	0.216	0.158	0.243	0.158

Fonte: Elaborada pelo autor.

E.2 Modelos G-AT e GR-AT Aplicados aos Dados de Período

Nesta Seção, apresentamos uma comparação de modelos segundo critérios de bondade de ajuste, para as ligações assimétricas transmutadas apresentadas na Seção D.2. Os resultados se encontram na Tabela 33. Pelos resultados, percebe-se que os modelos assimétricos têm melhor desempenho que os modelos tradicionais Logito, Probit e Cloglog, segundo os critérios adotados. Porém, os modelos potência e reversa de potência apresentaram melhores resultados (veja Tabela 19).

Tabela 33 – Comparação de modelos usando as ligações G-AT e GR-AT

	Logito	Probit	Cloglog	G-AT	GR-AT
WAIC	67184.47	67485.64	93117.61	29092.64	53348.53
DIC	67068.71	67378.05	93068.05	28849.86	53256.73
EBIC	67208.19	67509.28	93142.37	29116.89	53373.32
EAIC	67189.87	67490.96	93124.06	29098.58	53355.01
CPO	-33592.24	-33742.82	-46558.80	-14546.32	-26674.26
Lm	24144.205	24147.553	29499.396	9569.954	20199.742

Fonte: Elaborada pelo autor.

CAPÍTULO 5

Neste Apêndice, apresentamos as expressões dos gradientes necessários para as implementações do algoritmo do modelo DAGAR, para cada tipo de função de ligação utilizada. [Datta *et al.* \(2019\)](#) apresentam o modelo DAGAR inserindo um parâmetro a mais, a precisão marginal, que denotaremos aqui de τ_ϕ .

Como apresentado em [Datta *et al.* \(2019\)](#), assumindo uma distribuição a priori $\text{Gamma}(a, b)$ para τ_ϕ , temos que a distribuição log-posteriori resultante seria

$$\begin{aligned}
 \log p(\tau_\phi|\mathbf{\Omega}) &\propto \log p(\tau_\phi) + \log \prod_{s=1}^r [p(\phi_s|\mathbf{\Omega})] \\
 &\propto (a-1)\log(\tau_\phi) - b\tau_\phi + \log \prod_{s=1}^r \left[\frac{|\tau_\phi \mathbf{\Omega}|^{1/2}}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2} \phi_s^\top \tau_\phi \mathbf{\Omega} \phi_s\right\} \right] \\
 &\propto (a-1)\log(\tau_\phi) - b\tau_\phi + \frac{r \times m}{2} \log \tau_\phi - \frac{\tau_\phi}{2} \sum_{s=1}^r \phi_s^\top \mathbf{\Omega} \phi_s, \tag{F.1}
 \end{aligned}$$

em que $|\cdot|$ seria o determinante da matriz, e na expressão acima, $|\tau_\phi \mathbf{\Omega}| = \tau_\phi^m |\mathbf{\Omega}|$, expressa um resultado decorrente de álgebra matricial. O gradiente da log-posteriori em F.1 é dado por

$$\begin{aligned}
 \frac{\partial \log p(\tau_\phi|\mathbf{\Omega})}{\partial \tau_\phi} &= \frac{a-1}{\tau_\phi} - b + \frac{m \times r}{2\tau_\phi} - \frac{\sum_{i=1}^m \phi_i^\top \mathbf{\Omega} \phi_i}{2} \\
 &= \frac{2(a-1) + m \times r}{2\tau_\phi} - b - \frac{\sum_{i=1}^m \phi_i^\top \mathbf{\Omega} \phi_i}{2}. \tag{F.2}
 \end{aligned}$$

De [Datta *et al.* \(2019\)](#), $|\mathbf{\Omega}| = |(\mathbf{I} - \mathbf{B})^\top \mathbf{F}(\mathbf{I} - \mathbf{B})| = (1 - \rho^2)^m / \prod_{i=1}^m (1 + (n_{\pi(i)} - 1)\rho^2)$, e atribuindo uma distribuição a priori $U(0, 1)$ para ρ , então a log-posteriori de ρ corresponde a

$$\begin{aligned}
 \log p(\rho|\mathbf{\Omega}) &\propto \log(1) + \log \prod_{s=1}^r \left[\frac{|\tau_\phi \mathbf{\Omega}|^{1/2}}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2} \phi_s^\top \tau_\phi \mathbf{\Omega} \phi_s\right\} \right] \\
 &\propto \frac{m}{2} [r \log(1 - \rho^2) - \sum_{s=1}^r \log(1 + (n_{\pi(i)} - 1)\rho^2)] - \frac{\tau_\phi}{2} \sum_{s=1}^r \phi_s^\top \mathbf{\Omega} \phi_s, \tag{F.3}
 \end{aligned}$$

