

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Modelos estocásticos de transmissão para análises genéticas de características epidemiológicas

Milena Nascimento Lima

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

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Milena Nascimento Lima

Modelos estocásticos de transmissão para análises genéticas de características epidemiológicas

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

Área de Concentração: Estatística

Orientador: Prof. Dr. Osvaldo Anacleto Junior

Coorientador: Profa. Dra. Daiane Aparecida Zuanetti

USP/UFSCar – São Carlos
Setembro de 2023

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

L732m Lima, Milena Nascimento
Modelos estocásticos de transmissão para
análises genéticas de características epidemiológicas
/ Milena Nascimento Lima; orientador Osvaldo
Anacleto Junior; coorientadora Daiane Aparecida
Zuanetti. -- São Carlos, 2023.
92 p.

Tese (Doutorado - Programa Interinstitucional de
Pós-graduação em Estatística) -- Instituto de Ciências
Matemáticas e de Computação, Universidade de São
Paulo, 2023.

1. Inferência Bayesiana. 2. Modelos
estocásticos. 3. Genética Quantitativa. 4.
Características epidemiológicas. 5. Melhoramento
genético. I. Junior, Osvaldo Anacleto, orient. II.
Zuanetti, Daiane Aparecida, coorient. III. Título.

Milena Nascimento Lima

**Stochastic transmission models for genetic analyses of
epidemiological traits**

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 Doctorate Interagency Graduate Program in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Osvaldo Anacleto Junior

Co-advisor: Profa. Dra. Daiane Aparecida Zuanetti

**USP/UFSCar – São Carlos
September 2023**



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 Milena Nascimento Lima, realizada em 01/08/2023.

Comissão Julgadora:

Prof. Dr. Osvaldo Anacleto Junior (ICMC/USP)

Prof. Dr. Carlos Tadeu Pagani Zanini (UFRJ)

Profa. Dra. Teresa Cristina Martins Dias (UFSCar)

Profa. Dra. Lia Hanna Martins Morita (UFMT)

Prof. Dr. Francisco Louzada Neto (USP)

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

Para minha mãe.

AGRADECIMENTOS

Prezado(a)s leitor(a)s,

Considerando as emoções, nunca antes vividas, que se apresentam simultaneamente com a realização desta experiência, peço licença para utilizar neste espaço uma linguagem mais pessoal.

Começo agradecendo a Deus que, tendo sempre planos maiores e melhores que os meus, me concedeu o privilégio de ingressar nas universidades *UFSCar* e *USP* como aluna de doutorado e estabeleceu as condições essenciais para que esta tese fosse desenvolvida.

Em sequência, agradeço a minha família: Anita, Benedito, Roberto, Cássia e Alisson, Renê, Luzia e Débora, que acreditaram, torceram e contribuíram como puderam para que eu avançasse nos estudos.

Agora os agradecimentos acadêmicos. Bom pessoal, alcançar esta linha de chegada significa que um caminho longo, árduo e transformador foi percorrido. E eu me encontro feliz, principalmente porque não estive sozinha na caminhada. Tive a honra de ser acompanhada por dois professores incríveis, Osvaldo Anacleto e Daiane Zuanetti, que me orientaram em perfeita conduta pedagógica e humana. Cada planejamento feito, sugestões dadas, testes propostos, correções realizadas, helps, reuniões, encorajamentos, empatia, contribuíram para que eu prosperasse em cada etapa da jornada e me tornasse a protagonista no cenário da aprendizagem. Neste parágrafo não cabem todas as palavras que possam externar o quanto sou grata aos dois. Certamente passaria horas elogiando eles para vocês. Então eu resumo tudo nesta frase: Muito obrigada, Osvaldo e Daiane! Aproveito para agradecer também a todos os professores que passaram pela minha vida, desde a alfabetização até os professores do *PIPGES* e membros desta banca examinadora.

Agradeço ainda aos demais colaboradores do programa: (i) Coordenadores e secretários, por toda dedicação e cuidado para que tudo esteja sempre funcionando perfeitamente e assim possamos nos dedicar melhor à pesquisa; (ii) aos colegas de curso e de laboratório, pelas ideias trocadas, helps e momentos de descontração; (iii) e à *CAPES*, por ter financiado este trabalho.

E assim alcançamos a última etapa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (*CAPES*) - Código de Financiamento 001.

“Ergue do pó o necessitado e do monte de cinzas faz ressurgir o abatido; Ele os faz assentar-se com príncipes e lhes concede um lugar de honra, porque ao SENHOR pertencem os fundamentos da terra, e sobre eles estabeleceu o mundo.”
(1º Samuel 2:8 ; Salmos 113:7)

RESUMO

LIMA, M. N. **Modelos estocásticos de transmissão para análises genéticas de características epidemiológicas**. 2023. 92 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, 2023.

As epidemias podem comprometer significativamente a produção animal e gerar grandes impactos econômicos. Além disso, as atuais práticas de tratamento e controle de doenças infecciosas em animais de criação nem sempre apresentam a eficácia desejada. Nesses casos, o melhoramento genético oferece uma alternativa viável através do estudo das variações genéticas das características do hospedeiro que afetam a transmissão da doença, especialmente a susceptibilidade e a infectividade. No entanto, apesar dos avanços, o desafio para estimar efeitos genéticos que controlam a infectividade, principalmente, continua a existir. Sendo assim, o objetivo geral nesta tese foi contribuir com o estado da arte no desenvolvimento de modelos estatísticos que possam capturar a dinâmica da transmissão de doenças infecciosas e consequentemente aprimorar a estimação de efeitos genéticos na infectividade. Apresentamos uma nova versão do modelo dinâmico não linear de efeitos genéticos indiretos (*dynamic non-linear indirect genetic effects model - dnIGE*) e um método inferencial para estimar seus parâmetros. Nossa metodologia inclui uma estrutura de covariância na distribuição dos efeitos genéticos e ambientais da susceptibilidade e infectividade, que antes eram consideradas independentes, e utiliza técnicas computacionais Bayesianas modernas para estimar os efeitos genéticos e herdabilidades associados a estas características. Os resultados mostram que o modelo *dnIGE* estendido pode estimar com precisão, herdabilidades e valores genéticos associados à susceptibilidade e infectividade, mesmo quando existe uma correlação genética entre essas características. Nossa metodologia proposta oferece impactos potenciais em áreas como o controle de doenças na pecuária por meio de reprodução seletiva e também na previsão e controle do surgimento de surtos de doenças em populações humanas.

Palavras-chave: Doenças Infecciosas, Susceptibilidade, Infectividade, Genética quantitativa, Inferência Bayesiana.

ABSTRACT

LIMA, M. N. **Stochastic transmission models for genetic analyses of epidemiological traits.** 2023. 92 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, 2023.

Epidemics can significantly affect animal production and generate large economic impacts. Furthermore, current practices for treating and controlling infectious diseases in farmed animals do not always show the desired effectiveness. In these cases, the quantitative genetics offers a viable alternative through the study of genetic variations of host characteristics that affect disease transmission, especially susceptibility, and infectivity. However, despite advances, the challenge to estimate genetic effects that mainly control infectivity continues to exist. Therefore, the general objective of this thesis was to contribute to the state of the art in the development of statistical models that can capture the dynamics of transmission of infectious diseases and consequently improve the estimation of genetic effects on infectivity. We present a new version of the *dynamic non-linear indirect genetic effects model (dnIGE)* and an inferential method to estimate its parameters. Our methodology includes a covariance structure on the distribution of genetic and environmental effects of susceptibility and infectivity, which were previously considered independent and uses modern Bayesian inference to estimate the genetic effects and heritabilities associated with these traits. Results show that the extended *dnIGE* model can accurately estimate heritabilities and genetic values associated with susceptibility and infectivity, even when there is a genetic correlation between these traits. Our proposed methodology offers potential impacts in areas such as disease control in livestock through selective breeding and also in predicting and controlling the emergence of disease outbreaks in human populations.

Keywords: Infectious diseases, Susceptibility, Infectivity, Quantitative Genetics, Bayesian Inference.

LISTA DE ILUSTRAÇÕES

Figura 1 – Soluções numéricas para um modelo <i>SIR</i> determinístico com $\beta = 0.83$ e $\gamma = 0.12$	31
Figura 2 – Interações entre um animal susceptível com <i>I</i> infectados.	36
Figura 3 – Estrutura dos dados de infecção considerados para o desenvolvimento do modelo <i>dnIGE</i>	42
Figura 4 – Grupos independentes de tamanho 3 representando as diferentes interações entre os animais.	44
Figura 5 – Representação hierárquica do modelo para grupo de tamanho 2.	45
Figura 6 – Intervalos de credibilidade de 90% para herdabilidades de susceptibilidade para grupos de tamanho 10 (cenário base). As linhas tracejadas indicam herdabilidades verdadeiras e os pontos representam as médias a posteriori.	58
Figura 7 – Intervalos de credibilidade de 90% para herdabilidades de infectividade para grupos de tamanho 10 (cenário base). As linhas tracejadas indicam herdabilidades verdadeiras e os pontos representam as médias a posteriori.	59
Figura 8 – Influência das variações genéticas da infectividade nas estimativas do modelo <i>dnIGE</i> em comparação com a abordagem alternativa de Korsgaard, Madsen e Jensen (1998).	60
Figura 9 – Efeito do tamanho e do tipo de composição dos grupos nas estimativas do modelo <i>dnIGE</i>	62
Figura 10 – Intervalos de credibilidade para a herdabilidade da susceptibilidade quando o método de alocação é <i>random</i> e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.	65
Figura 11 – Intervalos de credibilidade para a herdabilidade da susceptibilidade quando o método de alocação é <i>2FAM</i> e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.	66
Figura 12 – Intervalos de credibilidade para a herdabilidade da infectividade quando o método de alocação é <i>random</i> e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.	67
Figura 13 – Intervalos de credibilidade para a herdabilidade da infectividade quando o método de alocação é <i>2FAM</i> e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.	68

Figura 14 – Intervalos de credibilidade para o coeficiente de correlação genética quando o método de alocação é <i>random</i> e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.	69
Figura 15 – Intervalos de credibilidade para o coeficiente de correlação genética quando o método de alocação é <i>2FAM</i> e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.	70
Figura 16 – Efeito das correlações genéticas nas acurácias de predição do modelo <i>dnIGE</i> estendido.	71
Figura 17 – Comparação entre os métodos de alocação em diversas correlações genéticas.	72

LISTA DE TABELAS

Tabela 1 – Notações referentes ao método de alocação <i>2FAM</i>	52
Tabela 2 – Configurações do exemplo.	53
Tabela 3 – Combinação de famílias para formação de 10 grupos de tamanho 4 em um bloco.	53
Tabela 4 – Configurações da população de base nos métodos de alocação.	55
Tabela 5 – Valores de herdabilidade, tamanho da amostra, e tamanho do grupo considerados nas epidemias simuladas para validação do modelo <i>dnIGE</i> original e estendido.	56
Tabela 6 – Configurações para o modelo <i>dnIGE</i> estendido.	63

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Desafios estatísticos em genética quantitativa de doenças infecciosas	23
1.2	Objetivos e contribuições da tese	24
1.3	Estrutura da tese	24
2	CONCEITOS DE GENÉTICA QUANTITATIVA E DOENÇAS INFECCIOSAS	25
2.1	O modelo básico da Genética Quantitativa	25
2.1.1	<i>Componentes de variação</i>	26
2.2	Genética Quantitativa de doenças infecciosas: panorama e alguns conceitos	27
3	MODELOS EPIDÊMICOS	29
3.1	Modelos epidêmicos: um breve histórico e aplicações	29
3.2	O modelo <i>SIR</i>	30
3.3	Formulação estocástica do modelo <i>SIR</i>	32
3.3.1	<i>Processo de Poisson</i>	33
3.3.2	<i>A transição de S para I</i>	35
3.3.3	<i>A transição de I para R</i>	36
3.3.4	<i>Evolução e espalhamento da epidemia</i>	37
4	MODELOS ESTATÍSTICOS PARA ANÁLISES GENÉTICAS DE INFECTIVIDADE ANIMAL	39
4.1	O modelo dinâmico não linear de efeitos genéticos indiretos (<i>dnIGE</i>)	40
4.1.1	<i>O Modelo</i>	42
4.1.2	<i>Componentes de variância para a susceptibilidade (g_j) e infectividade (f_j)</i>	44
4.2	O modelo <i>dnIGE</i> estendido	46
4.2.1	<i>Função de verossimilhança</i>	47
4.2.2	<i>Distribuição conjunta a posteriori</i>	49
5	EFEITO DE DIFERENTES DESENHOS EXPERIMENTAIS NA QUALIDADE DAS ESTIMATIVAS DO MODELO <i>dnIGE</i>	51
5.1	Simulação das populações de base	54

5.2	Software para estimação dos parâmetros do modelo <i>dnIGE</i>	56
5.3	Resultados do modelo <i>dnIGE</i> original	57
5.3.1	<i>Estimativas de herdabilidade</i>	57
5.3.2	<i>Estimativas de acurácia de predição</i>	59
5.3.3	<i>Efeitos de tamanho do grupo e herdabilidade nas estimativas de acurácias de predição</i>	60
5.4	Resultados do modelo <i>dnIGE</i> estendido	63
5.4.1	<i>Estimativas de herdabilidade</i>	64
5.4.2	<i>Efeito das correlações genéticas entre susceptibilidade e infectividade nas acurácias de predição</i>	71
6	CONSIDERAÇÕES FINAIS	73
6.1	Síntese dos principais resultados	73
6.2	Contribuições teóricas	73
6.3	Contribuições para o melhoramento genético de características epidemiológicas	74
6.4	Recomendações para investigações futuras	74
	REFERÊNCIAS	77
APÊNDICE A	DISTRIBUIÇÕES A POSTERIORI CONDICIONAIS COMPLETAS PARA INFERÊNCIA DO VETOR DE PARÂMETROS DO MODELO DNIGE ESTENDIDO .	85
A.1	Condicionais completas das variâncias genéticas	86
A.2	Condicionais completas das variâncias ambientais	87
A.3	Condicionais completas dos efeitos genéticos	89
A.4	Condicionais completas dos efeitos ambientais	90
A.5	Condicionais completas dos coeficientes de correlação	92
A.6	Condicional completa da taxa de contato efetivo β	92

INTRODUÇÃO

Nesta tese de doutorado o foco é na modelagem da transmissão de doenças infecciosas em animais de criação. Surtos de doenças infecciosas podem comprometer significativamente a produção animal, resultando em consideráveis perdas econômicas e ameaças persistentes à sociedade.

A literatura epidemiológica é abundante em nos mostrar exemplos sobre isso. Por exemplo, a peste suína africana, causada pelo *Asfvirus*, ressurgiu na China em 2018, se expandiu e prevalece nos continentes europeu, africano e asiático. Desde então milhões de animais já foram exterminados, incluindo os abates preventivos, o que evidencia a capacidade do vírus gerar graves perdas produtivas e econômicas (ZHAI *et al.*, 2019; CHEN *et al.*, 2022). Apesar de ter sido erradicada no Brasil em 1981, a doença é considerada de grande importância devido sua alta transmissibilidade e ausência de vacina. Em uma estimativa feita pela Embrapa (2019), uma possível chegada dessa enfermidade ao Brasil poderia acarretar num prejuízo de 5,5 bilhões de dólares em apenas um ano de epidemia. A gripe aviária H5N1, que reapareceu em 2003, se espalhou da Ásia para Europa e África resultando em enormes impactos na indústria avícola. Em alguns países do sudeste asiático, onde a doença é endêmica, as perdas econômicas totais chegaram a quase 8 bilhões de euros (CHANG *et al.*, 2007; EUROPEAN COMMISSION, 2006). A febre aftosa, doença infecciosa causada pelo aftovírus da família *Picornaviridae*, assolou os Estados brasileiros Mato Grosso do Sul e Paraná durante o surto que ocorreu em 2005. A ocorrência desta doença resultou no descarte de vários animais, gerando impactos negativos nas exportações de carne *in natura* provenientes destes Estados diretamente afetados (GARCIA *et al.*, 2015). De acordo com o MAPA | Ministério da Agricultura, Pecuária e Abastecimento (2018), o último surto no Brasil ocorreu em 2016, mas esta doença continua sendo alvo de permanente pesquisa e preocupação mundial (KNIGHT-JONES; RUSHTON, 2013; OIE | Organização Mundial para a Saúde Animal, 2018).

Neste contexto, as atuais práticas de tratamento e controle de epidemias nem sempre

apresentam a eficácia desejada. Alguns fatores que comprometem essa eficácia são, por exemplo, a evolução da resistência do agente patogênico a antibióticos ou vacinas e o surgimento de impactos ambientais indesejáveis como consequência de alguns tratamentos químicos utilizados para o controle da epidemia (GIBSON; BISHOP, 2005). Nesses casos, o melhoramento genético - que é um corpo de conhecimento que depende da Genética Quantitativa e da Estatística para a estimação de parâmetros dos modelos que incorporam as principais propriedades genético-epidemiológicas - oferece uma alternativa viável ao controle de doenças infecciosas através do estudo das variações genéticas das características que influenciam na disseminação da doença (BISHOP *et al.*, 2010). Esta técnica permite a exploração de sequências de *DNA*, mapeando as regiões genéticas que são responsáveis pela expressão das características epidemiológicas presentes nos animais submetidos à propagação da doença. Essa estratégia vem sendo relatada em muitos estudos genéticos e genômicos, e reconhecida como um fator determinante para a seleção artificial (DOESCHL-WILSON *et al.*, 2011; NATH; WOOLLIAMS; BISHOP, 2008). Resultados recentes, inclusive, apontam para a possibilidade da erradicação de doenças infecciosas por meio da seleção genética (HULST; JONG; BIJMA, 2021).

A literatura de doenças infecciosas aponta para duas principais características do hospedeiro que afetam a transmissão das doenças infecciosas: a **susceptibilidade**, que é a propensão de um animal tornar-se o novo hospedeiro após o contato com um animal infectado ou material infeccioso; e a **infectividade**, que pode ser definida como a capacidade do hospedeiro transmitir a infecção por unidade de contato (mais detalhes no Capítulo 2). No entanto, a maioria dos estudos genético-quantitativos de doenças infecciosas se concentram na resistência do hospedeiro (BISHOP *et al.*, 2010), a qual apresenta uma relação inversamente proporcional à susceptibilidade, enquanto que a infectividade, relacionada diretamente ao risco de transmissão (LLOYD-SMITH *et al.*, 2005), é geralmente ignorada. Entretanto, estudar a variação na infectividade pode ser relevante para prever e prevenir surtos futuros de doenças em pecuária.

A estimação de parâmetros de modelos que acomodam as principais propriedades genético-epidemiológicas de hospedeiros, pode permitir a identificação, via arquitetura genética, não apenas dos animais que são altamente resistentes à infecção, mas também dos que possuem alto risco para transmitir a doença infecciosa aos seus companheiros de grupo. Neste sentido, os modelos estatísticos representam um componente crucial das estratégias de componentes genéticos, pois é através deles que é possível estimar os parâmetros genéticos relacionados às características epidemiológicas. Com isto, melhoristas podem definir as estratégias necessárias para a seleção da geração parental e reprodução, prever a resposta à seleção, analisar as eventuais consequências das decisões de acasalamento e dar prosseguimento às demais etapas do programa de melhoramento genético.

Por todos esses aspectos, vemos que os modelos estatísticos podem contribuir para o diagnóstico, controle, prevenção e previsão de surtos futuros de doenças infecciosas em pecuária.

1.1 Desafios estatísticos em genética quantitativa de doenças infecciosas

A análise de dados epidêmicos apresenta vários desafios do ponto de vista estatístico. Por exemplo, a probabilidade de um animal ser infectado durante uma epidemia depende da interação desse animal com outros já infectados, o que resulta em uma alta estrutura de dependência em dados epidêmicos (BECKER; BRITTON, 1999). Enquanto que modelos estocásticos epidêmicos ignoram potenciais variações genéticas entre indivíduos, estudos de genética quantitativa de doenças infecciosas utilizam modelos estatísticos que incorporam variação genética, porém não consideram a dinâmica intrínseca do processo de transmissão de doenças (LIPSCHUTZ-POWELL; WOOLLIAMS; DOESCHL-WILSON, 2014). Isso dificulta o estudo da heterogeneidade genética na infectividade.

A variação individual na infectividade, que é expressa a partir de interações sociais, resulta no conhecido e frequente fenômeno de superespalhamento, que ocorre quando uma minoria de indivíduos altamente infecciosos é responsável pela maioria das transmissões de patógenos durante uma epidemia (LLOYD-SMITH *et al.*, 2005). Neste contexto, considerar a infectividade, além da susceptibilidade, pode trazer maiores benefícios na presença desses indivíduos, pois estes não seriam capturados pela seleção baseada apenas na susceptibilidade. Desta forma, os superespalhadores podem ser identificados, eliminando assim uma importante fonte de infecção para outros animais (TSAIRIDOU *et al.*, 2019).

Um dos primeiros modelos que permite a estimação de efeitos genéticos na infectividade é o modelo dinâmico não-linear de efeitos genéticos indiretos (*dynamic non-linear indirect genetic effects models - dnIGE*)(ANACLETO *et al.*, 2015). Seu desenvolvimento provém da combinação entre modelos de efeitos genéticos indiretos e processos estocásticos para modelar a natureza aleatória, não linear e dinâmica das infecções, incorporando explicitamente o processo de transmissão da doença. Apesar disso, um grande desafio ainda reside em estimar parâmetros genéticos associados a esta característica (ANACLETO *et al.*, 2019).

Assim, percebe-se que embora a importância da utilização da variação genética na infectividade para o controle de doenças em pecuária e aquicultura já tenha sido reconhecida e demonstrada, os atuais modelos genético-quantitativos de doenças infecciosas revelam uma complexidade para estimar com precisão os efeitos genéticos que a controlam, requerendo mais investigações que busquem acomodar essa complexidade envolvida por meio da construção de modelos estatísticos adequados para estimar os parâmetros genéticos associados à mesma. A presente tese está incluída nessas investigações.

1.2 Objetivos e contribuições da tese

O objetivo geral nesta tese consiste em contribuir com o estado da arte no desenvolvimento de modelos estatísticos que possam capturar a dinâmica da transmissão de doenças e consequentemente aprimorar a estimação de efeitos genéticos na infectividade.

Apresentamos uma proposta para incorporar no modelo *dnIGE* uma estrutura de covariância entre susceptibilidade e infectividade, o que configura a originalidade deste estudo. Como será apresentado a seguir, na formulação original do modelo, essas características foram assumidas como independentes.

Além disso, propomos uma avaliação do efeito dos diferentes desenhos experimentais na qualidade das estimativas dos modelos *dnIGE* e *dnIGE* estendido, variando tamanhos de grupo e métodos de composição para maximizar as acurácias de predição.

1.3 Estrutura da tese

Após esta introdução, os Capítulos 2 e 3 apresentam o referencial teórico que sustenta esta tese. O Capítulo 2 revisa os conceitos de genética quantitativa e doenças infecciosas, que são essenciais para a implementação da arquitetura genética da susceptibilidade e infectividade. O Capítulo 3 revisa a metodologia que determina a maneira com que a epidemia se desenvolve. Em sequência, no Capítulo 4, descreve-se o percurso metodológico que estende o modelo *dnIGE*, apresenta-se a metodologia Bayesiana para a estimação de parâmetros neste modelo além de um estudo de simulação para avaliar sua qualidade. No Capítulo 5 avalia-se o efeito dos diferentes desenhos experimentais na qualidade das estimativas do modelo *dnIGE*, nas versões original e estendida, e depois finaliza-se com a conclusão, que consta no Capítulo 6.

CONCEITOS DE GENÉTICA QUANTITATIVA E DOENÇAS INFECCIOSAS

De acordo com [Oldenbroek e Waaij \(2015\)](#), o melhoramento genético consiste na criação seletiva de animais com características hereditárias desejáveis para reprodução, com a intenção de melhorar a característica de interesse na geração seguinte (prole). Esta técnica depende da Genética Quantitativa e de modelos estatísticos para estimar a contribuição do efeito genético na expressão do fenótipo.

Desta forma, apresentamos neste capítulo os conceitos de Genética Quantitativa e de doenças infecciosas, necessários para o desenvolvimento deste trabalho.

2.1 O modelo básico da Genética Quantitativa

Os seres vivos de uma mesma espécie, sendo os animais o foco desse trabalho, são diferentes entre si. As diversas características apresentadas por eles raramente condizem somente com suas informações genéticas, pois dependem de fatores genéticos e ambientais. O ambiente corresponde a um conjunto de fatores referentes ao meio em que os animais são criados, afetando suas características.

Desta forma, o modelo básico da Genética Quantitativa, que é amplamente usado em melhoramento genético, é constituído unicamente pelo genótipo e pelo ambiente sendo descrito pela seguinte expressão matemática

$$P = G + E, \quad (2.1)$$

em que P representa o fenótipo, G é o valor genotípico do animal e E representa os desvios do ambiente ([FALCONER; MACKAY, 1996](#); [LYNCH; WALSH *et al.*, 1998](#)).

Um dos principais objetivos na Genética Quantitativa é quantificar a contribuição de efeitos genéticos e ambientais na expressão de um fenótipo específico. Nesse sentido, a criação de

animais tem o interesse na identificação de genótipos que expressam os fenótipos mais desejáveis sob a perspectiva da produção animal. Existe também uma grande preocupação com a mudança genética de populações em cada geração, pois apenas uma parte das diferenças observadas no desempenho do animal é passível de ser transferida de uma geração à outra.

Assim, o termo correspondente ao valor genotípico no modelo básico da Genética Quantitativa (2.1), possui componentes subjacentes e pode ser modelado da seguinte forma

$$G = A + D + I,$$

em que A é o efeito genético aditivo, também chamado de valor genético, D é o efeito de dominância e I é o efeito de epistasia (FALCONER; MACKAY, 1996). Entretanto, os efeitos D e I serão desconsiderados neste trabalho, pois não são de grande interesse em problemas de melhoramento genético. Neste caso, os efeitos genéticos aditivos e ambientais passam a ser os únicos determinantes do fenótipo. Logo, o modelo básico da Genética Quantitativa (2.1) se reduz a

$$P = A + E. \quad (2.2)$$

É importante ressaltar que as características a serem estudadas nesta tese estão sob efeitos poligênicos, o que significa que estas são controladas por muitos genes que agem conjuntamente com um pequeno efeito individual sobre elas.

2.1.1 Componentes de variação

Uma vez que a geração parental é selecionada, temos o interesse em prever a variação genética a partir da variação fenotípica para avaliar o progresso genético dos animais nas próximas gerações. Neste contexto, a estatística se mostra uma ferramenta crucial para estimar a contribuição do efeito genético na característica de interesse (CARNEIRO-JÚNIOR, 2009).

A partir do modelo básico da Genética Quantitativa (2.2) e da suposição de independência entre os efeitos genéticos aditivos e ambientais, a variância fenotípica pode ser decomposta em

$$\sigma_P^2 = \sigma_A^2 + \sigma_E^2, \quad (2.3)$$

em que σ_P^2 é a variância dos valores fenotípicos, σ_A^2 é a variância dos valores genéticos aditivos e σ_E^2 é a variância dos desvios de ambiente.

A partir desta composição surge a definição que descreve o quanto da variação em uma determinada característica pode ser atribuída à variação genética. A herdabilidade, em sentido restrito, mede a contribuição dos efeitos genéticos aditivos para a variação fenotípica total. Matematicamente, a herdabilidade no sentido restrito é definida por

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}. \quad (2.4)$$

Segundo [Charmantier, Garant e Kruuk \(2014\)](#) a herdabilidade de uma característica situa-se entre 0 e 1, indicando o grau de semelhança genética entre parentes em uma população. Desta forma, $h^2 = 0$ implica que não há variabilidade genética e portanto, $\sigma_p^2 = \sigma_E^2$, o que torna o processo seletivo ineficaz. Por outro lado, se $h^2 = 1$ ou está próxima de 1, então a variação ambiental é desprezível, de modo que a variação fenotípica é o resultado exclusivamente da variabilidade genética entre os indivíduos, ou seja, $\sigma_p^2 = \sigma_A^2$.

2.2 Genética Quantitativa de doenças infecciosas: panorama e alguns conceitos

De acordo com a [WHO | Organização Mundial de Saúde \(2016\)](#), as doenças infecciosas são causadas por microrganismos patogênicos como, por exemplo, vírus, bactérias, protozoários ou helmintos. Estes patógenos, que podem ser microparasitas (pequenos, geralmente unicelulares) ou macroparasitas (de forma maior), são os agentes biológicos que podem ser transmitidos entre os seres vivos. Uma infecção acontece quando os patógenos se instalam e se multiplicam em um organismo hospedeiro, podendo resultar na ocorrência da doença infecciosa. A transmissão de patógenos pode ocorrer, por exemplo, a partir do contato com animais infectados ou contato com material infeccioso expelido por indivíduos infectados. O tipo de transmissão a que este trabalho se refere é a transmissão direta.

A previsão, prevenção, diagnóstico, controle e erradicação das doenças infecciosas baseiam-se na análise dos fatores que influenciam a transmissão do patógeno. Em particular, o interesse em melhoramento genético é explorar a variação genética em características epidemiológicas para produzir animais mais preparados para potenciais surtos epidêmicos. As duas principais características epidemiológicas que afetam a transmissão de doenças infecciosas são descritas a seguir.

Numa população submetida a uma epidemia, existem animais que possuem a habilidade de bloquear ou restringir a replicação de agentes infecciosos. Esta característica é chamada de resistência e possui uma relação inversamente proporcional à susceptibilidade. Assim, uma baixa susceptibilidade está associada à alta resistência à doença e vice-versa. Em muitos casos a susceptibilidade à doença é uma característica poligênica ([BISHOP; WOOLLIAMS, 2014](#)). No entanto, também existem evidências de efeitos de um único gene na resistência como, por exemplo, em Necrose Pancreática Infecciosa (*IPN*), uma doença economicamente relevante em salmão do Atlântico ([ØDEGÅRD *et al.*, 2011](#); [HOUSTON *et al.*, 2008](#); [MOEN *et al.*, 2009](#)).

A definição de resistência varia entre estudos, e sua forma de medição depende da doença, do mecanismo de resposta do hospedeiro que expressa a variação genética e também do planejamento amostral ([BISHOP *et al.*, 2010](#)). Ela pode ser medida, por exemplo, por testes diagnósticos da doença ou mortalidade devido à mesma ([BISHOP; WOOLLIAMS, 2014](#); [BROOKS-POLLOCK *et al.*, 2015](#)). Em melhoramento genético de doenças infecciosas, a

susceptibilidade é considerada uma das principais características por ser mais fácil de ser mensurada em comparação com outras características epidemiológicas.

No entanto, existem evidências teóricas mostrando que as ferramentas de análise genética baseadas em susceptibilidade capturam somente uma parte da variação genética disponível em dados epidemiológicos (BISHOP; DOESCHL-WILSON; WOOLLIAMS, 2012; LIPSCHUTZ-POWELL *et al.*, 2012a; BISHOP; WOOLLIAMS, 2014). Nesse contexto, a infectividade, que é a capacidade do hospedeiro causar infecção em seus companheiros de grupo, é uma característica a ser considerada. Sua relação diretamente proporcional ao risco de transmissão pode levar ao fenômeno conhecido como superespalhamento, que ocorre quando um pequeno grupo de indivíduos altamente infecciosos é responsável pela maioria das infecções.

Por muito tempo foi um desafio modelar variação genética na infectividade pela falta de métodos estatísticos que pudessem separar com precisão as diferentes fontes de variação individual inerentes aos dados de doenças infecciosas. Um dos motivos para isso é que a infectividade não é expressa diretamente como fenótipos usuais, mas sim a partir de interações sociais, o que significa que ela impacta no *status* da doença nos animais que convivem com o animal que a possui e não no animal em si. Estudos teóricos recentes mostram que infectividade pode ser uma característica relevante em estudos de melhoramento genético (RAPHAKA *et al.*, 2018; TSAIRIDOU *et al.*, 2019).

MODELOS EPIDÊMICOS

Como o foco nesta pesquisa é a modelagem da transmissão de doenças sob a ótica de melhoramento genético, neste capítulo descreveremos conceitos e modelos que serão utilizados para representar processos epidêmicos.

3.1 Modelos epidêmicos: um breve histórico e aplicações

Os modelos epidêmicos podem nos ajudar a focar nos aspectos do processo de transmissão da doença infecciosa que são, em certo sentido, fundamentais, na medida em que ajudam a explicar algumas das principais características epidemiológicas em uma determinada população (ANDERSON; MAY, 1979; LEKONE; FINKENSTÄDT, 2006).

Como exemplo, no estudo de doenças infecciosas em humanos, as análises de dados de epidemias provenientes de modelos também podem ter resultados práticos diretos ao sugerir intervenções de saúde pública, ajudando a quantificar seus prováveis impactos, e dar suporte para a avaliação de riscos de doenças infecciosas emergentes, uma vez que fornecem informações sobre tendências (MOGHADAS *et al.*, 2009; DAY *et al.*, 2006). No estudo de doenças infecciosas em animais de criação, os modelos genético-quantitativos de características epidemiológicas apresentam grande relevância para o auxílio no diagnóstico, controle, previsão e prevenção de surtos futuros de doenças (BISHOP *et al.*, 2010; LLOYD-SMITH *et al.*, 2005).

Além disso, uma vantagem imediata é que uma única estrutura de modelagem geralmente se aplica a muitas outras infecções (KEELING; ROHANI, 2011). A modelagem pode ainda ser usada para comparar diferentes doenças na mesma população, a mesma doença em diferentes populações ou em diferentes períodos da história. Estudos comparativos dessas situações podem ser encontrados, por exemplo, em Yorke e London (1973a) e London e Yorke (1973b).

As contribuições iniciais de modelagem para investigar a disseminação de doenças infecciosas eram feitas geralmente para doenças específicas. Bernoulli desenvolveu o primeiro

modelo quando a população europeia foi afetada por uma epidemia de varíola, que se espalhou rapidamente, no século *XVIII*. Com o passar do tempo, outros cientistas também deixaram suas contribuições. Sir Ronald Ross expôs o primeiro modelo de transmissão da malária em 1911. Poucos anos depois, suas ideias foram aplicadas por William Ogilvy Kermack e Anderson Gray McKendrick que estudaram a transmissão de uma doença infecciosa em humanos (FOPPA, 2016; BAILEY *et al.*, 1975). Atualmente existem diversas generalizações desses modelos a fim de torná-los cada vez mais compatíveis com a realidade.

Em geral, os modelos epidêmicos consideram uma população particionada de acordo com os estágios da doença de interesse e cada partição recebe o nome desse estágio (susceptível, infectado, recuperado, entre outros). Estas considerações refletem nas ideias básicas por trás do modelo de Bernoulli que foram e ainda permanecem utilizadas até os dias atuais nos chamados modelos compartimentais (KEELING; ROHANI, 2011), que têm como um dos principais objetivos estimar níveis de transição entre possíveis estágios da doença de interesse.

3.2 O modelo *SIR*

Dentre os modelos compartimentais existentes, o *SIR* é um dos mais úteis e versáteis, que pode ser elaborado de várias maneiras. Desenvolvido por Kermack e McKendrick (1927), este modelo é atribuído ao estudo de doenças infecciosas em que o indivíduo adquire imunidade após ser curado da infecção.

No contexto deste trabalho, considere uma população animal susceptível à determinada doença infecciosa. Será assumida uma população de tamanho fixo (ou seja, não há nascimentos, mortes ou efeitos migratórios), fechada ao longo do tempo e completamente homogênea (sem estrutura etária, espacial ou social). Considere também que esta população contém um animal infectado, experimentalmente ou não. Este animal, que é denominado caso índice (ou introdutório), fará contatos com determinada quantidade de membros desta população, transmitindo a infecção durante todo seu período infeccioso. Esta descrição representa o início de uma epidemia, cuja evolução pode depender das características epidemiológicas dos hospedeiros, como susceptibilidade e infectividade, já descritas no Capítulo 2.

Assim, na medida em que a epidemia se espalha, o número de animais susceptíveis na população entra em declínio. Em contrapartida, o número de infectados cresce e cada animal infectado passará pelo curso clínico da doença até determinado tempo t . No modelo *SIR* assume-se que o período de incubação do agente infeccioso é instantâneo e a duração da infecciosidade corresponde à duração da doença no organismo hospedeiro. No final, os animais doentes serão removidos do grupo dos infectados e o número de recuperados crescerá, ver Figura 1. Nesse sentido, é importante ressaltar que uma vez recuperados, os animais adquirem imunidade, o que não permite que se tornem susceptíveis novamente.

Desta forma, o modelo *SIR* divide a população em três compartimentos: S (para os

susceptíveis), I (para os infectados) e R (para os recuperados). Neste caso, as letras representam não só as características epidemiológicas, mas também os respectivos números de animais em cada compartimento, que mudam ao longo do tempo. Então, $S(t)$, $I(t)$, e $R(t)$ são funções dependentes do tempo, cuja somatória representa o tamanho da população de animais hospedeiros, que permanece constante para todo tempo $t \geq 0$.

Muitos modelos compartimentais são definidos através de equações diferenciais determinísticas. O modelo básico determinístico SIR , por exemplo, é definido pelas seguintes equações diferenciais ordinárias

$$\begin{cases} \frac{dS}{dt} = -\beta SI \\ \frac{dI}{dt} = \beta SI - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases}$$

tal que $S+I+R=N$ com condição inicial no tempo $t = t_0$; $S(t_0) = S_0 \geq 0$, $I(t_0) = I_0 \geq 0$, $R(t_0) = R_0 \geq 0$.

A Figura 1, que nos mostra como a proporção de animais muda em cada compartimento com o decorrer do tempo, foi obtida considerando um modelo SIR determinístico e sem demografia (isto é, não considera mudanças em termos de crescimento ou decréscimo populacional durante o tempo de estudo) com as seguintes condições iniciais: $S(0) = 0.999$, $I(0) = 0.001$ e $R(0) = 0$.

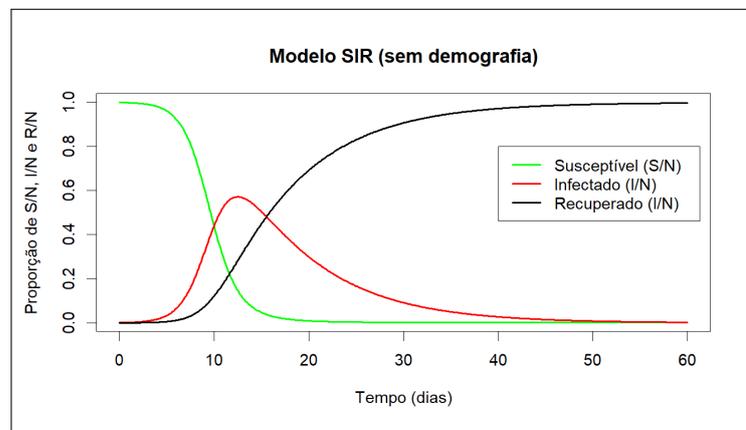


Figura 1 – Soluções numéricas para um modelo SIR determinístico com $\beta = 0.83$ e $\gamma = 0.12$.

Fonte: Elaborada pelo autor.

No entanto, a formulação determinística do modelo SIR não se aproxima da realidade, pois geralmente considera que dadas as mesmas condições iniciais, as trajetórias de S , I e R sempre serão as mesmas ao longo do tempo, o que não condiz com a real dinâmica de patógenos (KEELING; ROHANI, 2011). Por este e outros motivos, sua formulação estocástica, que incorpora a variação aleatória em parâmetros ou processos, tem sido bastante atraente nos últimos anos. De acordo com Berrai (2016), os modelos estocásticos são muito mais consistentes, em princípio, com a realidade biológica porque não se limitam em parâmetros médios e valores numéricos, que são apenas estimativas, mas em probabilidades de ocorrência de eventos como,

por exemplo, a probabilidade de um grande surto, a probabilidade de mudanças de *status* (saudável/doente) de um indivíduo, conhecendo sua situação nos passos anteriores, etc. Esses modelos podem integrar todas as fontes desejadas de heterogeneidade (espacial, temporal, genética, etc) e permitem a incorporação da natureza aleatória dos processos epidêmicos.

Por tudo isso, os modelos estocásticos são capazes de estimar o tempo em que uma doença infecciosa provavelmente persistirá (com ou sem intervenção epidemiológica) em uma população. Mediante as vantagens citadas acima, elaboramos a formulação estocástica de modelos, iniciando pelo modelo *SIR*, na seguinte seção.

3.3 Formulação estocástica do modelo *SIR*

Nos modelos compartimentais como o *SIR*, é necessário especificar como os animais transitam entre os compartimentos. Uma vez que a principal característica das doenças infecciosas em comparação às demais doenças é o fato de poderem ser transmitidas através de contatos, um componente importante para a formulação de modelos compartimentais é o entendimento da estrutura de contatos entre indivíduos da população alvo. Em modelos compartimentais para animais, uma suposição usual é a chamada *mistura homogênea*, que assume que o contato entre dois animais quaisquer ocorre aleatoriamente com a mesma probabilidade, independentemente de qualquer estrutura social na população (KEELING; ROHANI, 2011; BISHOP *et al.*, 2010).

Supondo então que cada animal contacta, por unidade de tempo, em média com c outros animais, de forma aleatória, isto representa a taxa de contato total (c), que é definida como a taxa de contato *per-capita* e por unidade de tempo e mede a frequência de encontro entre eles. No entanto, de acordo com Hal e Keynes (2008), medir essas taxas com precisão é extremamente difícil, seja qual for a via de transmissão, principalmente porque geralmente elas variam entre indivíduos e raramente podem ser documentadas com precisão. Além disso, nesta medida estão inclusos, também, aqueles contatos que não resultam na transmissão da doença, seja pela diversidade genética do animal hospedeiro ou do patógeno. Desta maneira, é fundamental medir também a eficácia com que ocorre a transferência de patógenos de um animal infeccioso para um susceptível. O risco de transmissão p , também chamado de taxa de ataque secundário, é a probabilidade de que, dada a exposição a um animal infeccioso, ocorra uma transferência de agentes patogênicos suficiente para causar a doença contagiosa em um animal susceptível (conhecido ou presumido). Em outras palavras, é a probabilidade de ocorrer contato efetivo, que é o contato que resulta na transmissão da doença. Desta forma, se x representa o número de casos secundários dentro de um grupo, e se o número de susceptíveis for indicado por s , então uma estimativa do risco de transmissão em condições de contato próximo é dada por

$$\hat{p} = \frac{x}{s}. \quad (3.1)$$

Ao combinar a taxa de contato e o risco de transmissão surge um conceito crucial, que é o da taxa de contato efetivo (β), um parâmetro de transmissão que pode ser entendido da

seguinte maneira: em uma população na qual todos os indivíduos interagem com todos os demais (isto é, sem estruturas de contato social relevantes), se um indivíduo susceptível, escolhido aleatoriamente, experimenta c contatos por unidade de tempo com outros indivíduos e se p é a probabilidade de transmissão da doença quando esse indivíduo entra em contato com um indivíduo infeccioso, então a taxa de contato efetivo é definida pela relação

$$\beta = cp. \quad (3.2)$$

Ou seja, a taxa de contato efetivo é diretamente proporcional ao risco de transmissão p .

Após o entendimento destas quantidades, precisamos analisar como as transições acontecem. No modelo *SIR* existem duas possíveis transições, a de $S \rightarrow I$ e a de $I \rightarrow R$. Sendo assim, faremos de imediato uma abordagem ao processo de Poisson, que estabelece alguns pontos-chave da teoria de processos estocásticos que serão necessários para modelar a primeira transição, ou seja, a de $S \rightarrow I$.

3.3.1 Processo de Poisson

De acordo com [Insua, Ruggeri e Wiper \(2012\)](#), um processo de contagem $Y(t)$, $t \geq 0$, é um processo estocástico que conta o número de eventos ocorridos até o momento t . Denotamos por $Y(s, t]$, com $s < t$, o número de eventos ocorridos no intervalo de tempo $(s, t]$.

Definição 3.1. O processo de contagem $Y(t)$, $t \geq 0$, é um processo de Poisson com a função de intensidade $\lambda(t)$ se as seguintes propriedades forem válidas:

- (i) $Y(0) = 0$ e $Y(s) \leq Y(t)$, $\forall s \leq t$;
- (ii) O processo possui incrementos independentes e estacionários, ou seja, para intervalos sem sobreposição (disjuntos) $(s, s + \Delta t]$ e $(t, t + \Delta t]$ temos que $Y(s, s + \Delta t]$ e $Y(t, t + \Delta t]$ são independentes, e identicamente distribuídos, e
- (iii) O número de eventos em qualquer intervalo de comprimento t segue uma distribuição de Poisson com média λt . Isto é, $\forall s, t \geq 0$,

$$\mathbb{P}[(Y(s+t) - Y(s)) = y] = \frac{(\lambda t)^y}{y!} e^{-\lambda t}, \quad y = 0, 1, \dots \quad (3.3)$$

o que explica λ ser chamada de taxa do processo de Poisson ([ROSS et al., 1996](#)).

No entanto, existem situações em que a ocorrência de eventos pode ser mais (ou menos) provável em determinados intervalos de tempo do que em outros, como ocorre por exemplo, durante uma epidemia em determinada população. Neste caso, um processo de Poisson não homogêneo (*PPNH*) pode ser utilizado para modelar o número de contatos efetivos nesta população.

Para formalizar a definição do *PPNH*, inicialmente considere $\lambda : [0, \infty) \rightarrow [0, \infty)$ uma função integrável em intervalos finitos e defina

$$\Lambda(t) = \int_0^t \lambda(y) dy \quad \forall \quad t \geq 0. \quad (3.4)$$

Definição 3.2 (*PPNH*). Um processo de contagem $Y(t)$, $t \geq 0$ é dito ser um processo de Poisson não homogêneo com função de taxa $\lambda(\cdot)$ se

- (i) $Y(0) = 0$;
- (ii) $Y(t)$ possui incrementos independentes, mas não estacionários;
- (iii) $Y(s, t] \sim \text{Poisson}(\Lambda(t))$, em que $\Lambda(t)$ é definido de acordo com (3.4).

Definição 3.3 (Função de intensidade). A função de intensidade λ de um processo de Poisson $Y(t)$ indica quão provável é que um evento ocorra ao redor do tempo t . Sua definição matemática é dada por

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(Y(t, t + \Delta t] \geq 1)}{\Delta t}.$$

Este conceito também se aplica a processos pontuais mais gerais. Observe que quando a função de intensidade é constante $\lambda(t) = \lambda$, para todo t , o processo de Poisson é homogêneo (*PPH*) caso contrário, é não homogêneo (*PPNH*). O item (iii) da Definição 3.2 é uma consequência desta definição, em que para $n \in \mathbb{Z}^+$

$$\mathbb{P}(Y(s, t] = n) = \frac{(\Lambda(t))^n}{n!} \exp\{-\Lambda(t)\}. \quad (3.5)$$

Em particular, para um *PPH* com taxa λ , considera-se que $Y(s, t] \sim \text{Poisson}(\lambda(t-s))$ e os incrementos são estacionários, já que sua distribuição não depende do ponto inicial do intervalo, mas apenas de seu comprimento.

Definição 3.4 (Função de valor médio). Como $Y(s, t] \sim \text{Poisson}(\Lambda(t))$, a função de valor médio do *PPNH* é expressa por

$$\mathbb{E}(Y(t)) = \Lambda(t), \text{ para } t \geq 0.$$

Definição 3.5 (Tempo do evento). Os tempos dos eventos $\{T_n, n \in \mathbb{Z}^+\}$ de um processo de Poisson $Y(t)$ são definidos por

$$T_n = \begin{cases} \min\{t : Y(t) \geq n\}, & n > 0 \\ 0, & n = 0. \end{cases}$$

Definição 3.6 (Tempo entre eventos). Os tempos entre eventos de um processo de Poisson, $Y(t)$, são definidos através de

$$Z_n = \begin{cases} T_n - T_{n-1}, & n > 0 \\ 0, & n = 0. \end{cases}$$

Os tempos dos eventos e entre eventos têm a seguinte relação: $T_n = \sum_{i=1}^n Z_i$.

Teorema 3.1. *Para um PPH com taxa λ , os tempos entre eventos, incluindo o primeiro evento, têm uma distribuição Exponencial(λ), enquanto o n -ésimo tempo de evento, T_n , tem uma distribuição Gama(n, λ), para cada $n \geq 1$.*

Demonstração. Veja em [Insua, Ruggeri e Wiper \(2012, Cap. 5\)](#).

Teorema 3.2 (Superposição). *Considere k processos de Poisson independentes $Y_i(t)$, com função de intensidade $\lambda_i(t)$ e função de valor médio $\Lambda_i(t)$, $i = 1, 2, \dots, k$ (ver Definição 3.4). Então o processo $Y(t) = \sum_{i=1}^k Y_i(t)$, para $t \geq 0$, é o processo de Poisson com intensidade $\lambda(t) = \sum_{i=1}^k \lambda_i(t)$ e função de valor médio $\Lambda(t) = \sum_{i=1}^k \Lambda_i(t)$.*

Demonstração. Veja em [Kulkarni \(2016, pag. 177\)](#).

A partir destes resultados modelamos a transição SI na Subseção 3.3.2 a seguir.

3.3.2 A transição de S para I

Inicialmente, considere dois animais que podem interagir diversas vezes, sendo que um deles é susceptível e o outro é infeccioso. Neste caso, o número de contatos efetivos, ou seja, o número de contatos que podem resultar na transmissão da infecção, pode ser modelado por um processo de Poisson com intensidade correspondente à taxa de contato efetivo β .

Agora, considere a situação em que temos um animal susceptível contactando com I animais infectados, conforme ilustrado na Figura 2. Neste cenário, temos I interações possíveis, o que corresponde a uma combinação de I processos de Poisson independentes com intensidade β . Então, pelo Teorema da Superposição (3.2), para um tempo t , o número total de contatos que resultam em infecção segue um processo de Poisson com função intensidade dada por $\lambda(t) = \beta I(t)$. Desta forma, esse animal pode ter vários contatos com os infecciosos, que em determinado momento algum deles poderá resultar na ocorrência da doença infecciosa.

Analogamente, suponha que ao invés de termos somente um susceptível e I infectados num determinado tempo t , agora temos S susceptíveis e I infectados. Então, pelo princípio multiplicativo, o número de interações possíveis entre susceptíveis e infectados no tempo t é $S(t)I(t)$. Como cada contato efetivo é modelado por processo de Poisson com taxa β , então o número total de contatos efetivos que resultam dessas possíveis interações pode ser modelado também por um processo de Poisson com função intensidade dada por $\lambda(t) = \beta S(t)I(t)$. Esta função de intensidade representa o número de animais susceptíveis convertidos em infectados durante o tempo a que β se refere. É a força (ou risco) da infecção, isto é, o risco total em que os susceptíveis se encontram devido a todas as exposições a indivíduos infectados. Esta é a regra de transição que governa o processo de infecção - a transição entre os *status* susceptível e infeccioso.

E pelo Teorema (3.1) os tempos entre infecções, incluindo a primeira infecção, são modelados por uma distribuição exponencial de parâmetro $\lambda(t)$.

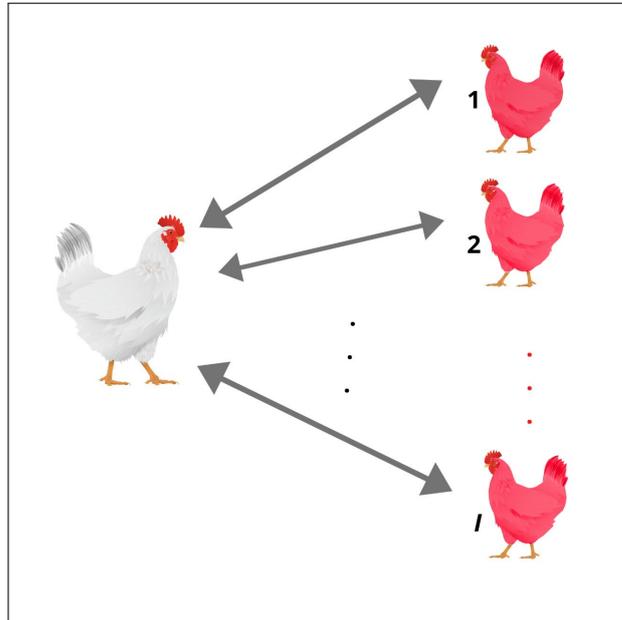


Figura 2 – Interações entre um animal suscetível com I infectados.

Fonte: Elaborada pelo autor.

3.3.3 A transição de I para R

Com o passar do tempo, os animais que adquiriram a infecção deixam o compartimento dos infectados, por recuperação ou morte. A transição $I \rightarrow R$ é assumida independente da transição $S \rightarrow I$. Nesse caso, todos os períodos infecciosos são independentes e identicamente distribuídos. Isto implica que, num período Δt , qualquer um dos animais infectados se recuperará com a mesma probabilidade, independentemente de como eles adquiriram a doença. Sendo assim, a taxa de recuperação ou de mortalidade, é denotada por γI , em que γ é constante por unidade de tempo e inversa ao tempo médio de duração do período infeccioso.

Desta forma, considere $Z \geq 0$ uma variável aleatória denotando o tempo que o animal permanece infectado até tornar-se recuperado. Além disso, assumamos que Z segue uma distribuição exponencial com parâmetro ou taxa $\gamma > 0$, isto é, sua função densidade de probabilidade é dada por

$$f(z) = \gamma e^{-z\gamma}, \quad z \geq 0.$$

O parâmetro γ é chamado de parâmetro de escala e mede o tempo médio de infecção, isto é, $\mathbb{E}(Z) = \frac{1}{\gamma}$. Sendo assim, a probabilidade de cada animal permanecer no compartimento I após z unidades de tempo desde que entrou nessa classe (com tempo médio de espera $\frac{1}{\gamma}$) é dada por

$$\mathbb{P}(Z \geq z) = e^{-\gamma z},$$

e a probabilidade de cada animal sair do compartimento I para entrar no compartimento R é dada por

$$\mathbb{P}(Z \leq z) = 1 - e^{-\gamma z}.$$

Para os casos em que esta transição não é necessária, o modelo se reduz a SI .