e a respectiva derivada é dada por

$$\begin{aligned} \frac{\partial \log p(\rho|\Omega)}{\partial \rho} &= -\frac{m}{2} \left(\frac{r}{1-\rho^2}(-2\rho) - \frac{\sum_{i=1}^m 2\rho(n_{\pi(i)}-1)}{1+(n_{\pi(i)}-1)\rho^2} \right) - \frac{\tau_\phi}{2} \sum_{i=1}^m \phi_i^\top \frac{\partial \Omega}{\partial \rho} \phi_i \\ &= m\rho \left(\frac{r}{1-\rho^2} + \frac{\sum_{i=1}^m (n_{\pi(i)}-1)}{1+(n_{\pi(i)}-1)\rho^2} \right) - \frac{\tau_\phi}{2} \sum_{i=1}^m \phi_i^\top \frac{\partial \Omega}{\partial \rho} \phi_i. \end{aligned} \quad (\text{F.4})$$

Da expressão acima, temos que

$$\begin{aligned} \frac{\partial \Omega}{\partial \rho} &= \frac{\partial(\mathbf{I} - \mathbf{B})^\top \mathbf{F}(\mathbf{I} - \mathbf{B})}{\partial \rho} \\ &= \frac{(\mathbf{I} - \mathbf{B})^\top \partial[\mathbf{F}(\mathbf{I} - \mathbf{B})]}{\partial \rho} + \frac{(\mathbf{I} - \mathbf{B})^\top \mathbf{F} \partial[(\mathbf{I} - \mathbf{B})]}{\partial \rho} \\ &= (\mathbf{I} - \mathbf{B})^\top \frac{\partial \mathbf{F}}{\partial \rho} (\mathbf{I} - \mathbf{B}) + (\mathbf{I} - \mathbf{B})^\top \mathbf{F} \frac{\partial(\mathbf{I} - \mathbf{B})}{\partial \rho} + \frac{\partial(\mathbf{I} - \mathbf{B})}{\partial \rho} \mathbf{F}(\mathbf{I} - \mathbf{B}). \end{aligned} \quad (\text{F.5})$$

Então, sabendo que $\mathbf{F} = \text{diag}(\tau_1, \dots, \tau_m)$ e que $b_{ij} = b_i$, para $i = 1, \dots, m$, a partir de 5.3, temos

$$\begin{aligned} \frac{\partial \tau_i}{\partial \rho} &= \frac{\partial}{\partial \rho} \left[\frac{(1 + (n_{<i} - 1)\rho^2)}{1 - \rho^2} \right] = \frac{2\rho(n_{<i} - 1)(1 - \rho^2) - [-2\rho(1 + (n_{<i} - 1)\rho^2)]}{(1 - \rho^2)^2} \\ &= \frac{2\rho(n_{<i} - 1)(1 - \rho^2) + 2\rho(1 + 2\rho(n_{<i} - 1)\rho^2)}{(1 - \rho^2)^2} = \frac{2\rho(n_{<i} - 1)(1 - \rho^2) + 2\rho + 2\rho^3(n_{<i} - 1)}{(1 - \rho^2)^2} \\ &= \frac{2\rho(n_{<i} - 1) - 2\rho^3(n_{<i} - 1) + 2\rho + 2\rho^3(n_{<i} - 1)}{(1 - \rho^2)^2} = \frac{2\rho(n_{<i} - 1) + 2\rho}{(1 - \rho^2)^2} \\ &= \frac{2\rho \times n_{<i}}{(1 - \rho^2)^2} \text{ para } i = 1, \dots, m, \end{aligned} \quad (\text{F.6})$$

e

$$\begin{aligned} \frac{\partial b_i}{\partial \rho} &= \frac{1 \times (1 + (n_{<i} - 1)\rho^2) - (\rho(2\rho(n_{<i} - 1)))}{(1 + (n_{<i} - 1)\rho^2)^2} \\ &= \frac{1 + (n_{<i} - 1)\rho^2 - 2\rho^2(n_{<i} - 1)}{(1 + (n_{<i} - 1)\rho^2)^2} = \frac{1 - (n_{<i} - 1)\rho^2}{(1 + (n_{<i} - 1)\rho^2)^2} \text{ para } i = 2, \dots, m; j \in N(i). \end{aligned} \quad (\text{F.7})$$