3.3.4 *Evolução e espalhamento da epidemia*

Sob a premissa de população homogênea, é possível obter uma relação entre a taxa de contato efetivo e o número básico de reprodução (R_0), o qual representa o número esperado de casos secundários produzidos por um único animal infeccioso durante todo o seu período de infecciosidade em uma população completamente susceptível (DIEKMANN; HEESTERBEEK; METZ, 1990). Suponhamos, então, que os contatos efetivos ocorram aleatoriamente com uma taxa de contato efetivo β na população e que a duração média do período infeccioso seja de $\frac{1}{\gamma}$. Então, o número total de contatos efetivos feitos por um indivíduo infeccioso é de $\frac{\beta}{\gamma}$. Mas se toda a população é susceptível, então cada um desses contatos efetivos resulta em uma infecção (por definição do que é um contato efetivo). Portanto, o número de infecções secundárias também é igual a $\frac{\beta}{\gamma}$ e, conseqüentemente, obtemos a relação

$$R_0 = \frac{\beta}{\gamma}.$$

Esta quantidade é um número real não negativo tal que:

- Se $R_0 \in [0, 1]$, o animal infectado contamina em média apenas um outro animal, ou menos que isso. Conseqüentemente, a infecção pode não se estabelecer e, caso se estabeleça, irá desaparecer rapidamente da população por envolver um número muito pequeno de animais;
- Se $R_0 \in (1, +\infty)$ a infecção pode, pelo menos inicialmente, se espalhar na população e tornar-se epidêmica, no sentido de que há uma probabilidade positiva de um grande surto. Neste caso, podemos notar que quanto maior o valor de R_0 maior a proporção de infecciosos na população. Portanto, R_0 também é, neste caso, um indicador de infecciosidade da doença.

Desta forma R_0 ocupa um papel central no estudo de doenças infecciosas pois ele indica se uma doença se espalhará em uma população.

Todos os conceitos de modelos compartimentais apresentados nesse capítulo são amplamente utilizados em aplicações do cotidiano. Vale a pena salientar que a pandemia do Coronavírus foi um grande exemplo disso, tendo esses conceitos inclusive sendo disseminados, por exemplo,

no noticiário e jornais. No entanto, estas aplicações geralmente são focadas em grandes populações e ignoram a heterogeneidade entre indivíduos, que é justamente o que modelos de Genética Quantitativa e programas de melhoramento genético visam explorar.

MODELOS ESTATÍSTICOS PARA ANÁLISES GENÉTICAS DE INFECTIVIDADE ANIMAL

Como mencionado no Capítulo 2, um dos desafios na modelagem da variação genética na infectividade é o fato dela não ser necessariamente mensurada no animal que a expressa, e sim observada nos seus companheiros de grupo. Diante disso, os chamados modelos de efeitos genéticos indiretos (*indirect genetic effects* - *IGEs*) apresentam grande relevância, pois permitem estimar herdabilidades e valores genéticos de características expressas a partir de interações sociais (GRIFFING, 1967; BIJMA; MUIR; ARENDONK, 2007). Em outras palavras, eles permitem avaliar a relação entre o genótipo de um animal e o fenótipo de seus companheiros de grupo. De acordo com Baud *et al.* (2017), também pode ser importante incorporar *IGEs* em modelos de Genética Quantitativa mesmo quando o principal interesse está em características expressas de forma direta.

Existem diversos exemplos documentados de estudos usando *IGEs* e mencionamos alguns. Ellen *et al.* (2008) e Peeters *et al.* (2012) estudaram *IGEs* para redução da mortalidade por canibalismo em galinhas poedeiras e domésticas (*Gallus gallus*), respectivamente. Bergsma *et al.* (2013) e Camerlink *et al.* (2018) avaliaram os efeitos para o crescimento e comportamento de porcos em decorrência do desmame.

As ideias básicas e exemplos de aplicações de modelos *IGE* para dados de epidemias geralmente escrevem o fenótipo de um animal como função do seu próprio efeito genético e ambiental de susceptibilidade (que são considerados efeitos diretos) e dos efeitos genéticos e ambientais de infectividade dos outros animais que vivem nesse mesmo ambiente (efeitos sociais e indiretos).

Os dados de doenças infecciosas em estudos de campo geralmente são binários, indicando se cada animal foi infectado após a exposição a uma doença infecciosa. Lipschutz-Powell *et al.* (2012a) demonstraram que, ao utilizar modelos *IGE* para tais dados de doenças infecciosas, a

variação genética na infectividade do hospedeiro pode ser capturada, embora não completamente. Esse resultado foi obtido a partir de um estudo de simulação com dados binários de doença, utilizando um modelo de efeitos mistos para resposta com distribuição normal, com efeitos aleatórios para representar efeitos genéticos poligênicos e efeitos ambientais de infectividade e susceptibilidade.

Um resultado similar foi obtido quando se assume efeitos monogênicos na infectividade. [Anche, Bijma e Jong \(2015\)](#) desenvolveram um modelo *IGE* baseado em modelos lineares generalizados (*generalized linear model - GLM*) com resposta binomial. O desenvolvimento do modelo foi baseado na relação entre *GLM* binomiais com função de ligação log-log complementar e modelos *SIR* ([BECKER, 2017](#)). No modelo, o *status* final do animal durante uma epidemia foi definido como variável resposta. A partir de estudos de simulação, foi verificado que essa classe de modelos subestima os efeitos genéticos de infectividade. Além disso, [Biemans, Jong e Bijma \(2017\)](#) verificaram que essa subestimação persiste quando um modelo similar é utilizado para dados longitudinais binários.

A desvantagem no uso de modelos *IGEs* baseados em modelos mistos se dá pela desconsideração da dinâmica no processo de transmissão da doença. É assumido que todos os membros do grupo expressam o efeito indireto (infectividade) em todos os tempos e que todos os animais são afetados pelo efeito indireto. No entanto, isto contrasta com a realidade, uma vez que apenas os membros infectados podem expressar infectividade e esta não é constante ao longo do tempo. Além disso, o número de animais susceptíveis também varia com o passar do tempo e o *status* binário da doença para cada animal submetido a epidemia é definido por uma função que depende da interação de sua susceptibilidade e da pressão da infecção variável no tempo exercida pelos animais infectados ([KEELING; ROHANI, 2011](#)). O problema da estimação de efeitos genéticos indiretos quando as interações variam no tempo foi discutido em [Bijma \(2014\)](#).

Tudo isso nos mostra que os modelos *IGE* convencionais falham em capturar a natureza dinâmica e não-linear dos processos da doença. A combinação de modelos genético-quantitativos de interações sociais *IGE* com modelos estocásticos compartimentais permitiu o desenvolvimento do modelo dinâmico não-linear de efeitos genéticos indiretos (*dynamic non-linear indirect genetic effects models - dnIGE*), cuja descrição é dada na seção a seguir.

4.1 O modelo dinâmico não linear de efeitos genéticos indiretos (*dnIGE*)

O modelo dinâmico não linear de efeitos genéticos indiretos (*dynamic non-linear indirect genetic effects model - dnIGE*), desenvolvido por [Anacleto et al. \(2015\)](#), é um dos primeiros modelos que incorpora explicitamente o processo de transmissão de doenças para estimar efeitos genéticos e herdabilidade de características epidemiológicas em animais de criação.

Considere uma população de animais sujeita a um surto de uma determinada doença infecciosa que segue um processo de transmissão representado por um modelo compartimental susceptível-infectado (*SI*), visto no Capítulo 3. Neste caso, uma vez infectado, o animal permanece infectado e capaz de transmitir a infecção durante todo seu tempo de vida. Considere também as seguintes suposições para o modelo *dnIGE*:

- (i) Os N participantes do estudo estão alocados em P grupos fechados, cada um de tamanho $N_k, k = 1, \dots, P$, em que $N = \sum_{k=1}^P N_k$;
- (ii) As epidemias em todos os grupos fechados são observadas dentro de um intervalo fixo $[0, T]$, em que o tempo zero representa o início da observação e T é o tempo final da observação;
- (iii) Além disso, estas epidemias podem resultar em surtos da doença por animais que foram infectados antes do intervalo fixo de observação ou pela introdução de animais infectados artificialmente (infecção experimental) no tempo $t = 0$;
- (iv) O tempo de infecção (que só pode ser observado no intervalo $[0, T]$) é denotado por τ_j e h_j é o indicador de não ser caso índice para o animal j , de forma que $\forall j = 1, \dots, N$ temos a seguinte definição:

$$h_j = \begin{cases} 0, & \text{se } \tau_j = 0 \quad (j \text{ é um caso índice}) \\ 1, & \text{se } \tau_j > 0 \quad (j \text{ não é caso índice}); \end{cases}$$

- (v) O *status* da doença (saudável/doente) pode ser periodicamente registrado em tempos $[t_0 = 0, t_1, \dots, t_{M-1}, t_M = T]$, em que M é o número de tempos da amostragem.

Quando os tempos de infecção τ_1, \dots, τ_N não podem ser registrados exatamente, os parâmetros do modelo podem ser estimados usando o *status* da doença registrado em cada um dos tempos de amostragem $[t_0 = 0, t_1, \dots, t_{M-1}, t_M = T]$.

Como exemplo, a Figura 3 representa um dos grupos onde uma epidemia pode ser observada. Podemos observar que o *animal 1* é caso índice, pois ele já inicia no estudo com o *status* de “infectado”, isto é, seu tempo de infecção é representado por $\tau_1 = 0$. O *animal 2* não é observado como infectado durante o tempo de observação, uma vez que $\tau_2 > T$ significa que o estudo termina e ele não adquire a doença no intervalo de observação (censura). Além disso, o *animal 3* está infectado (doente) no tempo de infecção τ_3 , registrado entre os tempos de amostragem t_1 e t_2 , enquanto que o tempo de infecção do *animal* N_k ocorre entre os tempos t_{M-2} e t_{M-1} .

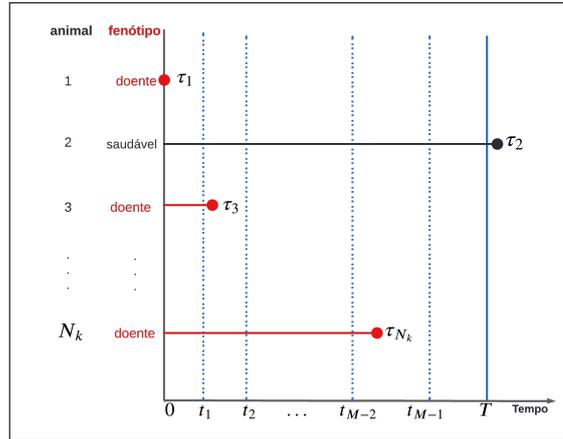


Figura 3 – Estrutura dos dados de infecção considerados para o desenvolvimento do modelo *dnIGE*.

Fonte: [Anacleto et al. \(2015\)](#).

Para a definição do modelo, inicialmente será assumido que todos os tempos de infecção são exatamente conhecidos, dados por $\tau = [\tau_1, \dots, \tau_I]^\top$, em que I é o número de infectados na população durante o intervalo $[0, T]$.

4.1.1 O Modelo

Considerando as características genéticas que influenciam na transmissão de doenças infecciosas, já descritas no Capítulo 2, os pressupostos, a formulação estocástica do modelo *SI* descritos no Capítulo 3, e as suposições específicas deste capítulo, fixamos um animal j e definimos p_j como sendo o número do grupo deste animal, sendo $\mathbf{p} = [p_1, \dots, p_N]^\top$. Assim, a taxa de infecção variável no tempo do animal j no tempo t , com $j = 1, \dots, N$, é dada por

$$\lambda_j(t) = g_j \beta \sum_{k:p_k=p_j} f_k \ell_k(t), \quad (4.1)$$

em que g_j é a susceptibilidade do j -ésimo animal, f_k é a infectividade do k -ésimo animal, $k : p_k = p_j$ no somatório representa o conjunto de todos os companheiros de grupo do animal j , e o indicador $\ell_k(t)$ é definido de forma que seleciona apenas os animais que foram infectados antes de j

$$\ell_k(t) = \begin{cases} 1, & \text{se } \tau_k < t \\ 0, & \text{caso contrário.} \end{cases}$$

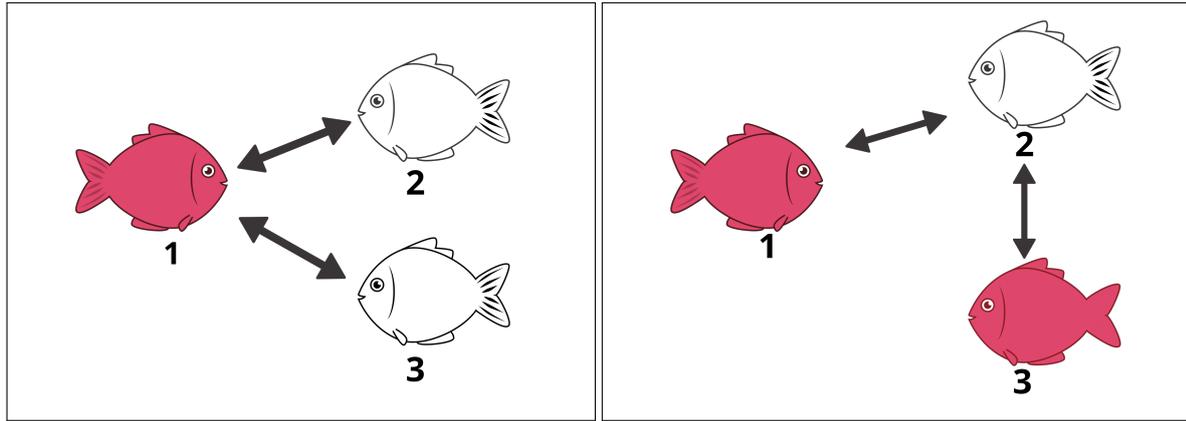
Desta forma, a taxa de infecção variável no tempo de um animal susceptível é uma função que depende não apenas da taxa de contato efetivo β da população, mas também de sua própria susceptibilidade e da infectividade de seus parceiros de grupo previamente infectados. Esta taxa de infecção é equivalente à função intensidade de um processo de Poisson não homogêneo para o número de contatos efetivos de cada animal na população, ver Definição 3.3.

Em populações homogêneas, quando $g_j = f_k = 1$, a taxa de infecção dada na Equação (4.1) representa a chamada força de infecção dependente da densidade (KEELING; ROHANI, 2011), que é a taxa na qual os animais são infectados durante uma epidemia. O modelo *dnIGE* assume que a taxa de infecção variável no tempo, definida em (4.1), é condicional em termos de fragilidade aleatória (AALEN; BORGAN; GJESSING, 2008), a g_j e f_k , que capturam heterogeneidade não observada na susceptibilidade e infectividade do hospedeiro, respectivamente, (ambos relativos). Esses termos de fragilidade representam desvios (ou efeitos aleatórios) em relação ao parâmetro populacional β e são, portanto, variação individual no processo de infecção. Assim, $\lambda_j(t)$ pode ser visto como a força individual da infecção agindo no animal j no tempo t e captura a heterogeneidade individual, os efeitos da média populacional e a dinâmica de transmissão não-linear do processo de infecção.

Como um exemplo da taxa de infecção do modelo *dnIGE*, considere o caso hipotético de um grupo único formado por somente dois animais, no qual o *animal 1* é um caso índice e o *animal 2* pode ser infectado pelo *animal 1*. Nesse caso, como o *animal 1* é um caso índice, seu tempo de infecção é zero ($\tau_1 = 0$) e consequentemente ele não expressa susceptibilidade (g_1). Analogamente, o *animal 2* por ser o último a ser infectado no grupo, não expressa infectividade (f_2). Usando a definição do modelo, temos que a taxa ou força de infecção do *animal 2* depende de sua própria susceptibilidade, da taxa de contato efetivo e da infectividade do *animal 1*. Matematicamente, $\lambda_2(t) = g_2\beta f_1$.

Agora, suponha um animal adicional no grupo, denominado *animal 3*. Considere que o interesse é definir a taxa de infecção agindo no *animal 2*, ainda definido como susceptível no exemplo. Uma situação que pode acontecer nesse caso é a composição do grupo constar apenas com um animal infectado, enquanto os outros dois são susceptíveis. Na segunda situação ocorre o inverso, ou seja, o grupo é composto por dois animais infectados e um susceptível, conforme podemos observar na Figura 4.

Note que estamos interessados somente nos contatos que resultam na transmissão da doença infecciosa. Desta forma, na Figura 4(a), as interações entre os animais susceptíveis 2 e 3 são desprezadas e a força de infecção do animal fixado coincide com a do exemplo anterior pois, uma vez que *animal 3* não expressa infectividade, o modelo atribui $\ell_3(t) = 0$ e desconta-o do somatório de infectados do grupo. Analogamente, na Figura 4(b) as interações entre os infectados 1 e 3 são descartadas e a força de infecção do *animal 2* é dada por $\lambda_2(t) = g_2\beta(f_1 + f_3)$.



(a) Apenas um animal entre os três está infectado.

(b) Dois estão infectados, um não.

Figura 4 – Grupos independentes de tamanho 3 representando as diferentes interações entre os animais.

Fonte: Elaborada pelo autor.

4.1.2 Componentes de variância para a susceptibilidade (g_j) e infectividade (f_j)

No modelo *dnIGE*, a força individual de infecção é decomposta a partir de uma estrutura de componentes de variância para a susceptibilidade e infectividade, descrita a seguir.

A susceptibilidade de um animal j é modelada como

$$\log(g_j) = a_{g,j} + e_{g,j}, \quad (4.2)$$

em que $a_{g,j}$ é o efeito genético aditivo (ou valor genético) para a susceptibilidade do animal j e $e_{g,j}$ é o seu efeito ambiental (ou valor ambiental ou, ainda, erro aleatório) para a susceptibilidade.

Observe que os efeitos médios de susceptibilidade são capturados pelo nível populacional da taxa de contato efetivo β na Equação (4.1). Como os casos índice ($\tau_j = 0$) não expressam susceptibilidade dentro do período de observação, as Equações (4.1) e (4.2) são consideradas neste contexto somente para cada j tal que $h_j = 1$.

Portanto, assumindo n_0 casos índice, o vetor \mathbf{a}_g de dimensão $(N - n_0)$ representa os efeitos genéticos aditivos de susceptibilidade, com $\mathbf{a}_g | \sigma_{A,g}^2 \sim N(\mathbf{0}, \sigma_{A,g}^2 \mathbf{A}_g)$, em que \mathbf{A}_g é a matriz de relação genética aditiva, baseada em *pedigree*, entre os animais susceptíveis da população, excluindo linhas e colunas relacionadas aos animais que são casos índice. Além disso, \mathbf{e}_g é o vetor de dimensão $(N - n_0)$ dos efeitos ambientais de susceptibilidade, com $\mathbf{e}_g | \sigma_{E,g}^2 \sim N(\mathbf{0}, \sigma_{E,g}^2 \mathbf{I}_g)$, em que \mathbf{I}_g é a matriz identidade de dimensão $(N - n_0)$.

Da mesma forma, um modelo de componentes de variância para a infectividade de um animal infectado j é definido como

$$\log(f_j) = a_{f,j} + e_{f,j}, \quad (4.3)$$

em que $a_{f,j}$ é o efeito genético aditivo da infectividade do animal j e $e_{f,j}$ é o efeito ambiental da infectividade.

A Equação (4.3) é definida para todo j que pode representar a infectividade, considerando casos índices e outros animais que são infectados durante o período de observação. Definindo I_f como o número de animais que podem expressar infectividade na população, \mathbf{a}_f representa o vetor I_f -dimensional de efeitos genéticos aditivos de infectividade, tal que $\mathbf{a}_f | \sigma_{A,f}^2 \sim N(\mathbf{0}, \sigma_{A,f}^2 \mathbf{A}_f)$ e \mathbf{A}_f é a matriz de relação genética aditiva, baseada em *pedigree*, entre os animais infectados da população excluindo linhas e colunas relacionadas a animais que não expressam infectividade. Adicionalmente, \mathbf{e}_f representa o vetor I_f -dimensional de efeitos ambientais de infectividade com $\mathbf{e}_f | \sigma_{E,f}^2 \sim N(\mathbf{0}, \sigma_{E,f}^2 \mathbf{I}_f)$ e \mathbf{I}_f sendo a matriz identidade de dimensão I_f .

Através da combinação de um processo não homogêneo de Poisson com a estrutura de componentes de variância apresentada, o modelo *dnIGE* estende o chamado modelo *IGE*, que permite estimar herdabilidades e valores genéticos de características expressas a partir de interações sociais que afetam fenótipos com distribuição normal (GRIFFING, 1967; BIJMA; MUIR; ARENDONK, 2007).

A Figura 5 mostra uma representação hierárquica da estrutura de componentes de variância para o exemplo hipotético de dois animais definido anteriormente. Observamos que o tempo de infecção do *animal 2* (τ_2) depende da taxa de transmissão β , da susceptibilidade deste animal (g_2) e da infectividade do *animal 1* (f_1). Por sua vez, estes fenótipos dependem dos efeitos genéticos ($a_{g,2}$) e ($a_{f,1}$) e ambientais ($e_{g,2}$) e ($e_{f,1}$), que dependem respectivamente das variâncias genéticas e ambientais da susceptibilidade e infectividade, representadas nas matrizes de covariância Σ_a e Σ_e . O vetor de parâmetros desconhecidos do *dnIGE* é dado por

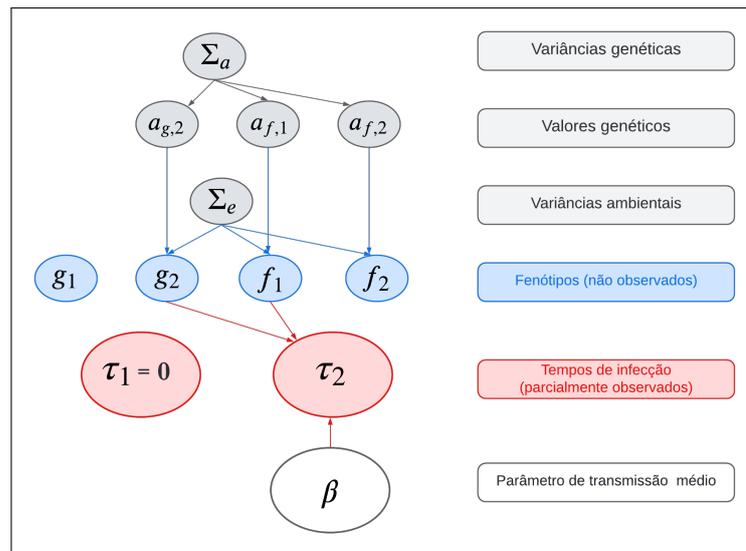


Figura 5 – Representação hierárquica do modelo para grupo de tamanho 2.

Fonte: Elaborada pelo autor.

$\boldsymbol{\theta} = \left[\mathbf{a}_g^\top, \mathbf{a}_f^\top, \mathbf{e}_g^\top, \mathbf{e}_f^\top, \sigma_{A,g}^2, \sigma_{A,f}^2, \sigma_{E,g}^2, \sigma_{E,f}^2, \beta \right]^\top$ e sua estimação Bayesiana será apresentada com

mais detalhes na Seção 4.2, junto com uma proposta de extensão desse modelo.

Para avaliar a qualidade do modelo, [Anacleto et al. \(2015\)](#) desenvolveram um estudo de simulação no qual dados de epidemias foram gerados para populações divididas em grupos a partir das suposições do *dnIGE*, com animais geneticamente relacionados (meios-irmãos). Na geração da população, os pais não tinham parentesco e um modelo poligênico foi considerado para a arquitetura genética das características.

Neste estudo verificou-se que o modelo *dnIGE* fornece estimativas precisas dos parâmetros genéticos tanto para susceptibilidade como também para infectividade sendo que, com herdabilidades maiores, o modelo capta variações genéticas de forma mais eficiente para ambas as características (acurácias de predição são maiores). Além disso, o efeito genético na infectividade foi melhor estimado quando os grupos eram menores, enquanto que o efeito genético na susceptibilidade seguiu tendência oposta.

Como mostrado em estudos de simulação em [Tsairidou et al. \(2019\)](#) e [Raphaka et al. \(2018\)](#), embora seja mais difícil capturar variações genéticas na infectividade, programas de melhoramento genético baseado em valores de acurácia estimados pelo modelo *dnIGE* podem contribuir para a redução de surtos epidêmicos e controle de doenças infecciosas.

4.2 O modelo *dnIGE* estendido

No modelo *dnIGE* original assume-se que tanto os efeitos genéticos quanto os efeitos ambientais da infectividade são independentes desses mesmos efeitos da susceptibilidade. Neste trabalho propõe-se uma extensão deste modelo, na qual é assumida uma estrutura de covariância na distribuição dos efeitos genéticos e ambientais dessas características.

Considere um modelo de risco paterno para os componentes de variância de susceptibilidade e infectividade ([MRODE, 2014](#)), ou seja, dada a exposição da prole à infecção, este modelo é tal que as equações do chamado modelo animal definidas em (4.2) e (4.3) são substituídas pelas seguintes:

$$\begin{cases} \log(g_j) = a_{g,s(j)} + e_{g,j}, & \text{se } j \text{ não é um caso índice e,} \\ \log(f_j) = a_{f,s(j)} + e_{f,j}, & \text{se } j \text{ pode expressar infectividade,} \end{cases} \quad (4.4)$$

em que $s(j)$ representa o pai do animal j , $j = 1, 2, \dots, N$.

Considere, ainda, que não existe grau de parentesco entre os animais selecionados para a reprodução e somente os genitores masculinos serão avaliados a partir dos registros da observação da prole submetida à epidemia. Nesse caso, os vetores de efeitos genéticos são definidos por $\mathbf{a} = [\mathbf{a}_g^\top, \mathbf{a}_f^\top]^\top$, em que S é o número de pais, \mathbf{a}_g e \mathbf{a}_f representam, respectivamente, os efeitos

genéticos de susceptibilidade e infectividade; e o vetor de efeitos ambientais é definido por $\mathbf{e} = [\mathbf{e}_g^\top, \mathbf{e}_f^\top]^\top$.

Desta forma, a distribuição dos efeitos genéticos e ambientais é dada por

$$\mathbf{a} | \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A \sim N_{2S} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{S,g}^2 & \rho_A \sigma_{S,g} \sigma_{S,f} \\ \rho_A \sigma_{S,g} \sigma_{S,f} & \sigma_{S,f}^2 \end{bmatrix} \otimes \mathbf{I}_S \right), \quad (4.5)$$

$$\mathbf{e} | \sigma_{E,g}^2, \sigma_{E,f}^2, \rho_E \sim N_{2N} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{E,g}^2 & \rho_E \sigma_{E,g} \sigma_{E,f} \\ \rho_E \sigma_{E,g} \sigma_{E,f} & \sigma_{E,f}^2 \end{bmatrix} \otimes \mathbf{I}_N \right), \quad (4.6)$$

em que $\sigma_{S,g}^2 = \frac{1}{4} \sigma_{A,g}^2$ e $\sigma_{S,f}^2 = \frac{1}{4} \sigma_{A,f}^2$ são as variâncias genéticas paternais obtidas a partir da decomposição do valor genético de um animal em valores genéticos de seus pais e um termo Mendeliano; ρ_A e ρ_E são os coeficientes de correlação; e \otimes é o produto de Kronecker (HENDERSON; SEARLE, 1981). Nesses vetores alocamos consecutivamente o efeito genético para susceptibilidade e infectividade de cada animal, assim como seus efeitos ambientais.

4.2.1 Função de verossimilhança

Seja $Y_j(w)$ o número de contatos efetivos do j -ésimo animal até o tempo w e $\lambda_j(t)$ a taxa de infecção variável no tempo desse animal j no tempo t , para $j = 1, \dots, N$, dada pela Equação (4.1) e em função dos efeitos genéticos de seu pai $s(j)$. De acordo com a Equação (3.5),

$$Y_j(w) \sim \text{Poisson} \left(\int_0^w \lambda_j(t) dt \right)$$

e, pela Definição (3.3), a probabilidade do animal j ter pelo menos um contato efetivo no tempo τ_j , é dada por $\Lambda_j(\tau_j)$. Assim, a probabilidade do tempo de infecção do j -ésimo animal ser exatamente τ_j é dada por :

$$\begin{aligned} \mathbb{P}(Y_j(\tau_j) = 0) \Lambda_j(\tau_j) &= \exp \left\{ - \int_0^{\tau_j} \lambda_j(t) dt \right\} \Lambda_j(\tau_j) \\ &= \exp \left\{ - \int_0^{\tau_j} \exp \{ a_{g,s(j)} + e_{g,j} \} \beta \sum_{k:p_k=p_j} \exp \{ a_{f,s(k)} + e_{f,k} \} \ell_k(t) dt \right\} \\ &\quad \exp \{ a_{g,s(j)} + e_{g,j} \} \beta \sum_{k:p_k=p_j} \exp \{ a_{f,s(k)} + e_{f,k} \} \ell_k(\tau_j) \end{aligned} \quad (4.7)$$

Em contrapartida, a probabilidade do tempo de infecção do j -ésimo animal ser superior ao tempo final T do experimento é

$$\begin{aligned} \mathbb{P}(Y_j(T) = 0) &= \exp \left\{ - \int_0^T \lambda_j(t) dt \right\} \\ &= \exp \left\{ - \int_0^T \exp \{ a_{g,s(j)} + e_{g,j} \} \beta \sum_{k:p_k=p_j} \exp \{ a_{f,s(k)} + e_{f,k} \} \ell_k(t) dt \right\} \end{aligned} \quad (4.8)$$

Seja $\boldsymbol{\theta} = (\mathbf{a}^\top, \mathbf{e}^\top, \sigma_{S,g}^2, \sigma_{S,f}^2, \sigma_{E,g}^2, \sigma_{E,f}^2, \rho_A, \rho_E, \boldsymbol{\beta})^\top$ o vetor de parâmetros do modelo descrito e considerando que os animais, que não são casos índice e são infectados antes de T , contribuem com a probabilidade (4.7) para a função de verossimilhança (COLOSIMO; GIOLO, 2006, pag.87) e que os animais não infectados até T contribuem com a probabilidade (4.8), a função de verossimilhança é descrita como

$$\begin{aligned} \mathbb{L}(\boldsymbol{\tau}|\boldsymbol{\theta}) = & \prod_{\substack{j:h_j=1 \\ \tau_j \leq T}} \left\{ \exp\{a_{g,s(j)} + e_{g,j}\} \boldsymbol{\beta} \sum_{k:p_k=p_j} \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \right\} \\ & \prod_{\substack{j:h_j=1 \\ \tau_j \leq T}} \left\{ \exp\left(-\int_0^{\tau_j} \exp\{a_{g,s(j)} + e_{g,j}\} \boldsymbol{\beta} \sum_{k:p_k=p_j} \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(t) dt\right) \right\} \\ & \prod_{\substack{j:h_j=1 \\ \tau_j > T}} \left\{ \exp\left(-\int_0^T \exp\{a_{g,s(j)} + e_{g,j}\} \boldsymbol{\beta} \sum_{k:p_k=p_j} \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(t) dt\right) \right\}, \end{aligned}$$

em que $j : h_j = 1$ representa os animais que não são casos índice, $\tau_j \leq T$ representa o tempo de infecção no intervalo de observação e $\tau_j > T$ representa censura. Além disso, $k : p_k = p_j$ no somatório representa o conjunto de todos os companheiros de grupo do animal j , e o indicador $\ell_k(t)$ é definido de forma que seleciona apenas os animais que foram infectados antes de j

$$\ell_k(t) = \begin{cases} 1, & \text{se } \tau_k < t \\ 0, & \text{caso contrário.} \end{cases}$$

Desta forma, a função de verossimilhança é definida como um produto cujos fatores correspondem, respectivamente, à probabilidade de se observar os tempos de infecção dos animais que não são casos de índice e a probabilidade de não se observar ocorrências de infecções ($\tau_j > T$) nos susceptíveis remanescentes (este último é o caso em que período de observação $[0, T]$ termina e ainda restam animais que não expressam infectividade). Esta verossimilhança é baseada em trabalhos de inferência Bayesiana para modelos estocásticos *SIR* assumindo populações homogêneas (O'NEILL; ROBERTS, 1999).

Consequentemente, a função log-verossimilhança é dada por

$$\begin{aligned} \ell(\boldsymbol{\tau}|\boldsymbol{\theta}) = & \sum_{\substack{j:h_j=1 \\ \tau_j \leq T}} \log \left(\boldsymbol{\beta} \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \right) \\ & - \boldsymbol{\beta} \sum_{\substack{j:h_j=1 \\ \tau_j \leq T}} \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} (\tau_j - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \\ & - \boldsymbol{\beta} \sum_{\substack{j:h_j=1 \\ \tau_j > T}} \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} (T - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j). \end{aligned}$$

4.2.2 Distribuição conjunta a posteriori

A partir do Teorema de Bayes e explorando a estrutura paramétrica hierárquica do modelo *dnIGE*, a distribuição conjunta *a posteriori* de $\boldsymbol{\theta}$ é definida como

$$\begin{aligned} \mathbb{P}(\boldsymbol{\theta}|\boldsymbol{\tau}) \propto & \mathbb{L}(\boldsymbol{\tau}|\boldsymbol{\theta})\mathbb{P}(\mathbf{a}|\sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A)\mathbb{P}(\mathbf{e}|\sigma_{E,g}^2, \sigma_{E,f}^2, \rho_E) \\ & \mathbb{P}(\sigma_{S,g}^2)\mathbb{P}(\sigma_{S,f}^2)\mathbb{P}(\sigma_{E,g}^2)\mathbb{P}(\sigma_{E,f}^2)\mathbb{P}(\rho_A)\mathbb{P}(\rho_E)\mathbb{P}(\boldsymbol{\beta}). \end{aligned}$$

As distribuições *a priori* dos efeitos genéticos aditivos e dos efeitos ambientais estão definidas nas Equações (4.5) e (4.6). Para as variâncias genéticas aditivas e ambientais uma sugestão é usar as prioris não informativas de Bayes-Laplace (PAULINO; TURKMAN; MURTEIRA, 2003). Desta forma, $\mathbb{P}(\sigma_{S,g}^2) = \frac{1}{\sigma_{S,g}^2}$, $\mathbb{P}(\sigma_{S,f}^2) = \frac{1}{\sigma_{S,f}^2}$, $\mathbb{P}(\sigma_{E,g}^2) = \frac{1}{\sigma_{E,g}^2}$, $\mathbb{P}(\sigma_{E,f}^2) = \frac{1}{\sigma_{E,f}^2}$. Para os coeficientes de correlação, podemos definir distribuições uniformes, $\rho_A \sim U(-1, 1)$ e $\rho_E \sim U(-1, 1)$. Para a taxa de contato efetivo assume-se, geralmente, uma distribuição Gama, $\beta \sim \text{Gama}(10^{-3}, 10^{-3})$. No entanto, outras distribuições *a priori* podem ser assumidas para as variâncias e correlações ou diretamente para as matrizes de variância e covariância, como veremos nos estudos de simulação apresentados no próximo capítulo.

A inferência para os parâmetros do modelo *dnIGE* estendido pode ser realizada por meio da metodologia de *Monte Carlo via Cadeias de Markov (MCMC)* (GELMAN *et al.*, 2013), que fornece amostras da distribuição *a posteriori* conjunta, obtida na Expressão (4.9), a partir das distribuições condicionais completas de cada parâmetro e cada variável não observada. Neste sentido, uma alternativa é utilizar o amostrador *Hamiltoniano Monte Carlo (HMC)* baseado na linguagem de programação probabilística *STAN* (STAN | Development Team, 2018), que é uma das contribuições desta tese. Desta forma, o modelo *dnIGE* original foi computacionalmente implementado como um caso particular do modelo estendido, com base no vetor de parâmetros $\boldsymbol{\theta} = \left[\mathbf{a}_g^\top, \mathbf{a}_f^\top, \mathbf{e}_g^\top, \mathbf{e}_f^\top, \sigma_{S,g}^2, \sigma_{S,f}^2, \sigma_{E,g}^2, \sigma_{E,f}^2, \boldsymbol{\beta} \right]^\top$.

No próximo capítulo, vários estudos de simulação são apresentados para verificar o desempenho de estimação dos modelos *dnIGE* original e estendido.

EFEITO DE DIFERENTES DESENHOS EXPERIMENTAIS NA QUALIDADE DAS ESTIMATIVAS DO MODELO *dnIGE*

De acordo com [Sae-Lim e Bijma \(2016\)](#), [Pooley *et al.* \(2022\)](#), a comparação entre diferentes desenhos experimentais é importante porque auxilia na tomada de decisões, com relação ao planejamento experimental a ser realizado durante o melhoramento genético das características de efeitos sociais. Como visto na Seção 2.2, a infectividade é expressa a partir de interações sociais, logo é considerada um *IGE*. Desta forma, deve-se seguir o critério de que o espalhamento da doença infecciosa precisa ser observado numa população dividida em grupos independentes ([ANCHE; JONG; BIJMA, 2014](#)). Sendo assim, uma questão decisiva consiste na maneira com que essa divisão deve acontecer.

A literatura referente a modelos *IGE* apresenta alguns desenhos experimentais, usualmente utilizados para determinar como os indivíduos devem ser alocados em grupos para uma estimação adequada dos parâmetros dos modelos.

As populações às quais esta tese se refere, são constituídas de N animais geneticamente relacionados, gerados do acasalamento entre S pais e M mães para cada pai, configurando uma estrutura parental de meios-irmãos, cujos agrupamentos constituem as famílias. Sendo assim, descrevemos os principais métodos de alocação a seguir:

- (1) Alocação random.** A alocação *random* consiste em selecionar, ao acaso, os animais constituintes da amostra que serão alocados em cada um dos grupos independentes. Neste método, cada animal tem a mesma probabilidade de ser escolhido para compor um grupo específico do experimento. Desta forma, cada família contribui aleatoriamente com indivíduos em grupos, e, neste sentido, esta técnica garante que as possíveis diferenças, conhecidas ou desconhecidas, entre os grupos ocorrem devido ao acaso e não por algum

viés sistemático.

(2) **Alocação familiar por membros de duas famílias (2FAM).** Para melhor entendimento deste método, considere as Tabelas 1, 2 e 3.

Tabela 1 – Notações referentes ao método de alocação 2FAM.

Símbolo	Significado	Expressão matemática
S	Total de pais	-
M	Número de mães/pai	-
N	Tamanho da amostra (prole)	$N = S.M$
n_g	Número de grupos	-
n_w	Tamanho dos grupos	-
n_{gf}	Número de grupos por família	$n_{gf} = \frac{M}{(\frac{1}{2}n_w)}$
n_{fb}	Número de famílias por bloco	$n_{fb} = n_{gf} + 1$
n_{gb}	Número de grupos por bloco	$n_{gb} = (n_{gf}(1 + n_{gf}))/2$
n_b	Número de blocos	$n_b = N/(n_w.n_{gb})$

De acordo com [Bijma \(2010\)](#), a composição familiar com duas famílias por grupo (2FAM) consiste na seguinte forma:

- (i) Cada grupo deve ser composto por membros de duas famílias distintas, de maneira que cada família (formada por agrupamento de meios-irmãos) contribua com uma quantidade de membros equivalente à metade do tamanho do grupo ($\frac{1}{2}n_w$);
- (ii) Logo, temos que cada família pode contribuir para a formação de $n_{gf} = \frac{M}{(1/2.n_w)}$ grupos, de maneira que o tamanho n_w de cada grupo seja um número par e que cada família seja combinada com todas as outras n_{gf} famílias;
- (iii) A escolha das famílias em cada grupo também não é aleatória, mas realizada na estrutura de blocos, que são compostos por uma determinada quantidade de famílias, $n_{fb} = n_{gf} + 1$. O número de grupos em cada bloco é o número de possíveis combinações de famílias dentro do bloco, dado por $n_{gb} = (n_{gf}(1 + n_{gf}))/2$;
- (iv) Finalmente, a expressão que determina o número de blocos é obtida pela razão entre o tamanho da amostra (N) e o produto entre o tamanho do grupo e o número de grupos por bloco, $n_b = N/(n_w.n_{gb})$.

Como exemplo, as Tabelas 2 e 3 representam o esquema de alocação 2FAM para 1 dos 50 blocos que podem ser obtidos a partir de uma amostra de tamanho $N = 2000$, gerada de 250 famílias de tamanho 8. Nesse esquema é possível formar 500 grupos de tamanho 4, de maneira que cada bloco seja composto por 10 grupos. Esses grupos são formados pela combinação entre as 5 famílias que pertencem a cada bloco.

Tabela 2 – Configurações do exemplo.

Símbolo	Valores
S	250
M	8
N	2000
ng	500
nw	4
ngf	4
nfb	5
ngb	10
nb	50

Tabela 3 – Combinação de famílias para formação de 10 grupos de tamanho 4 em um bloco.

Família A	Família B			
	2	3	4	5
1	2/2	2/2	2/2	2/2
2		2/2	2/2	2/2
3			2/2	2/2
4				2/2

É importante ressaltar que o total de genitores masculinos determina o número de famílias, pois do cruzamento de S pais e M mães para cada pai, um filho é gerado em cada cruzamento, resultando nas famílias formadas por agrupamentos de meios-irmãos. Sendo assim, esse número de famílias deve ser sempre um inteiro maior ou igual a três, para que seja possível obter mais de uma combinação e consequentemente mais de um grupo. Por definição, o tamanho do grupo deve ser um número par. Então neste caso, se o número de famílias for um número ímpar, o tamanho da família deve ser par e de forma que seja possível obter um inteiro positivo divisível por n_w . Além disso, uma vez que as expressões matemáticas que constituem este método estão relacionadas, é preciso ter o cuidado na escolha desses valores para que não ocorra situações não realistas como, por exemplo, obter um número fracionário para o total de blocos ou grupos.

(3) Alocação familiar por membros de três famílias (3FAM). Neste esquema, desenvolvido por [Ødegård e Olesen \(2011\)](#), cada grupo é composto por membros de três famílias distintas, de forma que cada família contribui com um número de membros equivalente a um terço do tamanho do grupo ($\frac{1}{3} \cdot n_w$). Assim, cada família contribui para $n_{gf} = \frac{M}{(1/3 \cdot n_w)}$ grupos. A partir desta descrição obtemos a estrutura de blocos, compostos por determinado número de grupos em cada um deles. O número de famílias por bloco é dado pela expressão $n_{fb} = 2n_{gf} + 1$, de forma que cada família de um bloco seja combinada com outras n_{gf} famílias do mesmo bloco. Assim, o número de grupos por bloco é obtido pelo produto entre o número de grupos por família e sua redução em uma unidade, $n_{gb} = n_{gf}(n_{gf} - 1)$. O número total de blocos é obtido pela mesma expressão do método anterior.

[Sae-Lim e Bijma \(2016\)](#) analisaram o desempenho dos métodos de alocação 2FAM e 3FAM. Entre seus resultados consta que os dois desenhos pouco diferem na precisão das estimativas dos parâmetros genéticos diretos e sociais. Sendo assim, consideraremos nesta tese os métodos de alocação 2FAM e *random*. O primeiro com baixa diversidade genética dentro dos grupos e o segundo com grande variação genética dentro dos grupos.

Até o momento, no modelo *dnIGE* foi considerada unicamente a forma de composição *random* (ANACLETO *et al.*, 2015). Sendo assim, o objetivo do estudo de simulação a ser apresentado neste capítulo consiste em verificar se a alocação familiar (isto é, grupos formados por animais geneticamente relacionados) resulta em melhores estimativas dos parâmetros genéticos, conforme verificado em estudos prévios para modelos *IGE* baseados em modelos mistos, com variável resposta seguindo distribuição normal (BIJMA, 2010; ØDEGÅRD; OLESEN, 2011; SAE-LIM; BIJMA, 2016).

5.1 Simulação das populações de base

As populações de base foram geradas de acordo com as etapas a seguir:

- (i) **Arquitetura genética dos genitores.** Consideramos $S = 100$ pais e $M = 20$ mães para cada pai, de forma que não fossem aparentados, ou seja, não correlacionados geneticamente. Desta forma, seus efeitos genéticos aditivos de susceptibilidade e infectividade foram obtidos através da distribuição normal bivariada tal que

$$\begin{aligned} & \bullet \mathbf{a}_s | \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{S,g}^2 & \rho_A \sigma_{S,g} \sigma_{S,f} \\ \rho_A \sigma_{S,g} \sigma_{S,f} & \sigma_{S,f}^2 \end{bmatrix} \right), \text{ em que } \mathbf{a}_s = \begin{bmatrix} a_{g,s} \\ a_{f,s} \end{bmatrix}, \text{ do} \\ & \quad s\text{-ésimo pai, para } s = 1, \dots, S; \text{ e} \\ & \bullet \mathbf{a}_m | \sigma_{M,g}^2, \sigma_{M,f}^2, \rho_A \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{M,g}^2 & \rho_A \sigma_{M,g} \sigma_{M,f} \\ \rho_A \sigma_{M,g} \sigma_{M,f} & \sigma_{M,f}^2 \end{bmatrix} \right), \text{ em que } \mathbf{a}_m = \begin{bmatrix} a_{g,m} \\ a_{f,m} \end{bmatrix}, \\ & \quad \text{da } m\text{-ésima mãe, para } m = 1, \dots, M. \end{aligned}$$

As notações $\sigma_{S,g}^2$, $\sigma_{S,f}^2$, $\sigma_{M,g}^2$, $\sigma_{M,f}^2$ representam as variâncias genéticas de susceptibilidade e infectividade dos genitores masculinos e femininos, respectivamente; e ρ_A é o coeficiente de correlação genética entre susceptibilidade e infectividade. Nos vetores são alocados consecutivamente os efeitos genéticos de ambas as características correspondentes a cada animal. Os valores de variância e coeficientes de correlação utilizados em cada cenário de simulação são variados e definidos nas próximas seções.

- (ii) **Arquitetura genética da prole.** Tendo em vista que a geração parental é selecionada para reprodução, a prole consiste, então, de $N = 2000$ animais, configurando uma estrutura parental de meios-irmãos. De acordo com o modelo básico da Genética Quantitativa, definido no Capítulo 2, os efeitos genéticos aditivos de cada animal da prole podem ser obtidos pela soma entre a média dos efeitos genéticos aditivos dos genitores e um termo mendeliano, simulado de uma distribuição normal com média zero, e variância correspondente à metade da variância genética aditiva da característica de interesse. Os efeitos ambientais para ambas as características foram simulados de acordo com (4.6). Em seguida, utilizando os modelos de componentes de variância definidos nas Equações

(4.2) e (4.3), as fragilidades associadas a essas duas características foram calculadas para a população de base.

Tabela 4 – Configurações da população de base nos métodos de alocação.

Item	<i>2FAM</i>	<i>random</i>
S	100	100
M	20	20
N	2000	2000
n_g	200	200
n_w	10	10
n_{gf}	4	-
n_{fb}	5	-
n_{gb}	10	-
n_b	20	-

- (iii) **Alocação em grupos.** Após a geração da prole, alocamos esses animais em grupos independentes de tamanhos iguais e de acordo com os dois esquemas de alocação a serem avaliados, seguindo a configuração descrita na Tabela 4.
- (iv) **Geração da epidemia.** Posteriormente, a transmissão da doença infecciosa foi simulada, de forma que em cada grupo um animal foi escolhido, ao acaso, para ser o caso índice responsável pelo início do surto epidêmico. Utilizamos o método direto de Gillespie (GILLESPIE, 1977; KEELING; ROHANI, 2011) para gerar a epidemia em cada um deles. Os tempos entre as infecções foram simulados a partir de uma distribuição exponencial com parâmetro dado pela soma das taxas de infecção dos animais ainda não infectados, representada por $\lambda_j(t)$ na Equação (4.1), assumindo por simplicidade $\beta = 1$. Após especificar o tempo de ocorrência da próxima infecção, o animal correspondente ao tempo referido foi escolhido aleatoriamente entre os susceptíveis restantes, considerando suas taxas de infecção como pesos relativos. A suposição de independência entre os grupos garante que nenhuma transmissão da doença tenha ocorrido entre animais pertencentes a grupos diferentes.

Os cenários foram definidos variando a população, o tamanho do grupo e as herdabilidades de susceptibilidade e infectividade em escala de fragilidade (definindo a variação ambiental desconhecida como 1). A Tabela 5 apresenta todos os cenários considerados para avaliar o modelo *dnIGE*.

Variamos o tamanho dos grupos em 2, 10 e 20 e os valores das variâncias na simulação dos dados de forma a obter herdabilidade (h^2) de 0.40 e 0.80 para susceptibilidade e infectividade. Para cada cenário (combinação de tamanho de grupo e herdabilidade dos efeitos genéticos) geramos 20 conjuntos de dados diferentes (ou seja, réplicas, cuja quantidade foi determinada de acordo com (ANACLETO *et al.*, 2015)) e cada réplica foi analisada nas alocações *random*

e *2FAM*. Os efeitos ambientais para ambas as características foram simulados a partir de uma distribuição normal com média zero e variância unitária.

Tabela 5 – Valores de herdabilidade, tamanho da amostra, e tamanho do grupo considerados nas epidemias simuladas para validação do modelo *dnIGE* original e estendido.

Susceptibilidade h_s^2 e Infectividade h_f^2	Infectividade h_f^2	Tamanho da amostra	Tamanho do grupo
0.4	0.1	1980	2
	0.33	2000	10
0.8	0.67		
	0.8	2100	20

Nas simulações estocásticas para grupos de tamanho diferente de 10, definimos o número de famílias de maneira que o número de blocos fosse um número inteiro, mas o mais próximo possível de 100. As famílias sempre tiveram o mesmo tamanho, 20.

5.2 Software para estimação dos parâmetros do modelo *dnIGE*

Como já mencionado na Seção 4.2, consideramos um modelo de risco paterno para os componentes de variância de susceptibilidade e infectividade (MRODE, 2014). Para a estimação dos parâmetros do modelo *dnIGE* estendido, desenvolvemos um pacote do *R*, denominado *dnIGEmodel* (ANACLETO; ZUANETTI; LIMA, 2023), onde implementamos um algoritmo de inferência Bayesiana, baseado em *Stan* (CARPENTER *et al.*, 2017), que permite estimar os efeitos genéticos dos genitores e herdabilidades associadas à susceptibilidade e infectividade. Neste sentido, a metodologia *MCMC* de estimação foi implementada no *dnIGEmodel* utilizando o pacote *Rstan* (STAN | Development Team, 2022), no qual não precisamos especificar diretamente as distribuições condicionais completas. No entanto, para fins de ilustração e informação, essas distribuições são apresentadas no Apêndice A, considerando as distribuições *a priori* (4.5) e (4.6). É importante ressaltar que o *Stan* implementa o algoritmo *NUTS* (*no-u-turn sample*), que é uma variação do método *HMC* (HOFFMAN; GELMAN *et al.*, 2014).

Para cada conjunto de dados, duas cadeias com pontos iniciais diferentes foram geradas para avaliar a convergência das cadeias de *MCMC*. Isto foi garantido através da observação de gráficos de traços (*trace plots*) e do cálculo da estatística de convergência de Gelman-Rubin para os parâmetros do modelo (GELMAN *et al.*, 2013). Cada cadeia foi composta por 11000 iterações, de modo que seus primeiros 1000 valores foram descartados no *burn-in*, e cada quinto valor das iterações restantes foi utilizado para calcular as estimativas finais.

Antes de iniciarmos as próximas sessões, registramos aqui a observação de que uma herdabilidade de 0.8 para as fragilidades de susceptibilidade e infectividade corresponde a uma

herdabilidade consideravelmente menor (≈ 0.2 , conforme verificado em estudos de simulação) para o *status* de infecção binária registrado, o que está de acordo com os valores registrados na literatura (LIPSCHUTZ-POWELL *et al.*, 2012a).

5.3 Resultados do modelo *dnIGE* original

Para analisar o efeito dos tipos de alocação na estimação dos valores genéticos de susceptibilidade e infectividade paternos, diferentes cenários foram considerados. Primeiramente, assumimos independência entre os efeitos genéticos e ambientais de susceptibilidade e infectividade e utilizamos o modelo *dnIGE* original para analisar os conjuntos de dados. Escolhemos a distribuição $Gama(0.001;0.001)$ como distribuição *a priori* para o inverso das variâncias (precisões) envolvidas no modelo (GELMAN, 2006).

5.3.1 Estimativas de herdabilidade

Apresentamos nas Figuras 6 e 7 os resumos da distribuição *a posteriori* para a herdabilidade de susceptibilidade e infectividade em escala de fragilidade, que nos mostram os intervalos de credibilidade de 90% para esses parâmetros ao ajustar o modelo *dnIGE* original com tamanho de grupo 10 (cenário base). Cada intervalo foi obtido a partir das amostras das variâncias fornecidas pelo *MCMC* em cada replicação, o que permitiu a obtenção das amostras para as herdabilidades, estimativas (médias *a posteriori*) e quartis. Essa configuração abrange os métodos de alocação *random* e *2FAM*.

Para a susceptibilidade, podemos observar que no método de alocação *random*, cerca de 85% dos intervalos contêm as herdabilidades verdadeiras utilizadas para simular os dados, enquanto que no método *2FAM* essa porcentagem é próxima de 90%. Para a infectividade, aproximadamente 95% dos intervalos de credibilidade contêm os verdadeiros valores da herdabilidade na alocação *random*, enquanto que essa porcentagem é de 85% na alocação *2FAM*. Isto significa que as probabilidades de cobertura obtidas são próximas às probabilidades de cobertura nominais, o que mostra que a herdabilidade das características é apropriadamente estimada em ambos os métodos de alocação.

Quanto à amplitude, podemos notar que os intervalos de credibilidade para herdabilidades de infectividade são mais largos do que os que são referentes à susceptibilidade, indicando que a estimação da variação genética para esta característica é mais difícil.

Resultados similares foram obtidos para grupos de tamanho 2 e 20.

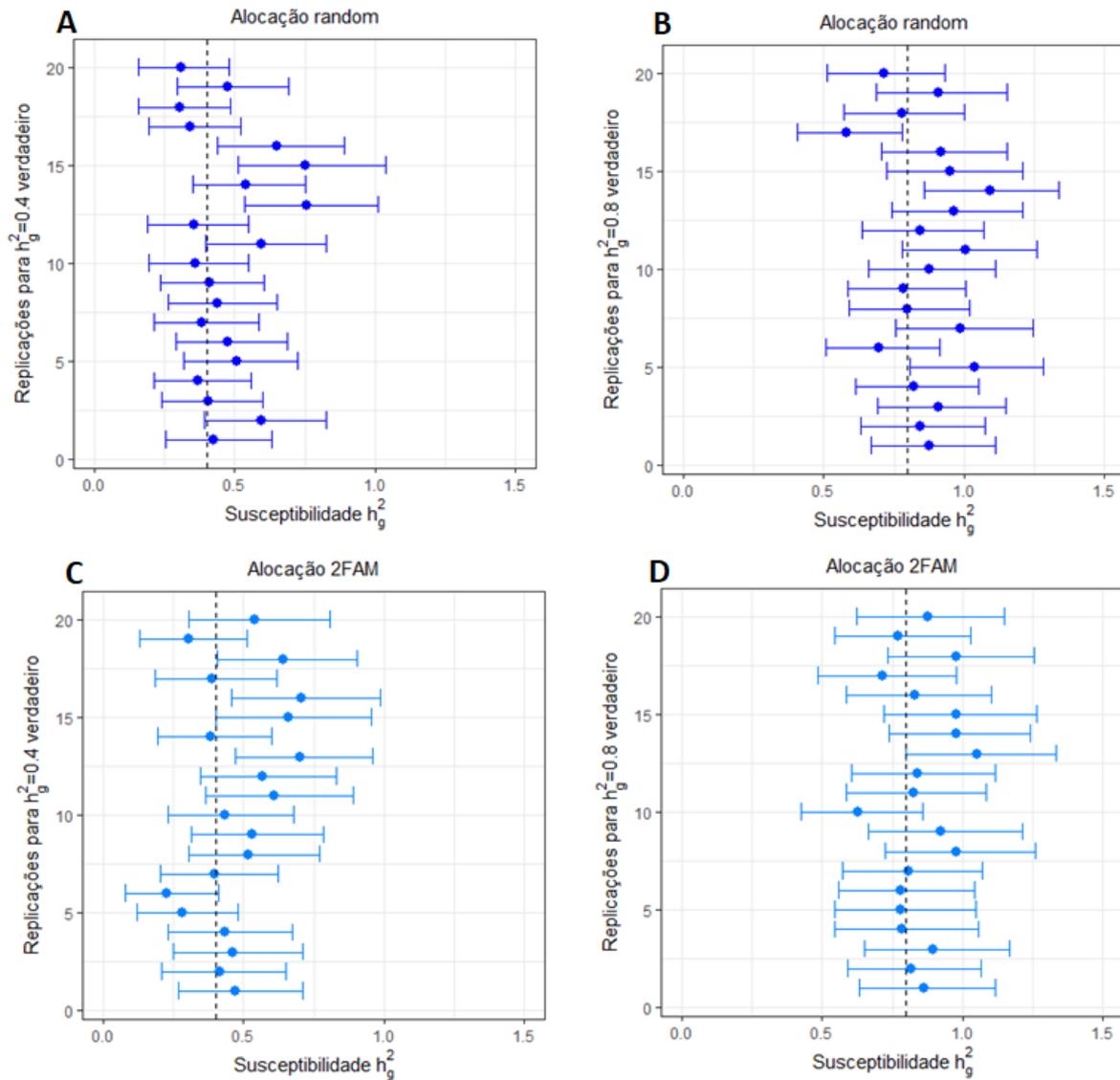


Figura 6 – Intervalos de credibilidade de 90% para herdabilidades de susceptibilidade para grupos de tamanho 10 (cenário base). As linhas tracejadas indicam herdabilidades verdadeiras e os pontos representam as médias a posteriori.

Fonte: Elaborada pelo autor.

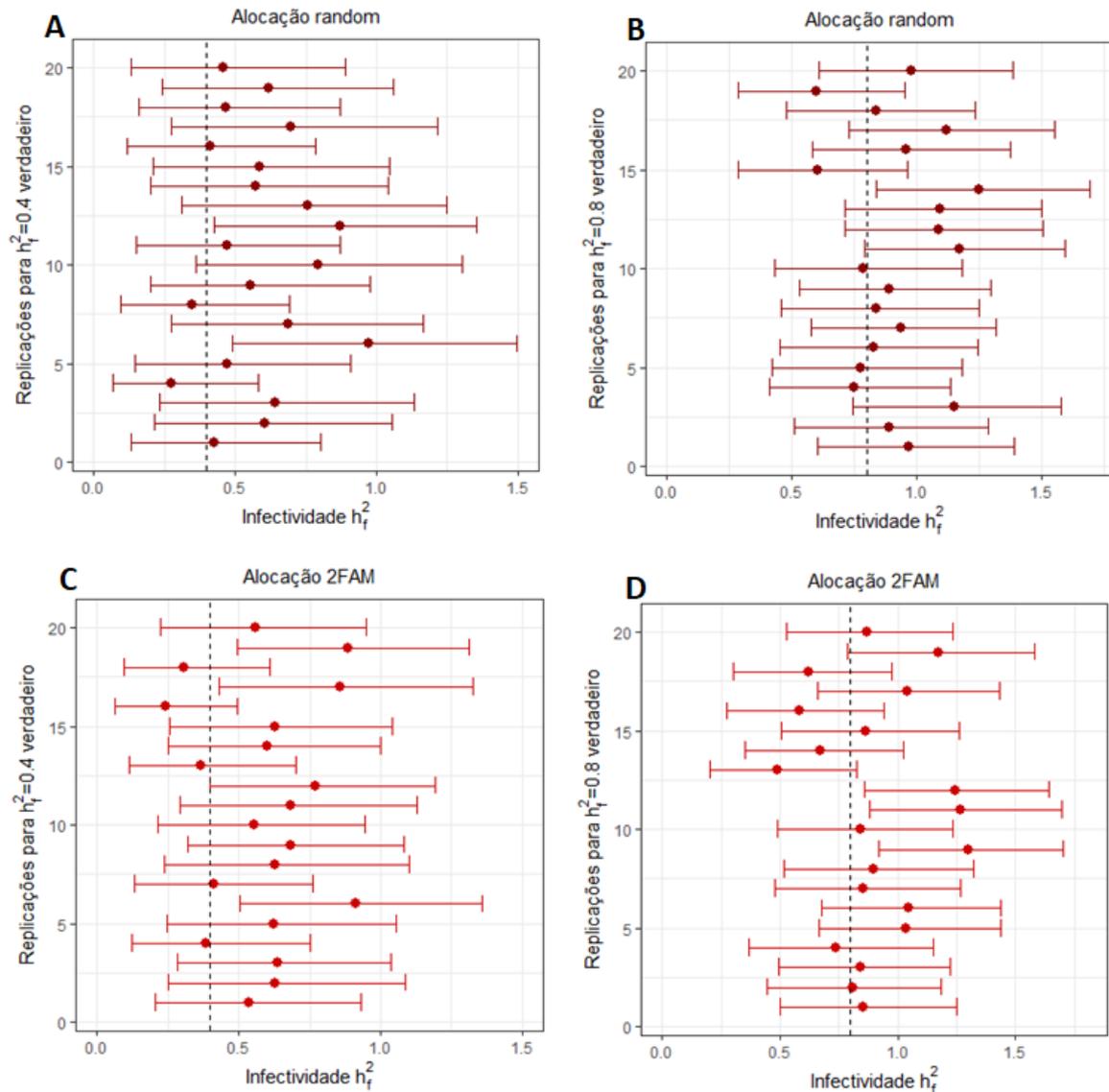


Figura 7 – Intervalos de credibilidade de 90% para herdabilidades de infectividade para grupos de tamanho 10 (cenário base). As linhas tracejadas indicam herdabilidades verdadeiras e os pontos representam as médias a posteriori.

Fonte: Elaborada pelo autor.

5.3.2 Estimativas de acurácia de predição

A Figura 8 mostra o impacto da diversidade genética na infectividade nas acurácias de predição dos efeitos genéticos aditivos paternos de susceptibilidade e infectividade, obtidos a partir do modelo *dnIGE* e do modelo alternativo (que considera somente a susceptibilidade) para os métodos de alocação *random* e *2FAM*. Neste cenário, fixamos grupos de tamanho 10, herdabilidade de susceptibilidade 0.8 e variamos os valores de herdabilidade de infectividade em 0.10, 0.33, 0.67, e 0.80. As acurácias de predição são definidas como a média das médias (obtidas entre as 20 replicações) das correlações entre os efeitos genéticos paternais verdadeiros e estimados, de forma que quanto mais próximo de 1 este valor estiver, mais próximo o valor

estimado estará do valor real. As barras representam essa precisão média da previsão, enquanto que as linhas representam essa estimativa mais ou menos seu erro padrão em 20 réplicas.

Assim como em [Anacleto et al. \(2015\)](#), na Figura 8 (A) percebe-se que há uma tendência crescente nas acurácias de predição para a infectividade, com precisão crescente (refletido por erros padrão menores) quando ocorre o aumento da herdabilidade na infectividade. Além disso, as previsões de susceptibilidade não foram afetadas pela diversidade genética na infectividade. Por outro lado, quando se utilizou o modelo alternativo - que ignora a diversidade genética na infectividade - a variação genética nesta característica levou à redução da acurácia de predição da susceptibilidade nesse modelo, bem como a precisão dessas estimativas. O mesmo comportamento pode ser observado na Figura 8 (B), que considera a alocação familiar. Os gráficos mostram também que a queda nas acurácias de predição da susceptibilidade, para valores maiores da infectividade, é mais acentuada no método de alocação *2FAM*.

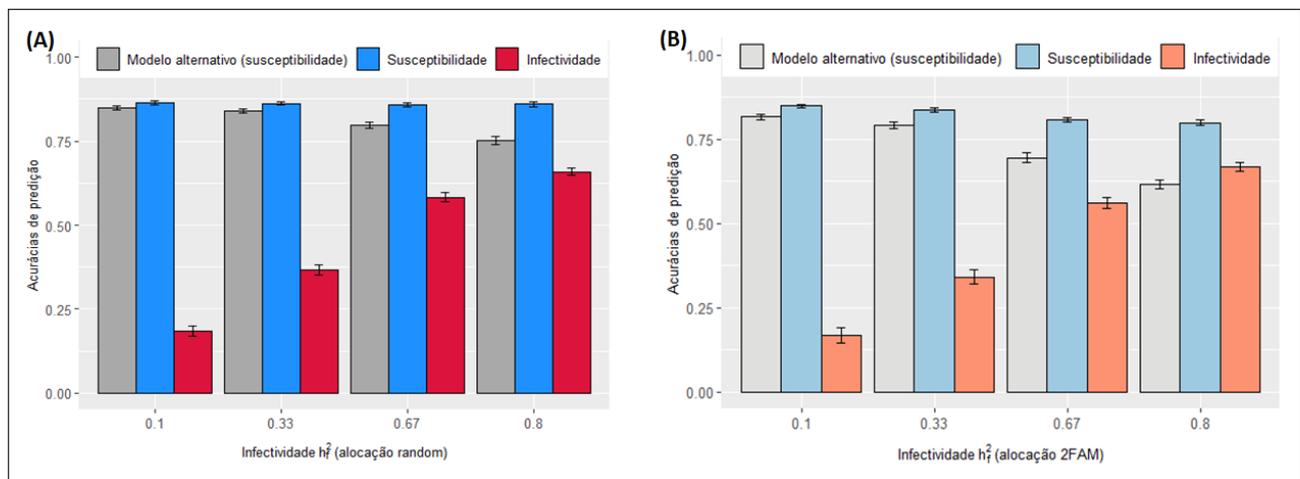


Figura 8 – Influência das variações genéticas da infectividade nas estimativas do modelo *dnIGE* em comparação com a abordagem alternativa de [Korsgaard, Madsen e Jensen \(1998\)](#).

Fonte: Elaborada pelo autor.

Estes resultados reforçam que a variação genética na infectividade pode afetar negativamente as precisões dos modelos que explicam a variação genética apenas na susceptibilidade, e enfatizam que o modelo *dnIGE* pode prever eficientemente os efeitos genéticos, tanto na susceptibilidade quanto na infectividade, com precisão de predição de risco genético paterno de infectividade dependendo da variação genética desta característica.

5.3.3 Efeitos de tamanho do grupo e herdabilidade nas estimativas de acurácias de predição

A Figura 9 nos mostra simultaneamente os efeitos do tipo de alocação, do tamanho do grupo e da variação genética (representada por diferentes valores de herdabilidade) no desempenho do modelo *dnIGE* para estimativas de acurácias de predição. As herdabilidades

consideradas para susceptibilidade e infectividade foram 0.4 e 0.8 para os grupos de tamanho 2, 10 e 20 e os desenhos experimentais testados em cada grupo foram as alocações *random* e familiar com membros de duas famílias.

Percebe-se que melhores acurácias de predição são obtidas com valores maiores de herdabilidade, tanto para a susceptibilidade, Figura 9 (A) e (B), quanto para a infectividade, Figura 9 (C) e (D). Além disso, a qualidade das estimativas foi, em geral, inferior para a infectividade quando comparada às estimativas da susceptibilidade, exceto quando o tamanho dos grupos é 2, que apresenta acurácia semelhante para ambas as características. Isto é refletido pelas acurácias médias mais baixas e erros padrão maiores das estimativas de herdabilidade associadas à infectividade.

Quanto à comparação entre os tipos de alocação, percebemos que a precisão da predição dos efeitos genéticos paternos de ambas as características pouco diferem quando o tamanho do grupo é muito pequeno, não importando, neste caso, o método de alocação. No entanto, na medida em que os grupos crescem, as acurácias de predição de susceptibilidade são maiores, principalmente para o tipo *random* de alocação. O mesmo não acontece para as acurácias de predição de infectividade, cuja melhoria é obtida com a redução do tamanho dos grupos, tanto na alocação *2FAM* quanto *random*, sendo o método *2FAM* com melhor desempenho em grupos maiores.

Vemos ainda, que o aumento da precisão da predição dos efeitos genéticos de susceptibilidade com o crescimento do grupo na alocação *random* reflete um *trade-off* entre prever efeitos genéticos de infectividade e susceptibilidade com relação ao tamanho do grupo, já que as acurácias de predição para os efeitos genéticos de infectividade seguem tendência oposta neste método de alocação. De acordo com [Anacleto et al. \(2015\)](#), isso pode ser explicado intuitivamente da seguinte forma: como os animais altamente susceptíveis tendem a se infectar com pressão de infecção mais baixa e, conseqüentemente, mais cedo em relação a indivíduos menos susceptíveis, e a pressão de infecção individual aumenta com o número de infectados, então grupos maiores fornecem mais informações sobre a ordem de infecção. Isso resulta em melhor acurácia e precisão das estimativas dos parâmetros genéticos de susceptibilidade. Por outro lado, na ausência de informações detalhadas sobre quem infecta quem, grupos maiores aumentam o potencial de confusão entre as expressões de infectividade de diferentes animais infectados, se tornando, portanto, mais difícil separar as infectividades individuais dos indivíduos infectados que podem ter transmitido a infecção aos susceptíveis à medida que o tamanho do grupo aumenta.

Este estudo, ao mostrar que a precisão na previsão dos efeitos genéticos de infectividade tende a ser maior quando os membros do grupo são alocados de forma familiar, concorda com outros estudos já documentados na literatura ([HEIDARITABAR et al., 2019](#); [ØDEGÅRD; OLESEN, 2011](#); [PEETERS et al., 2012](#); [BIJMA, 2010](#)). Os resultados apresentados em [Ødegård e Olesen \(2011\)](#) mostram que a precisão do efeito genético direto é maior quando os grupos são

compostos aleatoriamente. Aqui verificamos que as acurácias de predição de susceptibilidade, que possui efeito genético direto, são mais eficientes em grupos do tipo *random* de alocação, estando portanto em concordância com estes resultados.

Desta forma, podemos concluir que o método *dnIGE* com alocação *random*, utilizado em [Anacleto et al. \(2015\)](#), não é o mais eficiente para estimar efeitos genéticos de infectividade. No entanto, é eficiente para estimação de efeitos genéticos de susceptibilidade.

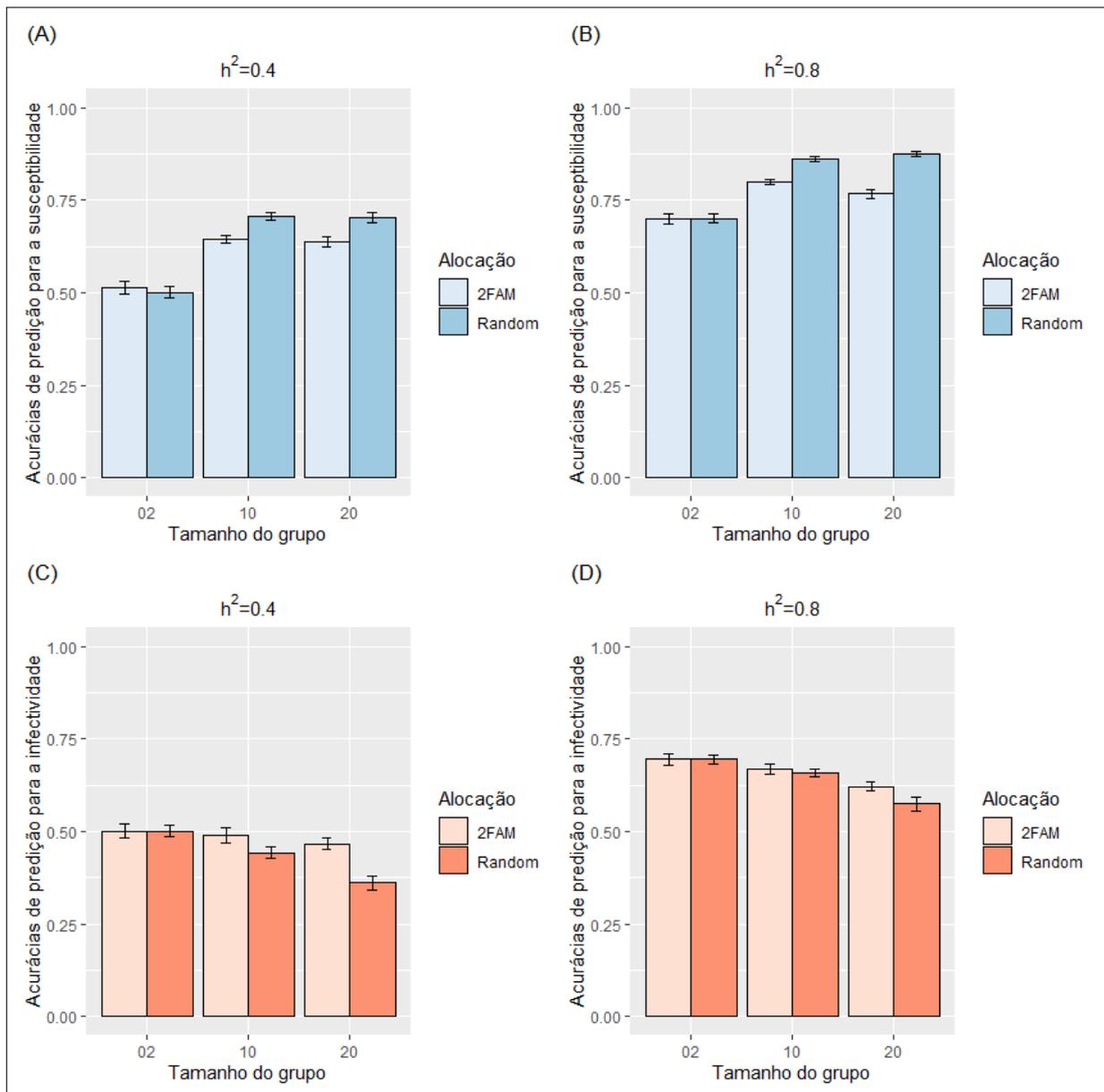


Figura 9 – Efeito do tamanho e do tipo de composição dos grupos nas estimativas do modelo *dnIGE*.

Fonte: Elaborada pelo autor.

5.4 Resultados do modelo *dnIGE* estendido

A população base deste segundo estudo foi gerada de forma análoga a do primeiro estudo, mas tomando como cenário base grupos de tamanho 20. Foram considerados 99 genitores masculinos e 20 genitoras para cada um deles. De cada cruzamento foi gerado um filho, resultando numa prole formada por 1980 animais. Em seguida, consideramos os métodos de alocação em grupos para esses animais. No método de alocação familiar, a contribuição de cada família foi de 10 de seus membros para compor a metade do tamanho de determinado grupo. O número de grupos por família foi 2 e um total de 33 blocos, sendo cada um deles formado por 3 famílias que se dividiram em 3 grupos. Fixamos grupos de tamanho 20 e herdabilidade de 0.80 para ambas as características, ver Tabela 6.

Tabela 6 – Configurações para o modelo *dnIGE* estendido.

Símbolo	<i>2FAM</i>	<i>random</i>
S	99	99
M	20	20
N	1980	1980
ng	99	99
n_w	20	20
n_{gf}	2	-
n_{fb}	3	-
n_{gb}	3	-
nb	33	-

Variamos os valores das correlações ρ_A e ρ_E em 0, ± 0.2 , ± 0.4 , ± 0.6 , ± 0.8 na simulação dos dados. Desta forma, cada cenário foi formado pela combinação de valores de ρ_A com a herdabilidade dos efeitos genéticos aditivos e grupos de tamanho 20. Geramos então 20 réplicas, que foram uma a uma analisadas nos dois métodos de alocação, *random* e *2FAM*.

Utilizamos a versão estendida do modelo *dnIGE*, apresentada na Seção 4.2, que considera a correlação genética entre susceptibilidade e infectividade, para analisar os conjuntos de dados. Tendo em vista que o modelo é hierárquico, escolhemos a distribuição *Lewandowski-Kurowicka-Joe (LKJ)* (LEWANDOWSKI; KUROWICKA; JOE, 2009) com parâmetro escalar unitário como sendo a distribuição *a priori* para as correlações genéticas envolvidas no modelo. Esta distribuição, que corresponde à distribuição uniforme no espaço de matrizes de correlação, foi atribuída no fator de *Cholesky* da matriz de variância e covariância dos efeitos genéticos. Aos desvios padrões atribuímos como distribuição *a priori* a distribuição de *Half-Cauchy (0,5)* (GELMAN, 2006).

Para a distribuição *a priori* dos efeitos ambientais, consideramos as diferentes contribuições da prole sobre cada característica. Para os casos em que os animais contribuem para ambas, *a priori* foi a distribuição normal bivariada, definida em (4.6). Para os casos em que os animais

contribuem apenas para uma delas, *a priori* foi uma distribuição normal univariada com média zero e desvio padrão σ_E .

5.4.1 Estimativas de herdabilidade

As Figuras 10, 11, 12, 13, 14, 15 mostram os resumos da distribuição *a posteriori* para a herdabilidade de susceptibilidade e infectividade em escala de fragilidade e para as correlações genéticas, através dos intervalos de credibilidade de 90% para esses parâmetros ao ajustar o modelo *dnIGE* estendido com tamanho de grupo 20 (cenário base). Essa configuração abrange os métodos de alocação *random* e *2FAM*, respectivamente.

Em geral, podemos observar que cerca de 90% dos intervalos contêm as herdabilidades e os coeficientes de correlação genética verdadeiros utilizados para simular os dados, indicando que as probabilidades de cobertura obtidas são próximas as probabilidades de cobertura nominais, indicando que as herdabilidades e correlações genéticas das características epidêmicas são apropriadamente estimadas em ambos os métodos de alocação.

Quanto às amplitudes, os intervalos de credibilidade para herdabilidades de infectividade permanecem maiores do que os que são referentes à susceptibilidade, indicando que a estimação da variação genética para esta característica ainda é mais difícil.

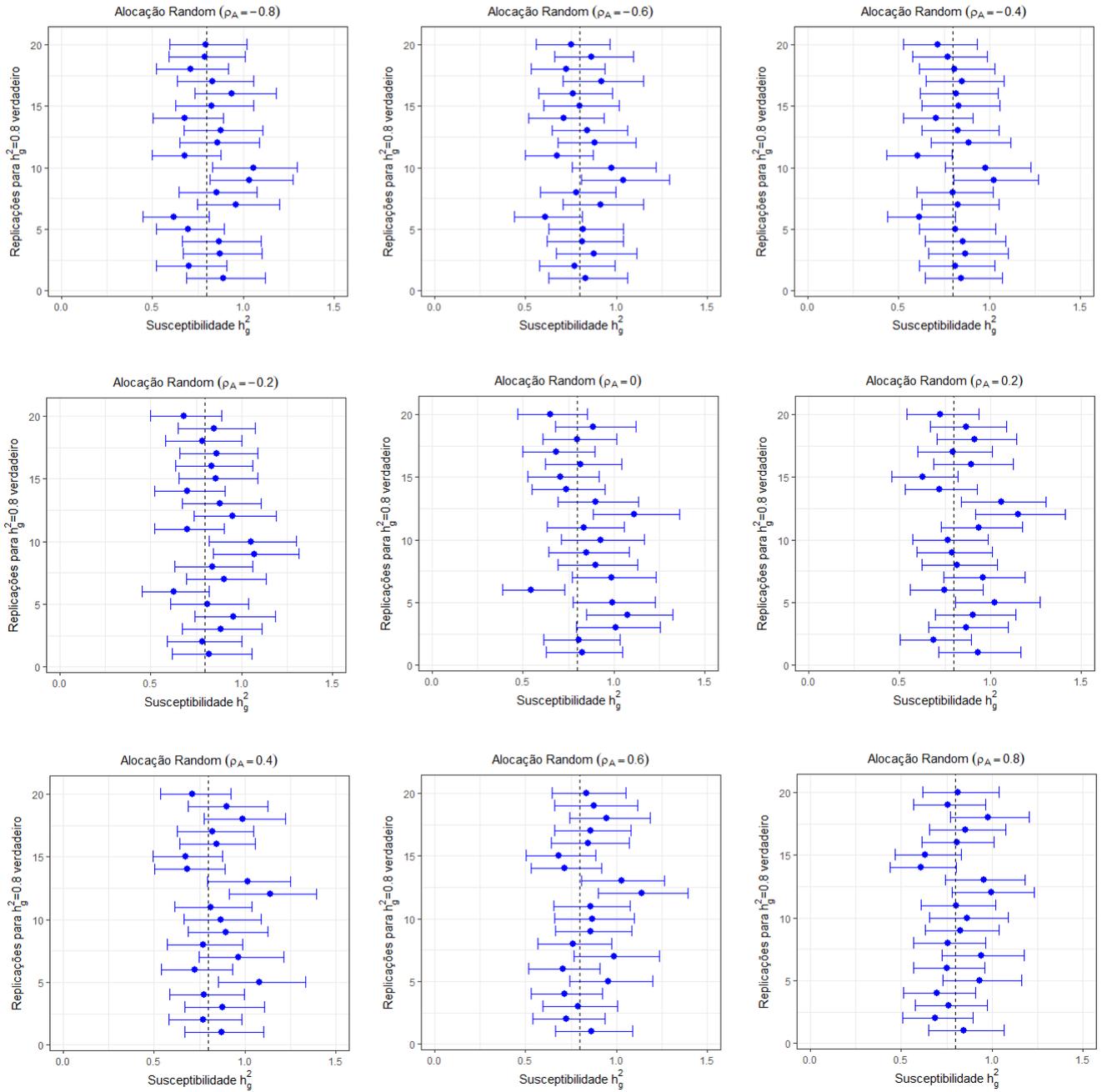


Figura 10 – Intervalos de credibilidade para a herdabilidade da susceptibilidade quando o método de alocação é *random* e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.

Fonte: Elaborada pelo autor.

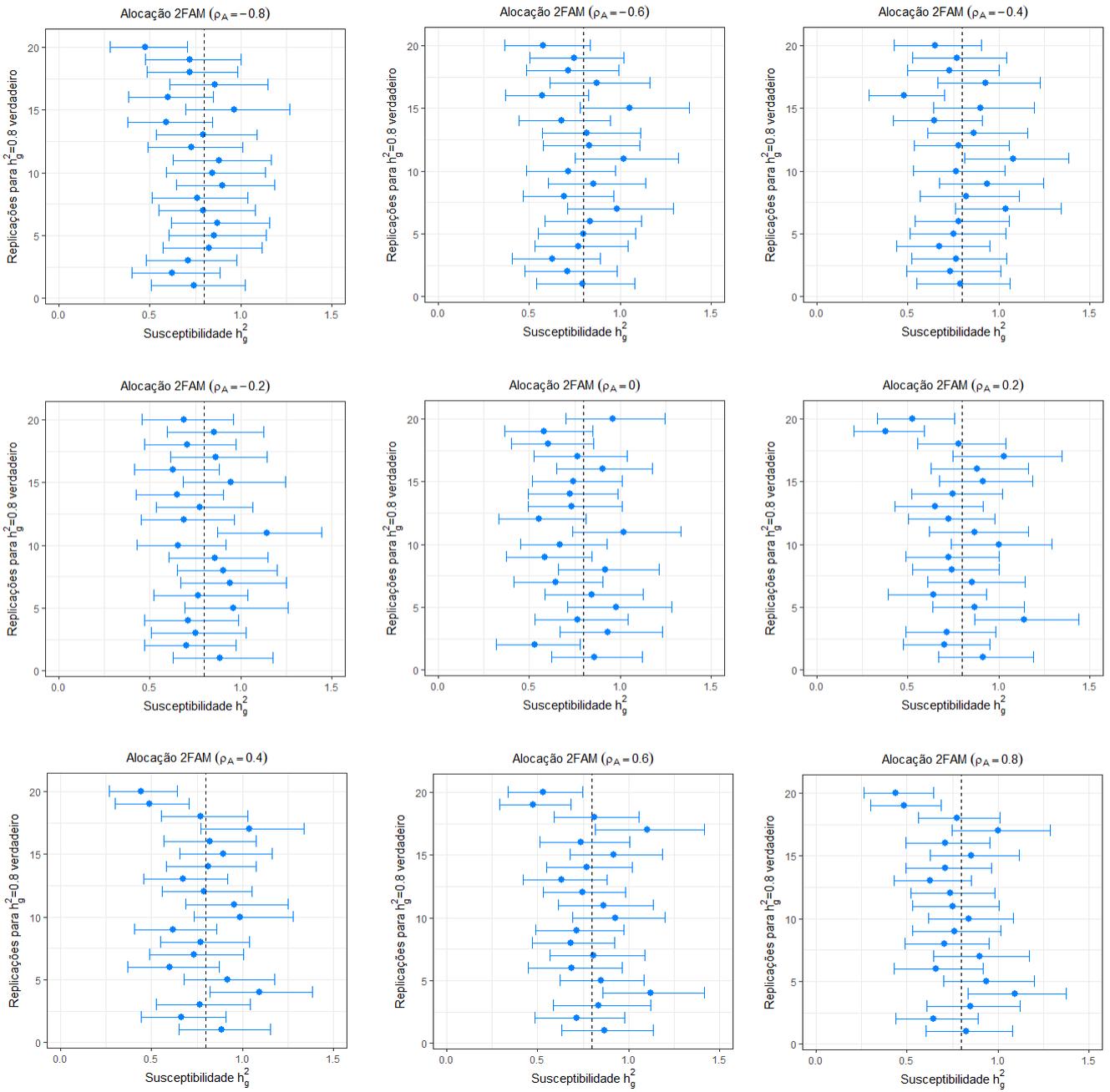


Figura 11 – Intervalos de credibilidade para a herdabilidade da susceptibilidade quando o método de alocação é 2FAM e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.

Fonte: Elaborada pelo autor.

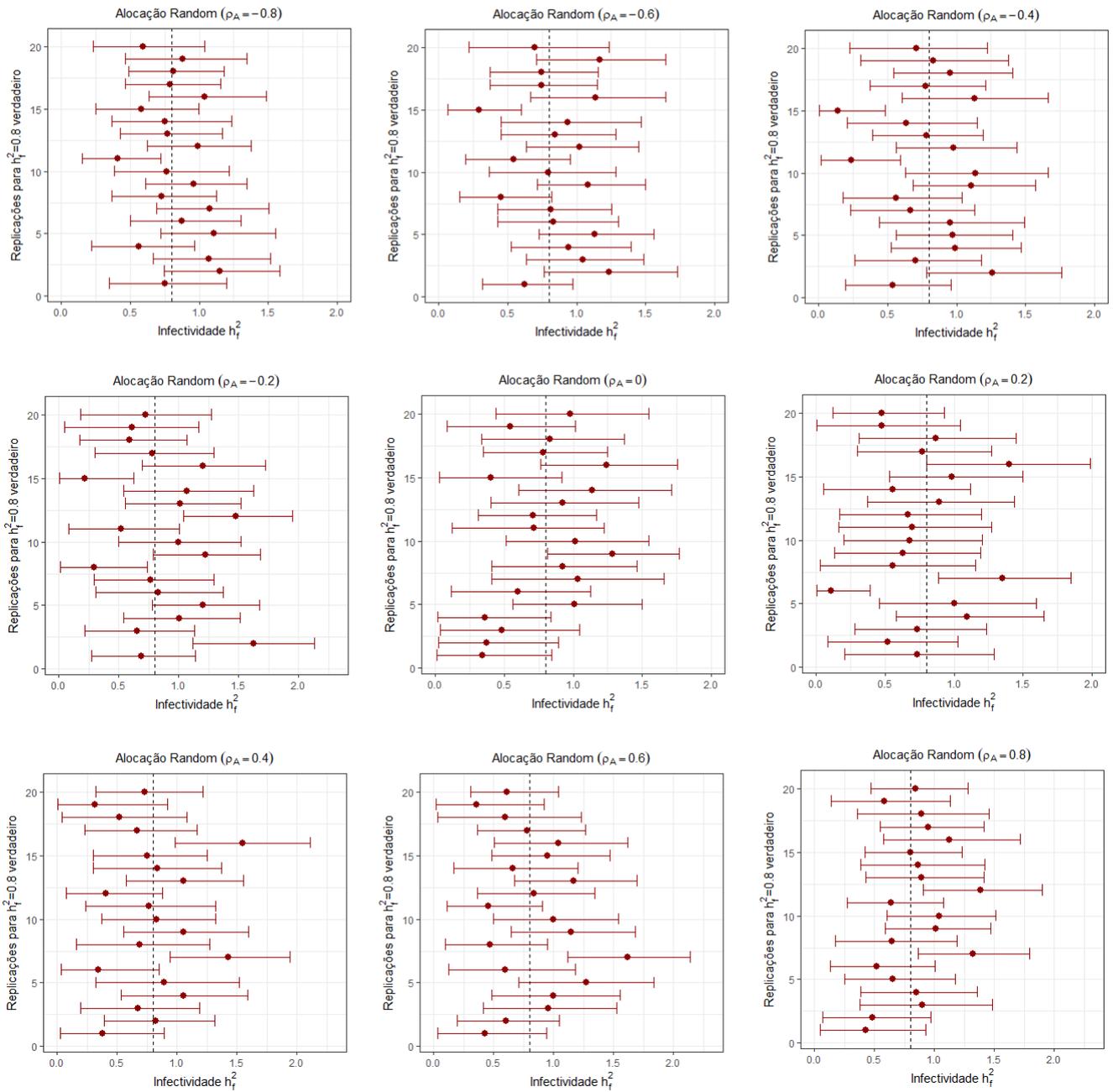


Figura 12 – Intervalos de credibilidade para a herdabilidade da infectividade quando o método de alocação é *random* e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.

Fonte: Elaborada pelo autor.

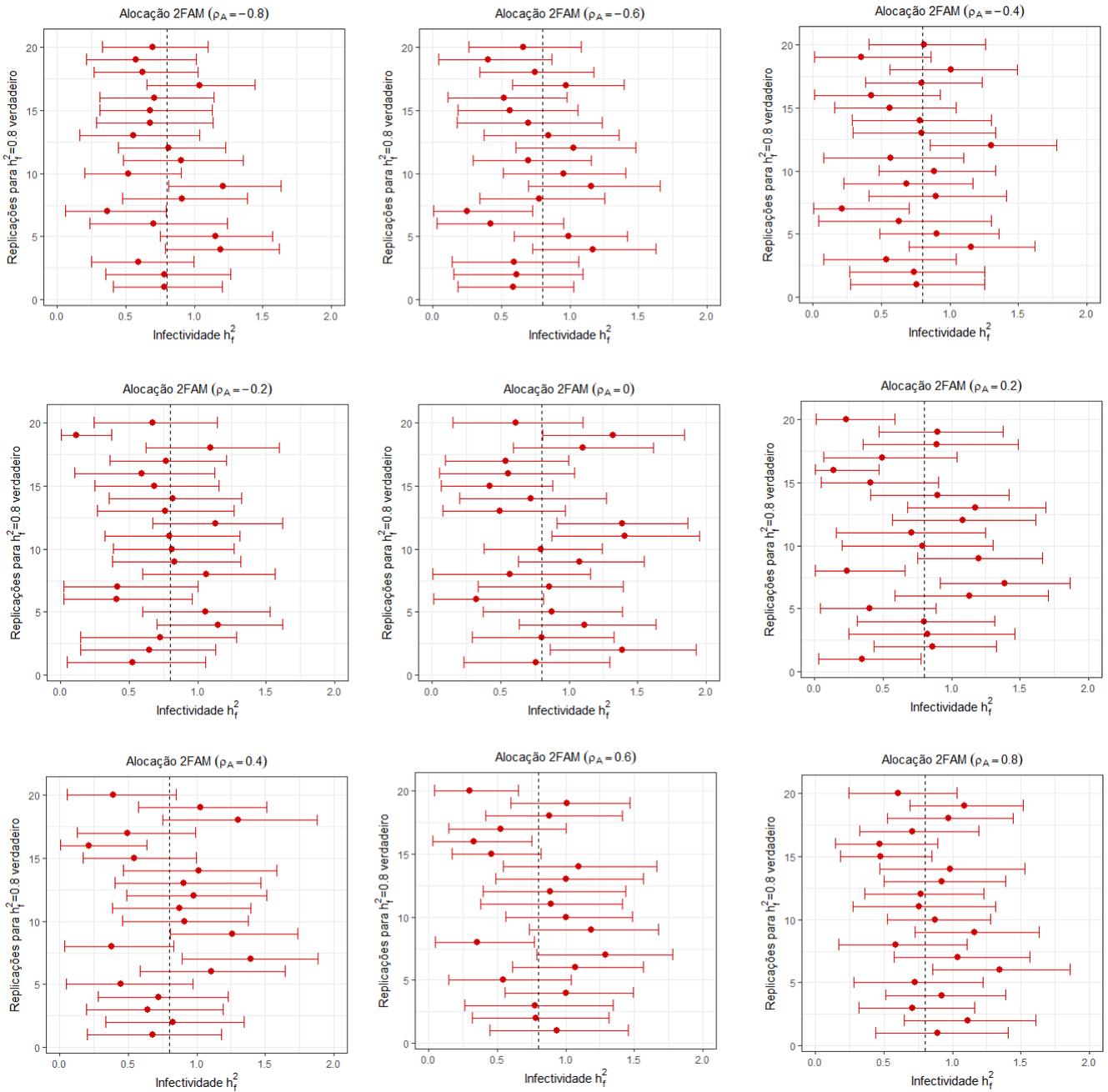


Figura 13 – Intervalos de credibilidade para a herdabilidade da infectividade quando o método de alocação é 2FAM e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.

Fonte: Elaborada pelo autor.

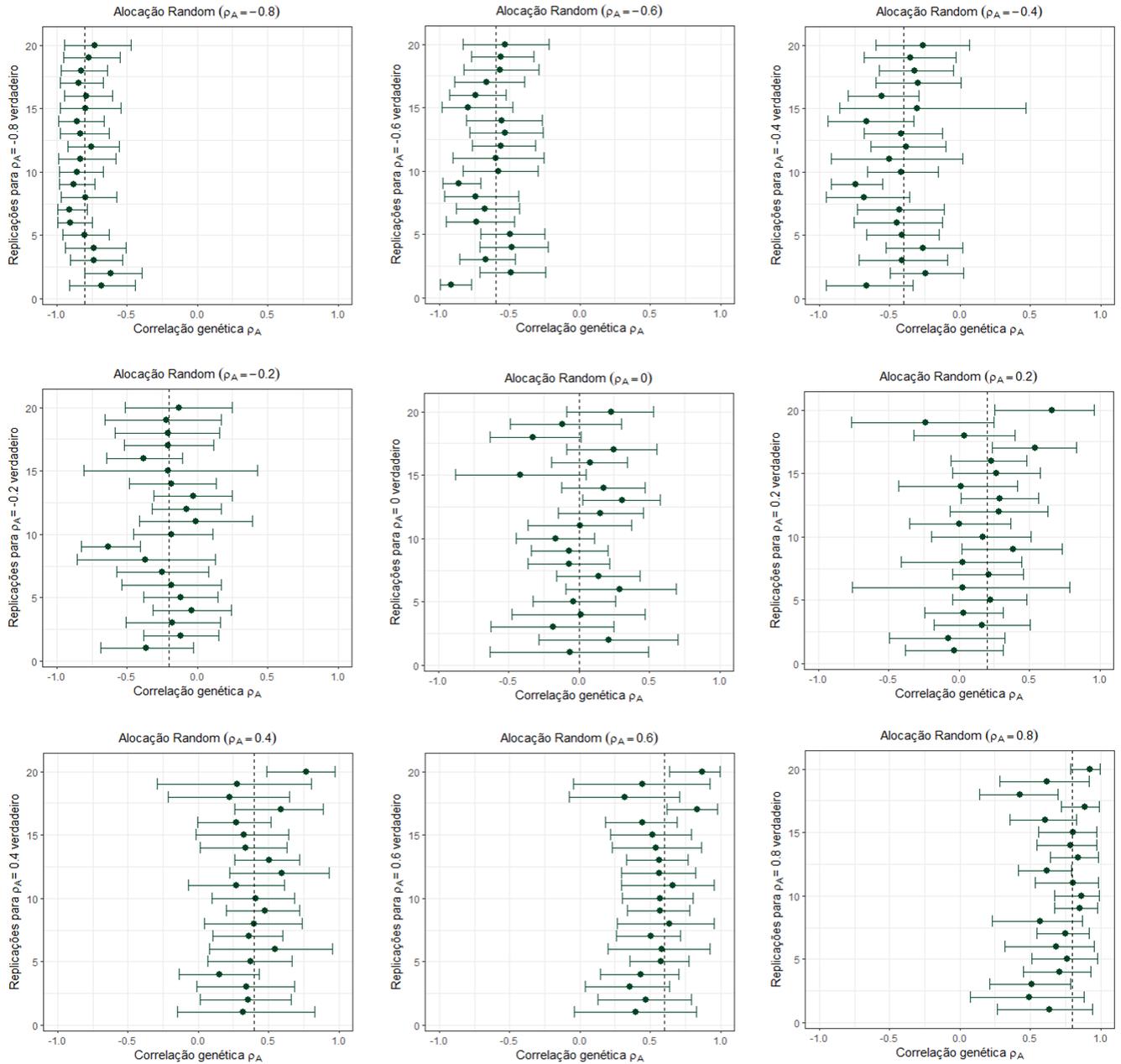


Figura 14 – Intervalos de credibilidade para o coeficiente de correlação genética quando o método de alocação é *random* e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.

Fonte: Elaborada pelo autor.

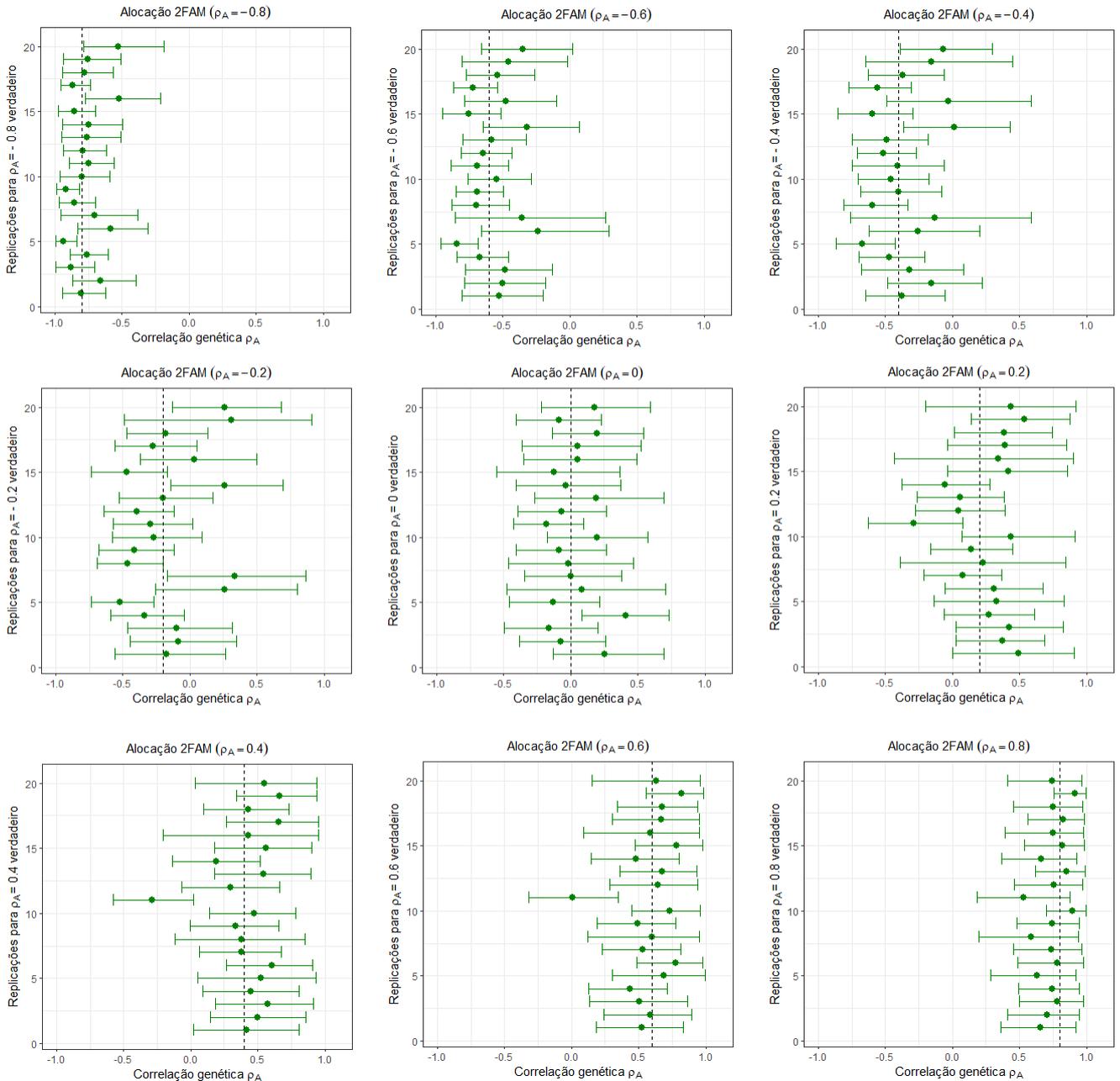


Figura 15 – Intervalos de credibilidade para o coeficiente de correlação genética quando o método de alocação é 2FAM e os valores de correlação (ρ_A) variam em $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, respectivamente.

Fonte: Elaborada pelo autor.

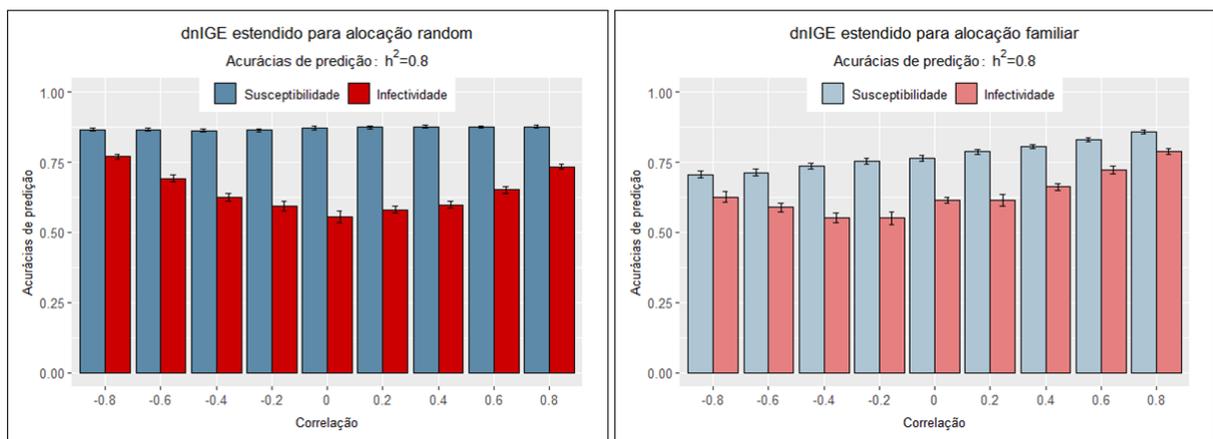
É importante ressaltar que, por definição, a herdabilidade das características assume valores que variam entre 0 e 1. No entanto, nas Figuras 6, 7, 10, 11, 12, 13 observamos que para alguns casos extremos existem intervalos de credibilidade que ultrapassam levemente esses limites. Uma possível explicação seria o fato de que as estimativas de herdabilidade não são obtidas de forma direta, mas calculadas pelas expressões $\hat{h}_g^2 = \frac{4\hat{\sigma}_{S,g}^2}{\hat{\sigma}_{S,g}^2 + \hat{\sigma}_{E,g}^2}$ e $\hat{h}_f^2 = \frac{4\hat{\sigma}_{S,f}^2}{\hat{\sigma}_{S,f}^2 + \hat{\sigma}_{E,f}^2}$, já que consideramos um modelo de risco paterno, em que seus parâmetros foram estimados a partir da

epidemia gerada na prole, assim como em [Anacleto et al. \(2015\)](#).

5.4.2 Efeito das correlações genéticas entre susceptibilidade e infectividade nas acurácias de predição

As Figuras 16 e 17 apresentam o efeito das correlações genéticas entre susceptibilidade e infectividade nas acurácias de predição do modelo *dnIGE* estendido. Vemos que, para o método de alocação *random* (Figura 16(a)), as acurácias de predição de susceptibilidade indicam que há sempre uma boa performance do modelo para predizer o efeito genético desta característica. Quando a força da correlação aumenta (tanto positiva quanto negativamente) há uma tendência de melhora nas acurácias de predição para a infectividade, tendo as correlações negativas uma leve vantagem sobre as correlações positivas. Por outro lado, para o método de alocação familiar (Figura 16(b)), as acurácias de predição de susceptibilidade seguem tendência positiva com respeito à direção da correlação. Quando a força da correlação aumenta positivamente, as acurácias de susceptibilidade aumentam. No entanto, quando a força da correlação entre susceptibilidade e infectividade aumenta negativamente, essas acurácias diminuem. Já as acurácias de predição da infectividade seguem o mesmo comportamento que as do método de alocação *random*, mas são melhores que as acurácias do método *random* quando as correlações são positivas ou zero.

Dessa maneira, a alocação *random* no modelo *dnIGE* estendido também apresenta melhor desempenho em predizer valores de susceptibilidade, ao passo que a alocação *2FAM* apresenta melhor desempenho em predizer valores de infectividade quando a correlação entre essas duas características é nula ou positiva. Quando a correlação entre essas características é negativa, a alocação *random* se mostra mais eficiente em predizer valores de infectividade.



(a) Acurácias de predição de susceptibilidade e infectividade quando o método de alocação é *random*.

(b) Acurácias de predição de susceptibilidade e infectividade quando o método de alocação é *2FAM*.

Figura 16 – Efeito das correlações genéticas nas acurácias de predição do modelo *dnIGE* estendido.

Fonte: Elaborada pelo autor.

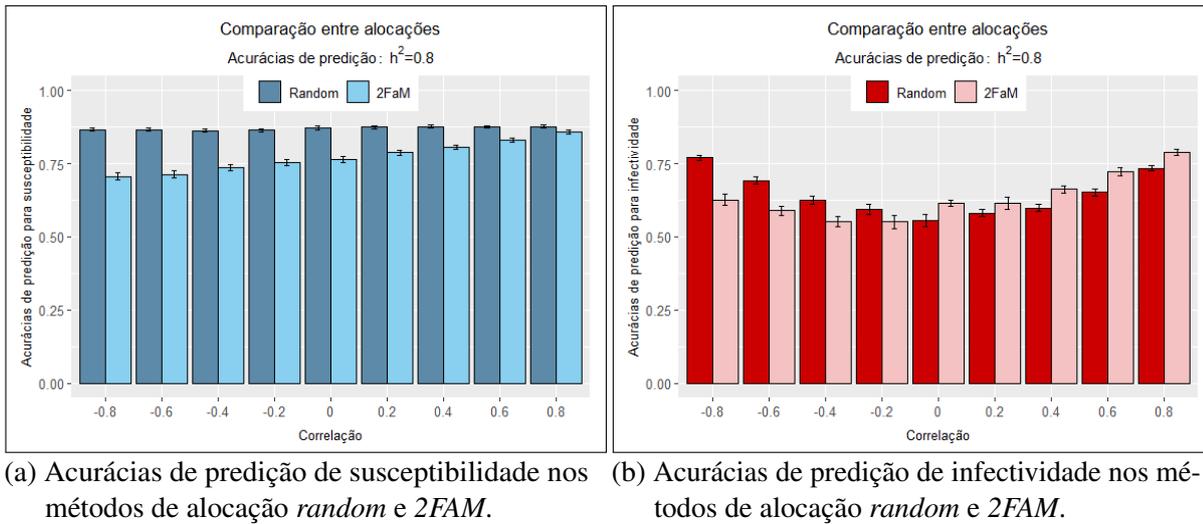


Figura 17 – Comparação entre os métodos de alocação em diversas correlações genéticas.

Fonte: Elaborada pelo autor.

É importante ressaltar que a obtenção de dados reais que permitam a estimação de efeitos genéticos em infectividade requer a realização de experimentos em larga escala, como discutido, por exemplo, em [Anacleto et al. \(2019\)](#), o que representa um grande desafio atualmente. Desta forma, nos limitamos à aplicação do modelo somente em dados simulados.

CONSIDERAÇÕES FINAIS

O presente capítulo traz uma síntese dos principais resultados da pesquisa realizada e discute suas contribuições teóricas, aplicações gerenciais e possíveis implicações sociais. Ainda são apresentadas as sugestões para investigações futuras, assim como potenciais aplicações.

6.1 Síntese dos principais resultados

Os resultados mostraram que a boa performance do modelo *dnIGE* estendido para a estimação de efeitos genéticos de infectividade está relacionada a cinco aspectos principais: tamanho da amostra, aumento da diversidade genética, redução do tamanho dos grupos, método de alocação e, finalmente, à presença de correlação genética entre as características epidemiológicas.

Tendo em vista que ambas as características são de grande importância em programas de melhoramento genético, esta pesquisa sugere que o método de alocação ideal a ser utilizado na prática é o *random*, pois além de ter fornecido boas estimativas para a infectividade, ainda refletiu sempre uma boa performance do modelo para prever os valores genéticos da susceptibilidade, enquanto que no método *2FAM* as acurácias de predição para esta característica são reduzidas quando a força da correlação diminui positivamente e quando aumenta negativamente.

6.2 Contribuições teóricas

Embora há muito tempo tenha sido demonstrado que na presença de características expressas por interações sociais, a resposta à seleção depende da covariância entre os efeitos genéticos diretos e indiretos (GRIFFING, 1967) - os quais correspondem à susceptibilidade e infectividade nesta tese - as pesquisas publicadas até o momento não investigaram as perspectivas do modelo *dnIGE* (ANACLETO *et al.*, 2015) sugerindo que estas características epidemiológicas estejam geneticamente correlacionadas. Sendo assim, apresentamos aqui esta contribuição

acadêmica, que na busca por comprovação empírica do modelo teórico proposto por [Bijma, Muir e Arendonk \(2007\)](#) atesta que a acurácia preditiva da infectividade é melhorada quando os parâmetros genéticos são fortemente correlacionados.

Além disso, a análise dos desenhos experimentais mostrou que quando susceptibilidade e infectividade são correlacionadas positivamente, o método *2FAM* apresenta as melhores acurácias de predição para a infectividade; e quando as correlações são negativas, o método *random* se torna o mais eficiente em prever os efeitos genéticos de susceptibilidade e infectividade.

6.3 Contribuições para o melhoramento genético de características epidemiológicas

Tendo em vista que o modelo estatístico aqui apresentado permite a identificação, via arquitetura genética, de animais altamente resistentes e altamente infecciosos em uma epidemia, esta tese contribui para a otimização de programas de melhoramento genético de características epidemiológicas, permitindo o bem-estar animal. Com isso, nossa metodologia oferece impactos potenciais para o controle de doenças infecciosas em pecuária, segurança alimentar global e também para a previsão e controle do surgimento de surtos futuros de epidemias. Produtores rurais e empresas que comercializam alimentos de origem animal podem ser beneficiados pela redução de custos decorrentes de surtos epidêmicos. A sociedade, por sua vez, pode ser beneficiada ao consumir esses alimentos sem risco de contaminação e por um custo reduzido.

6.4 Recomendações para investigações futuras

No cenário em que os atuais estudos genético-quantitativos de doenças infecciosas se concentram somente na susceptibilidade do hospedeiro, verificamos o efeito das variações genéticas de infectividade nas acurácias de predição (dos efeitos genéticos aditivos paternos de susceptibilidade e infectividade) do modelo *dnIGE* na versão original em comparação com a abordagem alternativa, que ignora a variação genética na infectividade. Realizamos esse teste considerando os métodos de alocação *random* e *2FAM*. Sendo assim, uma sugestão imediata consiste em fazer esta mesma verificação para ambos os métodos de alocação, considerando que as características epidemiológicas estão correlacionadas.

Outra sugestão consiste em aplicar o modelo nos casos em que os tempos de infecção são desconhecidos. Se é sabido que a infecção ocorreu num determinado intervalo de tempo $[\tau_i, \tau_j]$ é possível atribuir uma distribuição proposta uniforme entre τ_i e τ_j para estimar em que momento do intervalo cada animal é infectado. Desta forma, os parâmetros genéticos do modelo podem ser estimados usando o *status* da doença, registrado em cada um dos tempos da amostragem ([O'NEILL; ROBERTS, 1999](#)). Além disso, pode ser interessante verificar a relação da função de verossimilhança com os tempos entre as infecções.

Assim como na versão original do modelo *dnIGE*, nós nos concentramos em modelos compartimentais *SI* epidemiológicos. No entanto, dado que a força individual de infecção na Equação (4.1) é baseada em um processo de Poisson para o número de infecções que cada animal pode adquirir, temos que o modelo *dnIGE* estendido se enquadra na classe de modelos de eventos recorrentes, utilizados em análise de sobrevivência (KALBFLEISCH; PRENTICE, 2011). Isto sugere que o modelo pode ser aplicado a doenças que permitem a recuperação ou infecções recorrentes, mediante sua extensão para um modelo *SIR*. Consequentemente, isto pode permitir a estimação de R_0 e a aplicação em dados de sobrevivência.

Finalizando, a metodologia desta tese pode ser estendida para fins de estudos de características complexas em populações humanas. De acordo com Anacleto *et al.* (2015), o mecanismo de disseminação deve ser incorporado ao modelo *dnIGE*. Isso pode ser feito, por exemplo, assumindo uma estrutura de comunidade (como a presença de domicílios) ou usando modelos de rede de contato (DANON *et al.*, 2011), que foram aplicados com sucesso na modelagem de doenças infecciosas.

REFERÊNCIAS

AALEN, O.; BORGAN, O.; GJESSING, H. **Survival and event history analysis: a process point of view**. New York: Springer Science & Business Media, 2008. Citado na página 43.

ANACLETO, O.; CABALEIRO, S.; VILLANUEVA, B.; SAURA, M.; HOUSTON, R. D.; WOOLLIAMS, J. A.; DOESCHL-WILSON, A. B. Genetic differences in host infectivity affect disease spread and survival in epidemics. **Scientific Reports**, Nature Publishing Group, v. 9, n. 1, p. 4924, 2019. Disponível em: <<https://doi.org/10.1038/s41598-019-40567-w>>. Citado nas páginas 23 e 72.

ANACLETO, O.; GARCIA-CORTÉS, L. A.; LIPSCHUTZ-POWELL, D.; WOOLLIAMS, J. A.; DOESCHL-WILSON, A. B. A novel statistical model to estimate host genetic effects affecting disease transmission. **Genetics**, Genetics Soc America, v. 201, n. 3, p. 871–884, 2015. Disponível em: <<https://doi.org/10.1534/genetics.115.179853>>. Citado nas páginas 23, 40, 42, 46, 54, 55, 60, 61, 62, 71, 73 e 75.

ANACLETO, O.; ZUANETTI, D. A.; LIMA, M. N. **dnIGEmodel: R package to simulate data and estimate parameters of the dynamic non-linear indirect genetic effects (dnIGE) model**. 2023. R package. Disponível em: <<https://github.com/osvaldoanacleto/dnIGEmodel>>. Citado na página 56.

ANCHE, M.; JONG, M. D.; BIJMA, P. On the definition and utilization of heritable variation among hosts in reproduction ratio R_0 for infectious diseases. **Heredity**, Nature Publishing Group, v. 113, n. 4, p. 364, 2014. Disponível em: <<https://doi.org/10.1038/hdy.2014.38>>. Citado na página 51.

ANCHE, M. T.; BIJMA, P.; JONG, M. C. D. Genetic analysis of infectious diseases: estimating gene effects for susceptibility and infectivity. **Genetics Selection Evolution**, BioMed Central, v. 47, n. 1, p. 85, 2015. Disponível em: <<https://doi.org/10.1186/s12711-015-0163-z>>. Citado na página 40.

ANDERSON, R. M.; MAY, R. M. Population biology of infectious diseases: Part i. **Nature**, Nature Publishing Group, v. 280, n. 5721, p. 361–367, 1979. Disponível em: <<https://doi.org/10.1038/280361a0>>. Citado na página 29.

BAILEY, N. T. *et al.* **The mathematical theory of infectious diseases and its applications**. Hafner Press: Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975. Citado na página 30.

BARRY, J. R. **Probabilidade: um curso em nível intermediário**. Rio de Janeiro, Brasil: IMPA, 2008. Citado na página 86.

BAUD, A.; MULLIGAN, M. K.; CASALE, F. P.; INGELS, J. F.; BOHL, C. J.; CALLEBERT, J.; LAUNAY, J.-M.; KROHN, J.; LEGARRA, A.; WILLIAMS, R. W. *et al.* Genetic variation in the social environment contributes to health and disease. **PLoS Genetics**, Public Library of Science San Francisco, CA USA, v. 13, n. 1, p. e1006498, 2017. Disponível em: <<https://doi.org/10.1371/journal.pgen.1006498>>. Citado na página 39.

- BECKER, N. G. **Analysis of infectious disease data**. Florida: CRC Press, 2017. Citado na página 40.
- BECKER, N. G.; BRITTON, T. Statistical studies of infectious disease incidence. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 61, n. 2, p. 287–307, 1999. Disponível em: <<https://doi.org/10.1111/1467-9868.00177>>. Citado na página 23.
- BERGSMA, R.; MATHUR, P.; KANIS, E.; VERSTEGEN, M.; KNOL, E.; ARENDONK, J. V. Genetic correlations between lactation performance and growing-finishing traits in pigs. **Journal of Animal Science**, Oxford University Press, v. 91, n. 8, p. 3601–3611, 2013. Disponível em: <<https://doi.org/10.2527/jas.2012-6200>>. Citado na página 39.
- BERRAI, I. E. **Contribution à l'analyse déterministe et stochastique du modèle épidémiologique SIR**. Tese (Doutorado) — Université Hassan II Casablanca, Marrocos, 2016. Citado na página 31.
- BIEMANS, F.; JONG, M. C. de; BIJMA, P. A model to estimate effects of snps on host susceptibility and infectivity for an endemic infectious disease. **Genetics Selection Evolution**, BioMed Central, v. 49, n. 1, p. 53, 2017. Disponível em: <<https://doi.org/10.1186/s12711-017-0327-0>>. Citado na página 40.
- BIJMA, P. Estimating indirect genetic effects: precision of estimates and optimum designs. **Genetics**, Genetics Soc America, v. 186, n. 3, p. 1013–1028, 2010. Disponível em: <<https://doi.org/10.1534/genetics.110.120493>>. Citado nas páginas 52, 54 e 61.
- BIJMA, P. The quantitative genetics of indirect genetic effects: a selective review of modelling issues. **Nature Publishing Group**, 2014. Disponível em: <<https://doi.org/10.1038/hdy.2013.15>>. Citado na página 40.
- BIJMA, P.; MUIR, W. M.; ARENDONK, J. A. V. Multilevel selection 1: quantitative genetics of inheritance and response to selection. **Genetics**, Genetics Soc America, v. 175, n. 1, p. 277–288, 2007. Disponível em: <<https://doi.org/10.1534/genetics.106.062711>>. Citado nas páginas 39, 45 e 74.
- BISHOP, S.; DOESCHL-WILSON, A. B.; WOOLLIAMS, J. A. Uses and implications of field disease data for livestock genomic and genetics studies. **Frontiers in Genetics**, Frontiers, v. 3, p. 114, 2012. Disponível em: <<https://doi.org/10.3389/fgene.2012.00114>>. Citado na página 28.
- BISHOP, S. C.; AXFORD, R. F.; NICHOLAS, F. W.; OWEN, J. B. **Breeding for disease resistance in farm animals**. London: CABI, 2010. Citado nas páginas 22, 27, 29 e 32.
- BISHOP, S. C.; WOOLLIAMS, J. A. Genomics and disease resistance studies in livestock. **Livestock Science**, Elsevier, v. 166, p. 190–198, 2014. Disponível em: <<https://doi.org/10.1016/j.livsci.2014.04.034>>. Citado nas páginas 27 e 28.
- BROOKS-POLLOCK, E.; JONG, M. D.; KEELING, M.; KLINKENBERG, D.; WOOD, J. Eight challenges in modelling infectious livestock diseases. **Epidemics**, Elsevier, v. 10, p. 1–5, 2015. Disponível em: <<https://doi.org/10.1016/j.epidem.2014.08.005>>. Citado na página 27.
- CAMERLINK, I.; URSINUS, W. W.; BARTELS, A. C.; BIJMA, P.; BOLHUIS, J. E. Indirect genetic effects for growth in pigs affect behaviour and weight around weaning. **Behavior Genetics**, Springer, v. 48, n. 5, p. 413–420, 2018. Disponível em: <<https://doi.org/10.1007/s10519-018-9911-5>>. Citado na página 39.

- CARNEIRO-JÚNIOR, J. Melhoria genética animal. **Embrapa Acre-Capítulo em livro técnico (INFOTECA-E)**, In: GONÇALVES, RC; OLIVEIRA, LC de (Ed.). Embrapa Acre: ciência e tecnologia, 2009. Disponível em: <<https://www.embrapa.br/busca-de-publicacoes/-/publicacao/661781/melhoramento-genetico-animado>>. Citado na página 26.
- CARPENTER, B.; GELMAN, A.; HOFFMAN, M. D.; LEE, D.; GOODRICH, B.; BETANCOURT, M.; BRUBAKER, M. A.; GUO, J.; LI, P.; RIDDELL, A. Stan: a probabilistic programming language. **Grantee Submission**, ERIC, v. 76, n. 1, p. 1–32, 2017. Citado na página 56.
- CHANG, C.-C.; LEE, D.-H.; LIN, H.-C.; HSU, S.-S. The potential economic impact of avian flu pandemic on taiwan. **No 381-2016-22318**, 2007. Disponível em: <<https://ageconsearch.umn.edu/record/9803/>>. Citado na página 21.
- CHARMANTIER, A.; GARANT, D.; KRUK, L. E. **Quantitative genetics in the wild**. New York: OUP Oxford, 2014. Citado na página 27.
- CHEN, S.-N.; LI, C.-L.; LIN, J.-S.; ZHAI, S.-L.; SUN, M.-F. Diverse african swine fever viruses in china. **New Microbes and New Infections**, Elsevier, v. 46, 2022. Disponível em: <<https://doi.org/10.1016/j.nmni.2022.100976>>. Citado na página 21.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Editora Blucher, 2006. Citado na página 48.
- DANON, L.; FORD, A. P.; HOUSE, T.; JEWELL, C. P.; KEELING, M. J.; ROBERTS, G. O.; ROSS, J. V.; VERNON, M. C. Networks and the epidemiology of infectious disease. **Interdisciplinary perspectives on infectious diseases**, Hindawi, v. 2011, 2011. Disponível em: <<https://doi.org/10.1155/2011/284909>>. Citado na página 75.
- DAY, T.; PARK, A.; MADRAS, N.; GUMEL, A.; WU, J. When is quarantine a useful control strategy for emerging infectious diseases? **American Journal of Epidemiology**, Oxford University Press, v. 163, n. 5, p. 479–485, 2006. Disponível em: <<https://doi.org/10.1093/aje/kwj056>>. Citado na página 29.
- DIEKMANN, O.; HEESTERBEEK, J. A. P.; METZ, J. A. On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. **Springer**, Springer, v. 28, n. 4, p. 365–382, 1990. Disponível em: <<https://doi.org/10.1007/BF00178324>>. Citado na página 37.
- DOESCHL-WILSON, A. B.; DAVIDSON, R.; CONINGTON, J.; ROUGHSEGE, T.; HUTCHINGS, M. R.; VILLANUEVA, B. Implications of host genetic variation on the risk and prevalence of infectious diseases transmitted through the environment. **Genetics**, Genetics Soc America, v. 188, n. 3, p. 683–693, 2011. Disponível em: <<https://doi.org/10.1534/genetics.110.125625>>. Citado na página 22.
- ELLEN, E. D.; VISSCHER, J.; ARENDONK, J. A. van; BIJMA, P. Survival of laying hens: genetic parameters for direct and associative effects in three purebred layer lines. **Poultry Science**, Elsevier, v. 87, n. 2, p. 233–239, 2008. Disponível em: <<https://doi.org/10.3382/ps.2007-00374>>. Citado na página 39.
- EMBRAPA. **Pestes suínas: Desafio é manter a sanidade do rebanho**. 2019. Disponível em: <<https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1121843/relatorio-de-atividades-2019>>. Acesso em: 20-01-2021. Citado na página 21.

- EUROPEAN COMMISSION. **FACTSHEET-Avian influenza**. 2006. Disponível em: <https://ec.europa.eu/health/ph_threats/com/docs/factsheet_influenza_en.pdf>. Acesso em: 19-01-2021. Citado na página 21.
- FALCONER, D.; MACKAY, T. F. **Introduction to quantitative genetics**. England: Pearson, 1996. Citado nas páginas 25 e 26.
- FOPPA, I. M. **A Historical Introduction to Mathematical Modeling of Infectious Diseases: Seminal Papers in Epidemiology**. London: Academic Press, 2016. Citado na página 30.
- GARCIA, D. C. C.; SÁ, C. V. G. C. d.; PIMENTEL, C. M. M.; MELO, C. B. d. Impactos do surto de febre aftosa de 2005 sobre as exportações de carne bovina brasileira. **Ciência Animal Brasileira / Brazilian Animal Science**, 2015. Disponível em: <<https://revistas.ufg.br/vet/article/view/26158>>. Citado na página 21.
- GELMAN, A. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). **Bayesian Analysis**, 2006. Disponível em: <<https://doi.org/10.1214/06-BA117A>>. Citado nas páginas 57 e 63.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; DUNSON, D. B.; VEHTARI, A.; RUBIN, D. B. **Bayesian data analysis**. NW: Chapman and Hall/CRC, 2013. Citado nas páginas 49 e 56.
- GIBSON, J.; BISHOP, S. Use of molecular markers to enhance resistance of livestock to disease: a global approach. **Revue Scientifique Et Technique-Office International Des Epizooties**, OIE OFFICE INTERNATIONAL DES EPIZOOTIES, v. 24, n. 1, p. 343, 2005. Citado na página 22.
- GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. **The journal of Physical Chemistry**, ACS Publications, v. 81, n. 25, p. 2340–2361, 1977. Disponível em: <<https://doi.org/10.1021/j100540a008>>. Citado na página 55.
- GRIFFING, B. Selection in reference to biological groups i. individual and group selection applied to populations of unordered groups. **Australian Journal of Biological Sciences**, CSIRO, v. 20, n. 1, p. 127–140, 1967. Disponível em: <<https://doi.org/10.1071/BI9670127>>. Citado nas páginas 39, 45 e 73.
- HAL, W.; KEYNES, M. Modelling epidemics. In: **S320 : Infectious disease**. Cambridge: The Open University, 2008. Citado na página 32.
- HEIDARITABAR, M.; BIJMA, P.; JANSS, L.; BORTOLUZZI, C.; NIELSEN, H. M.; MADSEN, P.; ASK, B.; CHRISTENSEN, O. F. Models with indirect genetic effects depending on group sizes: a simulation study assessing the precision of the estimates of the dilution parameter. **Genetics Selection Evolution**, BioMed Central, v. 51, n. 1, p. 1–10, 2019. Disponível em: <<https://doi.org/10.1186/s12711-019-0466-6>>. Citado na página 61.
- HENDERSON, H. V.; SEARLE, S. R. The vec-permutation matrix, the vec operator and kronecker products: A review. **Linear and Multilinear Algebra**, Taylor & Francis, v. 9, n. 4, p. 271–288, 1981. Disponível em: <<https://doi.org/10.1080/03081088108817379>>. Citado na página 47.
- HOFFMAN, M. D.; GELMAN, A. *et al.* The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. **J. Mach. Learn. Res.**, v. 15, n. 1, p. 1593–1623, 2014. Disponível em: <<https://jmlr.org/papers/volume15/hoffman14a/hoffman14a.pdf>>. Citado na página 56.

HOUSTON, R. D.; HALEY, C. S.; HAMILTON, A.; GUY, D. R.; TINCH, A. E.; TAGGART, J. B.; MCANDREW, B. J.; BISHOP, S. C. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in atlantic salmon (*salmo salar*). **Genetics**, Genetics Soc America, v. 178, n. 2, p. 1109–1115, 2008. Disponível em: <<https://doi.org/10.1534/genetics.107.082974>>. Citado na página 27.

HULST, A. D.; JONG, M. C. de; BIJMA, P. Why genetic selection to reduce the prevalence of infectious diseases is way more promising than currently believed. **Genetics**, Oxford University Press, v. 217, n. 4, p. iyab024, 2021. Disponível em: <<https://doi.org/10.1093/genetics/iyab024>>. Citado na página 22.

INSUA, D.; RUGGERI, F.; WIPER, M. **Bayesian analysis of stochastic process models**. Reino Unido: John Wiley & Sons, 2012. v. 978. Citado nas páginas 33 e 35.

KALBFLEISCH, J. D.; PRENTICE, R. L. **The statistical analysis of failure time data**. Alemanha: John Wiley & Sons, 2011. v. 360. Citado na página 75.

KEELING, M. J.; ROHANI, P. **Modeling infectious diseases in humans and animals**. Estados Unidos: Princeton University Press, 2011. Citado nas páginas 29, 30, 31, 32, 40, 43 e 55.

KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. **Proceedings of the Royal Society of London. Series A.**, The Royal Society London, v. 115, n. 772, p. 700–721, 1927. Disponível em: <<https://doi.org/10.1098/rspa.1927.0118>>. Citado na página 30.

KNIGHT-JONES, T.; RUSHTON, J. The economic impacts of foot and mouth disease—what are they, how big are they and where do they occur? **Preventive Veterinary Medicine**, Elsevier, v. 112, n. 3-4, p. 161–173, 2013. Disponível em: <<https://doi.org/10.1016/j.prevetmed.2013.07.013>>. Citado na página 21.

KORSGAARD, I. R.; MADSEN, P.; JENSEN, J. Bayesian inference in the semiparametric log normal frailty model using gibbs sampling. **Genetics Selection Evolution**, BioMed Central, v. 30, n. 3, p. 241, 1998. Disponível em: <<https://doi.org/10.1186/1297-9686-30-3-241>>. Citado nas páginas 15 e 60.

KULKARNI, V. G. **Modeling and analysis of stochastic systems**. Estados Unidos: Crc Press, 2016. Citado na página 35.

LEKONE, P. E.; FINKENSTÄDT, B. F. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. **Biometrics**, Wiley Online Library, v. 62, n. 4, p. 1170–1177, 2006. Disponível em: <<http://www.jstor.org/stable/4124539>>. Citado na página 29.

LEWANDOWSKI, D.; KUROWICKA, D.; JOE, H. Generating random correlation matrices based on vines and extended onion method. **Journal of multivariate analysis**, Elsevier, v. 100, n. 9, p. 1989–2001, 2009. Disponível em: <<https://doi.org/10.1016/j.jmva.2009.04.008>>. Citado na página 63.

LIPSCHUTZ-POWELL, D.; WOOLLIAMS, J. A.; BIJMA, P.; DOESCHL-WILSON, A. B. Indirect genetic effects and the spread of infectious disease: are we capturing the full heritable variation underlying disease prevalence? **PloS One**, Public Library of Science, v. 7, n. 6, p. e39551, 2012a. Disponível em: <<https://doi.org/10.1371/journal.pone.0039551>>. Citado nas páginas 28, 39 e 57.

- LIPSCHUTZ-POWELL, D.; WOOLLIAMS, J. A.; DOESCHL-WILSON, A. B. A unifying theory for genetic epidemiological analysis of binary disease data. **Genetics Selection Evolution**, BioMed Central, v. 46, n. 1, p. 15, 2014. Disponível em: <<https://doi.org/10.1186/1297-9686-46-15>>. Citado na página 23.
- LLOYD-SMITH, J. O.; SCHREIBER, S. J.; KOPP, P. E.; GETZ, W. M. Superspreading and the effect of individual variation on disease emergence. **Nature**, Nature Publishing Group, v. 438, n. 7066, p. 355, 2005. Disponível em: <<https://doi.org/10.1038/nature04153>>. Citado nas páginas 22, 23 e 29.
- LONDON, W.; YORKE, J. A. Recurrent outbreaks of measles, chickenpox and mumps: Ii. systematic differences in contact rates and stochastic effects. **American Journal of Epidemiology**, Oxford University Press, v. 98, n. 6, p. 469–482, 1973b. Disponível em: <<https://doi.org/10.1093/oxfordjournals.aje.a121576>>. Citado na página 29.
- LYNCH, M.; WALSH, B. *et al.* **Genetics and analysis of quantitative traits**. Estados Unidos: Sinauer Sunderland, MA, 1998. v. 1. Citado na página 25.
- MAPA | Ministério da Agricultura, Pecuária e Abastecimento . **Brasil Livre da Febre Aftosa**. 2018. Disponível em: <<https://www.gov.br/agricultura/pt-br/assuntos/saude-animal-e-vegetal/saude-animal/programas-de-saude-animal/febre-aftosa/brasil-livre-da-aftosa>>. Acesso em: 24-10-2020. Citado na página 21.
- MOEN, T.; BARANSKI, M.; SONESSON, A. K.; KJØGLUM, S. Confirmation and fine-mapping of a major qtl for resistance to infectious pancreatic necrosis in atlantic salmon (*salmo salar*): population-level associations between markers and trait. **BMC Genomics**, Springer, v. 10, n. 1, p. 368, 2009. Disponível em: <<https://doi.org/10.1186/1471-2164-10-368>>. Citado na página 27.
- MOGHADAS, S. M.; PIZZI, N. J.; WU, J.; YAN, P. Managing public health crises: the role of models in pandemic preparedness. **Influenza and other respiratory viruses**, Wiley Online Library, v. 3, n. 2, p. 75–79, 2009. Disponível em: <<https://doi.org/10.1111/j.1750-2659.2009.00081.x>>. Citado na página 29.
- MRODE, R. A. **Linear models for the prediction of animal breeding values**. Boston: Cabi, 2014. Citado nas páginas 46 e 56.
- NATH, M.; WOOLLIAMS, J.; BISHOP, S. Assessment of the dynamics of microparasite infections in genetically homogeneous and heterogeneous populations using a stochastic epidemic model. **Journal of Animal Science**, Oxford University Press, v. 86, n. 8, p. 1747–1757, 2008. Disponível em: <<https://doi.org/10.2527/jas.2007-0615>>. Citado na página 22.
- ØDEGÅRD, J.; BARANSKI, M.; GJERDE, B.; GJEDREM, T. Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. **Aquaculture Research**, Wiley Online Library, v. 42, p. 103–114, 2011. Disponível em: <<https://doi.org/10.1111/j.1365-2109.2010.02669.x>>. Citado na página 27.
- ØDEGÅRD, J.; OLESEN, I. Comparison of testing designs for genetic evaluation of social effects in aquaculture species. **Aquaculture**, Elsevier, v. 317, n. 1-4, p. 74–78, 2011. Disponível em: <<https://doi.org/10.1016/j.aquaculture.2011.04.016>>. Citado nas páginas 53, 54 e 61.

OIE | Organização Mundial para a Saúde Animal. **Foot and mouth disease (FMD)**. 2018. Disponível em: <<https://www.woah.org/en/disease/foot-and-mouth-disease/>>. Acesso em: 24-10-2020. Citado na página 21.

OLDENBROEK, K.; WAAIJ, L. van der. **Textbook animal breeding and genetics for BSc students**. Wageningen UR: Centre for Genetic Resources The Netherlands and Animal Breeding and Genomics Centre, 2015. Disponível em: <<https://library.wur.nl/WebQuery/wurpubs/524548>>. Citado na página 25.

O'NEILL, P. D.; ROBERTS, G. O. Bayesian inference for partially observed stochastic epidemics. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, Wiley Online Library, v. 162, n. 1, p. 121–129, 1999. Disponível em: <<https://doi.org/10.1111/1467-985X.00125>>. Citado nas páginas 48 e 74.

PAULINO, C. D. M.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. Citado na página 49.

PEETERS, K.; EPPINK, T. T.; ELLEN, E. D.; VISSCHER, J.; BIJMA, P. Indirect genetic effects for survival in domestic chickens (*Gallus gallus*) are magnified in crossbred genotypes and show a parent-of-origin effect. **Genetics**, Oxford Academic, v. 192, n. 2, p. 705–713, 2012. Disponível em: <<https://doi.org/10.1534/genetics.112.142554>>. Citado nas páginas 39 e 61.

POOLEY, C.; MARION, G.; BISHOP, S.; DOESCHL-WILSON, A. Optimal experimental designs for estimating genetic and non-genetic effects underlying infectious disease transmission. **Genetics Selection Evolution**, BioMed Central, v. 54, n. 1, p. 1–22, 2022. Disponível em: <<https://doi.org/10.1186/s12711-022-00747-1>>. Citado na página 51.

RAPHAKA, K.; SÁNCHEZ-MOLANO, E.; TSAIRIDOU, S.; ANACLETO, O.; GLASS, E. J.; WOOLLIAMS, J. A.; DOESCHL-WILSON, A.; BANOS, G. Impact of genetic selection for increased cattle resistance to bovine tuberculosis on disease transmission dynamics. **Frontiers in Veterinary Science**, Frontiers, v. 5, p. 237, 2018. Disponível em: <<https://doi.org/10.3389/fvets.2018.00237>>. Citado nas páginas 28 e 46.

ROSS, S. M.; KELLY, J. J.; SULLIVAN, R. J.; PERRY, W. J.; MERCER, D.; DAVIS, R. M.; WASHBURN, T. D.; SAGER, E. V.; BOYCE, J. B.; BRISTOW, V. L. **Stochastic processes**. New York: Wiley, 1996. v. 2. Citado na página 33.

SAE-LIM, P.; BIJMA, P. Comparison of designs for estimating genetic parameters and obtaining response to selection for social interaction traits in aquaculture. **Aquaculture**, Elsevier, v. 451, p. 330–339, 2016. Disponível em: <<https://doi.org/10.1016/j.aquaculture.2015.09.017>>. Citado nas páginas 51, 53 e 54.

STAN | Development Team. **Stan Modeling Language Users Guide and Reference Manual, versão 2.18.0**. 2018. Disponível em: <<https://mc-stan.org>>. Acesso em: 11-01-2021. Citado na página 49.

STAN | Development Team. **RStan: the R interface to Stan**. 2022. R package version 2.21.5. Disponível em: <<https://mc-stan.org/>>. Citado na página 56.

TSAIRIDOU, S.; ANACLETO, O.; WOOLLIAMS, J.; DOESCHL-WILSON, A. Enhancing genetic disease control by selecting for lower host infectivity and susceptibility. **Heredity**, Nature Publishing Group, v. 122, n. 6, p. 742–758, 2019. Disponível em: <<https://doi.org/10.1038/s41437-018-0176-9>>. Citado nas páginas 23, 28 e 46.

WHO | ORGANIZAÇÃO MUNDIAL DE SAÚDE. **Infectious diseases**. 2016. Disponível em: <<https://www.emro.who.int/health-topics/infectious-diseases/index.html>>. Acesso em: 10-05-2020. Citado na página 27.

YORKE, J. A.; LONDON, W. P. Recurrent outbreaks of measles, chickenpox and mumps: I. seasonal variation in contact rates. **American Journal of Epidemiology**, Oxford University Press, v. 98, n. 6, p. 453–468, 1973a. Disponível em: <<https://doi.org/10.1093/oxfordjournals.aje.a121575>>. Citado na página 29.

ZHAI, S.-L.; WEN-KANG, W.; MING-FEI, S.; LV, D.-H.; ZHI-HONG, X. African swine fever spread in china. **The Veterinary Record**, Blackwell Publishing Ltd., v. 184, n. 18, p. 559, 2019. Disponível em: <<https://doi.org/10.1080/22221751.2019.1590128>>. Citado na página 21.

DISTRIBUIÇÕES A POSTERIORI CONDICIONAIS COMPLETAS PARA INFERÊNCIA DO VETOR DE PARÂMETROS DO MODELO DNIGE ESTENDIDO

Para a estimação do vetor de parâmetros θ , precisamos obter amostras da distribuição a posteriori conjunta (4.9).

Devido a complexidade desta distribuição vamos utilizar técnicas baseadas nos métodos *MCMC* que dependem do cálculo das distribuições condicionais completas. Desta forma, nas seções seguintes apresentamos, respectivamente, os cálculos das distribuições condicionais completas para as variâncias genéticas ($\sigma_{A,g}^2, \sigma_{A,f}^2$), variâncias ambientais ($\sigma_{E,g}^2, \sigma_{E,f}^2$), coeficientes de correlação (ρ_A, ρ_E), efeitos genéticos ($a_{g,s(j)}, a_{f,s(k)}$), efeitos ambientais ($e_{g,j}, e_{f,j}$), e para a taxa de contato efetivo (β). Em particular, a contribuição de cada animal no cálculo das condicionais completas das variâncias e efeitos ambientais depende de seu tipo (caso índice/não índice) e de seu *status* (infectado/não infectado) no final da epidemia (ver Quadro 1).

Quadro 1 – Eventos possíveis e respectivas características expressas em uma epidemia.

Eventos			Características expressas na epidemia	
			Susceptibilidade	Infectividade
É caso índice ($h_j = 0$)			Não	Sim
Não é caso índice ($h_j = 1$)	Foi infectado durante a epidemia	Foi o último a ser infectado no grupo ($u_j = 0$)	Sim	Não
		Não foi o último a ser infectado no grupo ($u_j = 1$)	Sim	Sim
	Não foi infectado durante a epidemia		Sim	Não

Fonte: Elaborada pelo autor.

A.1 Condicionais completas das variâncias genéticas

Definindo o vetor dos efeitos genéticos aditivos $\mathbf{a} = [\mathbf{a}_g^\top, \mathbf{a}_f^\top]^\top$, a distribuição condicional completa da variância genética na susceptibilidade é dada por,

$$\begin{aligned} \mathbb{P}(\sigma_{S,g}^2 | \cdot) &\propto \mathbb{P}(\sigma_{S,g}^2) \times \mathbb{P}(\mathbf{a} | \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A) \\ &\propto \mathbb{P}(\sigma_{S,g}^2) \times \mathbb{P}(\mathbf{a}_g | \mathbf{a}_f, \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A), \end{aligned}$$

em que “ \cdot ” denota o vetor de parâmetros desconsiderando $\sigma_{S,g}^2$.

A partir da suposição de que os pais não são relacionados geneticamente, temos que

$$\mathbb{P}(\sigma_{S,g}^2 | \cdot) \propto \left(\frac{1}{\sigma_{S,g}^2} \right) \times \prod_{i=1}^S \mathbb{P}(a_{g,i} | a_{f,i}, \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A),$$

em que S é o conjunto dos indivíduos (pais) que expressam susceptibilidade. Logo, temos que

$$a_{g,i}, a_{f,i} | \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{S,g}^2 & \rho_A \sigma_{S,g} \sigma_{S,f} \\ \rho_A \sigma_{S,g} \sigma_{S,f} & \sigma_{S,f}^2 \end{bmatrix} \right).$$

Logo, por propriedades da distribuição condicional da normal bivariada (BARRY, 2008),

$$a_{g,i} | a_{f,i}, \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A \sim N \left(\frac{\rho_A \sigma_{S,g}}{\sigma_{S,f}} a_{f,i}, \sigma_{S,g}^2 (1 - \rho_A^2) \right), \quad i = 1, \dots, S. \quad (\text{A.1})$$

Logo, mantendo os termos que dependem de $\sigma_{S,g}^2$ em (A.1), temos que,

$$\mathbb{P}(\sigma_{S,g}^2 | \cdot) \propto \left(\frac{1}{\sigma_{S,g}^2} \right)^{S/2+1} \exp \left\{ -\frac{1}{2\sigma_{S,g}^2 (1 - \rho_A^2)} \sum_{i=1}^S \left(a_{g,i} - \frac{\rho_A \sigma_{S,g}}{\sigma_{S,f}} a_{f,i} \right)^2 \right\}. \quad (\text{A.2})$$

Analogamente, a distribuição condicional completa da variância genética na infectividade é dada por:

$$\mathbb{P}(\sigma_{S,f}^2 | \cdot) \propto \left(\frac{1}{\sigma_{S,f}^2} \right)^{S/2+1} \exp \left\{ -\frac{1}{2\sigma_{S,f}^2 (1 - \rho_A^2)} \sum_{i=1}^S \left(a_{f,i} - \frac{\rho_A \sigma_{S,f}}{\sigma_{S,g}} a_{g,i} \right)^2 \right\}. \quad (\text{A.3})$$

A.2 Condicionais completas das variâncias ambientais

Na condicional completa da variância ambiental na susceptibilidade consideramos os casos em que os animais não sejam casos índice ($h_j = 1$) e que podem ser infectados durante a epidemia. Ou seja, somente animais que expressam susceptibilidade contribuem para esta condicional completa. Logo,

$$\begin{aligned} \mathbb{P}(\sigma_{E,g}^2 | \cdot) &\propto \mathbb{P}(\sigma_{E,g}^2) \mathbb{P}(\mathbf{e} | \sigma_{E,g}^2, \sigma_{E,f}^2, \rho_E) \\ &= \frac{1}{\sigma_{E,g}^2} \prod_{\substack{j:h_j=1 \\ \tau_j > 0}} \mathbb{P}(e_{g,j} | e_{f,j}, \sigma_{E,g}^2, \sigma_{E,f}^2, \rho_E), \end{aligned} \quad (\text{A.4})$$

em que:

$$e_{g,j} \sim \begin{cases} N\left(\frac{\rho_E \sigma_{E,g}}{\sigma_{S,f}} \cdot e_{f,j}, \sigma_{E,g}^2 (1 - \rho_E^2)\right), & \text{se } j \text{ expressa infectividade} \\ N(0, \sigma_{E,g}^2), & \text{se } j \text{ não expressa infectividade.} \end{cases} \quad (\text{A.5})$$

Logo:

$$\mathbb{P}(\sigma_{E,g}^2 | \cdot) \propto \begin{cases} \left(\frac{1}{\sigma_{E,g}^2}\right) \exp\left\{-\frac{1}{2\sigma_{E,g}^2(1-\rho_E^2)} \sum_{\substack{j:h_j=1 \\ \tau_j > 0}} \left(e_{g,j} - \frac{\rho_E \sigma_{E,g}}{\sigma_{E,f}} e_{f,j}\right)^2\right\}, & \text{se } j \text{ expressa infectividade;} \\ \left(\frac{1}{\sigma_{E,g}^2}\right) \exp\left\{-\frac{1}{2\sigma_{E,g}^2} \sum_{\substack{j:h_j=1 \\ \tau_j > 0}} e_{g,j}^2\right\}, & \text{se } j \text{ não expressa infectividade.} \end{cases} \quad (\text{A.6})$$

Analogamente, na condicional completa da variância ambiental na infectividade consideramos os animais que expressam infectividade durante a epidemia. Logo, são os casos índice e os que podem ser infectados naturalmente no período de observação que contribuem para esta condicional completa. Portanto,

$$\begin{aligned} \mathbb{P}(\sigma_{E,f}^2 | \cdot) &\propto \mathbb{P}(\sigma_{E,f}^2) \mathbb{P}(\mathbf{e} | \sigma_{E,f}^2, \sigma_{E,g}^2, \rho_E) \\ &= \frac{1}{\sigma_{E,f}^2} \prod_{\substack{j:h_j=1 \\ \tau_j > 0}} \mathbb{P}(e_{f,j} | e_{g,j}, \sigma_{E,f}^2, \sigma_{E,g}^2, \rho_E), \end{aligned} \quad (\text{A.7})$$

em que:

$$e_{f,j} \sim \begin{cases} N\left(\frac{\rho_E \sigma_{E,f}}{\sigma_{E,g}} \cdot e_{g,j}, \sigma_{E,f}^2 (1 - \rho_E^2)\right) & \text{se } j \text{ expressa susceptibilidade;} \\ N(0, \sigma_{E,f}^2) & \text{se } j \text{ não expressa susceptibilidade.} \end{cases}$$

Consequentemente,

$$\mathbb{P}(\sigma_{E,f}^2|\cdot) \propto \begin{cases} \left(\frac{1}{\sigma_{E,f}^2} \right) \exp \left\{ -\frac{1}{2\sigma_{E,f}^2(1-\rho_E^2)} \sum_{\substack{j:h_j=1 \\ \tau_j>0}} \left(e_{f,j} - \frac{\rho_E \sigma_{E,f}}{\sigma_{E,g}} e_{g,j} \right)^2 \right\}, & \text{se } j \text{ expressa susceptibilidade;} \\ \left(\frac{1}{\sigma_{E,f}^2} \right) \exp \left\{ -\frac{1}{2\sigma_{E,f}^2} \sum_{\substack{j:h_j=1 \\ \tau_j>0}} e_{f,j}^2 \right\}, & \text{se } j \text{ n\~{a}o expressa susceptibilidade.} \end{cases} \quad (\text{A.8})$$

A.3 Condicionais completas dos efeitos genéticos

Para encontrarmos a distribuição condicional completa do efeito genético aditivo de susceptibilidade de cada pai i , a log-verossimilhança é avaliada apenas para os descendentes do animal i que não sejam casos de índice. A representação desses animais é dada na seguinte forma:

$$\ell_{g,i} = \{j : (h_j = 1) \cap (s(j) = i)\}.$$

Então a log-condicional completa de $a_{g,i}$, $i = 1, \dots, S$ é dada por:

$$\begin{aligned} \log(\mathbb{P}(\mathbf{a}_{g,i}|\cdot)) &\propto \log(\mathbb{L}(\boldsymbol{\theta}) \times \mathbb{P}(\mathbf{a}|\sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A)) \\ &\propto \log(\mathbb{L}(\boldsymbol{\theta})) + \log(\mathbb{P}(\mathbf{a}_g|\mathbf{a}_f, \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A)) \\ &\propto \log(\mathbb{L}(\boldsymbol{\theta})) + \log\left(\prod_{i=1}^S \mathbb{P}(a_{g,i}|a_{f,i}, \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A)\right) \\ &\propto \text{(por propriedades da distribuição Normal)} \end{aligned} \tag{A.9}$$

$$\begin{aligned} &\propto \log(\mathbb{L}(\boldsymbol{\theta})) + \log\left(\exp\left\{-\frac{1}{2\sigma_{S,g}^2(1-\rho_A^2)} \sum_{i=1}^S \left(a_{g,i} - \frac{\rho_A \sigma_{S,g}}{\sigma_{S,f}} a_{f,i}\right)^2\right\}\right) \\ &\propto \log(\mathbb{L}(\boldsymbol{\theta})) - \sum_{i=1}^S \frac{\left(a_{g,i} - \frac{\rho_A \sigma_{S,g}}{\sigma_{S,f}} a_{f,i}\right)^2}{2\sigma_{S,g}^2(1-\rho_A^2)} \\ &\propto \sum_{\substack{j:j \in \ell_{g,i} \\ \tau_j \leq T}} a_{g,s(j)} - \beta \sum_{\substack{j:j \in \ell_{g,i} \\ \tau_j \leq T}} \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} (\tau_j - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \\ &\quad - \beta \sum_{\substack{j:j \in \ell_{g,i} \\ \tau_j > T}} \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} (T - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \\ &\quad - \sum_{i=1}^S \frac{\left(a_{g,i} - \frac{\rho_A \sigma_{S,g}}{\sigma_{S,f}} a_{f,i}\right)^2}{2\sigma_{S,g}^2(1-\rho_A^2)}. \end{aligned} \tag{A.10}$$

Na condicional completa do efeito genético aditivo na infectividade de cada pai i , a log-verossimilhança é avaliada para cada indivíduo que tenha um companheiro de grupo que seja descendente de algum pai i infectado antes do animal j . Estes animais são representados pelo conjunto:

$$\ell_{f,i} = \{j : \tau_j > \min\{(\tau_k : p_k = p_j) \cap (s(k) = i)\}\}.$$

Então a log-condicional completa de $a_{f,i}$, $i = 1, \dots, S$ é dada por:

$$\begin{aligned}
\log(\mathbb{P}(\mathbf{a}_{f,i}|\cdot)) &\propto \log(\mathbb{L}(\boldsymbol{\theta}) \times \mathbb{P}(\mathbf{a}|\sigma_{S,f}^2, \sigma_{S,g}^2, \rho_A)) \\
&\propto \log(\mathbb{L}(\boldsymbol{\theta})) + \log(\mathbb{P}(\mathbf{a}_f|\mathbf{a}_g, \sigma_{S,f}^2, \sigma_{S,g}^2, \rho_A)) \\
&\propto \log(\mathbb{L}(\boldsymbol{\theta})) + \log\left(\prod_{i=1}^S \mathbb{P}(a_{f,i}|a_{g,i}, \sigma_{S,f}^2, \sigma_{S,g}^2, \rho_A)\right) \\
&\propto \log(\mathbb{L}(\boldsymbol{\theta})) + \log\left(\exp\left\{-\frac{1}{2\sigma_{S,f}^2(1-\rho_A^2)} \sum_{i=1}^S \left(a_{f,i} - \frac{\rho_A \sigma_{S,f}^2}{\sigma_{S,g}^2} a_{g,i}\right)^2\right\}\right) \\
&\propto \log(\mathbb{L}(\boldsymbol{\theta})) - \sum_{i=1}^S \frac{\left(a_{f,i} - \frac{\rho_A \sigma_{S,f}^2}{\sigma_{S,g}^2} a_{g,i}\right)^2}{2\sigma_{S,f}^2(1-\rho_A^2)} \\
&\propto \sum_{\substack{j:j \in \ell_{f,i} \\ \tau_j \leq T}} \log\left(\sum_{k:p_k=p_j} \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j)\right) \\
&\quad - \beta \sum_{\substack{j:j \in \ell_{f,i} \\ \tau_j \leq T}} \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} (\tau_j - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \\
&\quad - \beta \sum_{\substack{j:j \in \ell_{f,i} \\ \tau_j > T}} \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} (T - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \\
&\quad - \sum_{i=1}^S \frac{\left(a_{f,i} - \frac{\rho_A \sigma_{S,f}^2}{\sigma_{S,g}^2} a_{g,i}\right)^2}{2\sigma_{S,f}^2(1-\rho_A^2)}. \tag{A.11}
\end{aligned}$$

A.4 Condicionais completas dos efeitos ambientais

A partir da suposição de independência dos efeitos ambientais, a log-condicional completa do efeito ambiental de susceptibilidade de cada animal j que não seja caso de índice é

$$\begin{aligned}
\log(\mathbb{P}(e_{g,j}|\cdot)) &\propto \log(\mathbb{L}(\boldsymbol{\theta})) + \log(\mathbb{P}(e_{g,j}|\sigma_{E,g}^2, \sigma_{E,f}^2, \rho_E, e_{f,j})) \\
&\propto \left[\beta e_{g,j} - \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} (\tau_j - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \right] \delta_j \\
&\quad - \left[\beta \exp\{a_{g,s(j)} + e_{g,j}\} \sum_{k:p_k=p_j} (T - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_j) \right] (1 - \delta_j) \\
&\quad + \log(\mathbb{P}(e_{g,j}|\sigma_E^2))
\end{aligned}$$

em que

$$\delta_j = \begin{cases} 1, & \text{se o animal } j \text{ foi infectado durante o período de observação;} \\ 0, & \text{caso contrário.} \end{cases}$$

Como $\mathbb{P}(e_{g,j}|\sigma_E^2, e_{f,j})$ depende do animal ter expresso a infectividade (A.5), temos que,

$$\log(\mathbb{P}(e_{g,j}|\sigma_E^2)) = \begin{cases} \frac{\sum_{\substack{j:h_j=1 \\ \tau_j>0}} \left(e_{g,j} - \frac{\rho_E \sigma_{E,g}}{\sigma_{E,f}} e_{f,j} \right)}{2\sigma_{E,g}^2(1-\rho_E^2)}, & \text{se } j \text{ expressa infectividade;} \\ -\frac{1}{2\sigma_{E,g}^2} \sum_{\substack{j:h_j=1 \\ \tau_j>0}} e_{g,j}^2, & \text{se } j \text{ não expressa infectividade.} \end{cases}$$

Na posteriori condicional do efeito ambiental de infectividade de cada animal infectado j , a log-verossimilhança é avaliada para seus companheiros de grupo que foram infectados depois de j , uma vez que indivíduos somente podem expressar infectividade após serem infectados e, se ainda houver susceptíveis em seus grupos após a infecção. Assim a log-condicional completa do efeito ambiental de infectividade do animal j , $j = 1, \dots, I$ é

$$\begin{aligned} \log(\mathbb{P}(e_{f,j}|\cdot)) &\propto \log(\mathbb{L}(\boldsymbol{\theta})) + \log(\mathbb{P}(e_{f,j}|\sigma_{E,f}^2, \sigma_{E,g}^2, \rho_E, e_{g,j})) \\ &= \sum_{\substack{i:p_i=p_j \\ \tau_i \geq \tau_j}} \delta_i \log \left(\sum_{k:p_k=p_j} \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_i) \right) + \log(\mathbb{P}(e_{f,j}|\sigma_E^2)) \\ &\quad - \beta \sum_{\substack{i:p_i=p_j \\ \tau_i \geq \tau_j}} \exp\{a_{g,s(i)} + e_{g,i}\} \sum_{k:p_k=p_i} (\tau_i - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_i) \delta_i \quad (\text{A.12}) \\ &\quad - \beta \sum_{\substack{i:p_i=p_j \\ \tau_i > T}} \left\{ \exp\{a_{g,s(i)} + e_{g,i}\} \sum_{k:p_k=p_i} (T - \tau_k) \exp\{a_{f,s(k)} + e_{f,k}\} \ell_k(\tau_i) \right\} (1 - \delta_i) \end{aligned}$$

em que,

$$\log(\mathbb{P}(e_{f,j}|\sigma_E^2)) = \begin{cases} \frac{-\sum_{\substack{j:h_j=1 \\ \tau_j>0}} \left(e_{f,j} - \frac{\rho_E \sigma_{E,f}}{\sigma_{E,g}} e_{g,j} \right)}{2\sigma_{E,g}^2(1-\rho_E^2)}, & \text{se } j \text{ expressa susceptibilidade;} \\ -\frac{1}{2\sigma_{E,f}^2} \sum_{\substack{j:h_j=1 \\ \tau_j>0}} e_{f,j}^2, & \text{se } j \text{ expressa susceptibilidade.} \end{cases}$$

A.5 Condicionais completas dos coeficientes de correlação

A distribuição condicional completa de ρ_A é dada por,

$$\begin{aligned}
\mathbb{P}(\rho_A|\cdot) &\propto \mathbb{P}(\rho_A) \times \mathbb{P}(\mathbf{a}|\sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A) \\
&= \frac{1}{2} \prod_{i=1}^S \mathbb{P}(a_{g,i}, a_{f,i}|\sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A) \\
&= \frac{1}{2} \prod_{i=1}^S \mathbb{P}(a_{f,i}|\sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A) \times \mathbb{P}(a_{g,i}|a_{f,i}, \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A) \\
&= \frac{1}{2} \prod_{i=1}^S (2\pi\sigma_{S,f}^2)^{-1/2} \exp\left\{-\frac{a_{f,i}^2}{2\sigma_{S,f}^2}\right\} \times (2\pi\sigma_{S,g}^2(1-\rho_A^2))^{-1/2} \exp\left\{-\frac{1}{2\sigma_{S,g}^2(1-\rho_A^2)} \sum_{i=1}^S \left(a_{g,i} - \frac{\rho_A\sigma_{S,g}}{\sigma_{S,f}} a_{f,i}\right)^2\right\} \\
&\propto (1-\rho_A^2)^{-S/2} \exp\left\{-\frac{1}{2\sigma_{S,g}^2(1-\rho_A^2)} \sum_{i=1}^S \left(a_{g,i} - \frac{\rho_A\sigma_{S,g}}{\sigma_{S,f}} a_{f,i}\right)^2\right\}. \tag{A.13}
\end{aligned}$$

Na condicional completa da correlação ambiental entre susceptibilidade e infectividade são considerados os animais que expressam susceptibilidade e infectividade. Portanto

$$\begin{aligned}
\mathbb{P}(\rho_E|\cdot) &= \frac{1}{2} \prod \mathbb{P}(e_{g,j}, e_{f,j}|\Sigma_E) \\
&\propto \prod (1-\rho_E^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_{E,g}^2(1-\rho_E^2)} \left(e_{g,i} - \frac{\rho_E\sigma_{E,g}}{\sigma_{E,f}} e_{f,i}\right)^2\right\} \tag{A.14}
\end{aligned}$$

A.6 Condicional completa da taxa de contato efetivo β

Para o caso onde todos os animais são infectados (prevalência = 1) temos que

$$\begin{aligned}
\mathbb{P}(\beta|\cdot) &\propto \mathbb{P}(\beta)\mathbb{L}(\boldsymbol{\theta}) \tag{A.15} \\
&\propto \beta^{a-1} \exp\{-b\beta\} \prod_{j:h_j=1} \beta \exp\{g_j\beta \sum (\tau_j - \tau_k) f_k \ell_k(\tau_j)\}.
\end{aligned}$$

Logo,

$$\beta|\cdot \sim \text{Gama}\left(a+I-n_0, \quad b + \sum_{j:h_j=1} g_j \sum_{k:p_k=p_j} (\tau_j - \tau_k) f_k \ell_k(\tau_j)\right).$$

Caso a prevalência seja menor do que 1, temos que,

$$\begin{aligned}
\mathbb{P}(\beta|\cdot) &\propto \beta^{a-1} \exp\{-b\beta\} \underbrace{\prod_{j:h_j=0} \beta \exp\{g_j\beta \sum (\tau_j - \tau_k) f_k \ell_k(\tau_j)\}}_{\text{Infectado}} \underbrace{\prod_{j:h_j=1} \beta \exp\{g_j\beta \sum (T - \tau_k) f_k \ell_k(t)\}}_{\text{N\~{a}o infectado}} \\
\beta|\cdot &\sim \text{Gama}\left(a+I-n_0, \quad b + \left[g_j \sum_{k:p_k=p_j} (\tau_j - \tau_k) f_k \ell_k(\tau_j) \right] + \left[g_j \sum_{k:p_k=p_j} (T - \tau_k) f_k \ell_k(t) \right] \right)
\end{aligned}$$

