

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP

GEORGE LUCAS MORAES PEZZOTT

**MODELOS ESPACIAIS DE CAPTURA-RECAPTURA PARA
POPULAÇÕES ABERTAS**

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

Orientador: Prof. Dr. Luis Ernesto Bueno Salasar

**São Carlos
Janeiro de 2019**

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP

GEORGE LUCAS MORAES PEZZOTT

SPATIAL CAPTURE-RECAPTURE MODELS FOR OPEN POPULATIONS

Doctoral dissertation submitted to the Des/UFSCar and to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the PhD degree Statistics – Interagency Program Graduate in Statistics UFSCar-USP.

Advisor: Prof. Dr. Luis Ernesto Bueno Salasar

São Carlos
January 2019



UNIVERSIDADE FEDERAL DE SÃO CARLOS

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

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato George Lucas Moraes Pezzott, realizada em 22/11/2018:

Prof. Dr. Luis Ernesto Bueno Salazar
UFSCar

Profa. Dra. Daiane Aparecida Zuanetti
UFSCar

Profa. Dra. Maria Elina Bichuette
UFSCar

Prof. Dr. Victor Fossaluzza
USP

Prof. Dr. Marcos Oliveira Prates
UFMG

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Marcos Oliveira Prates e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. Luis Ernesto Bueno Salazar

AGRADECIMENTOS

Primeiramente, agradeço a Deus por estar sempre comigo, por ter iluminado minha trajetória me dando sabedoria e saúde para concluir este trabalho.

À minha família, principalmente meus pais Lauro e Edna e meus irmãos Roger e Julio, por todo o amor que me deram em toda minha vida e, sobretudo, pelo incentivo para seguir com meus estudos.

À minha esposa, Vanessa, que amo tanto, pelo companheirismo, apoio e amizade durante todos estes anos, pelos momentos de alegria compartilhados e pelas palavras de conforto nos momentos difíceis que, sem dúvida, foram essenciais nesta jornada.

Ao meu orientador professor Luis Ernesto Bueno Salasar pela orientação, paciência, comprometimento e confiança depositada em mim na elaboração deste trabalho e, em especial, pela amizade que construímos durante minha pós-graduação.

Ao professor José Galvão Leite pelas ideias, correções e sugestões feitas neste trabalho e pelo privilégio de trabalhar com um doutor tão prestigiado na área acadêmica, tornando-o uma das minhas inspirações para seguir profissão nesta área.

Ao Bruno Gabriel Oliveira do Monte por disponibilizar os dados de sua tese de doutorado para motivação e aplicação da minha pesquisa além de toda colaboração, junto com a professora Maria Elina Bichuette, para melhorar meu entendimento sobre o assunto.

Aos professores, colegas e funcionários do PIPGES por todo conhecimento compartilhado, apoio e pelo ótimo ambiente de pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e, por isto, sou grato a esta fundação.

RESUMO

PEZZOTT, G. L. M. **Modelos espaciais de captura-recaptura para populações abertas.** 2019. 129 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, 2019.

Nesta tese propomos dois modelos espaciais de captura-recaptura para estimação da abundância populacional em população aberta. Os modelos estatísticos propostos ajustam-se a dados obtidos via amostragem de captura-recaptura com marcação individual realizada em diferentes locais dentro do habitat, levando em consideração as taxas de nascimentos e mortes durante o período de estudo e as localizações geográficas das capturas. No primeiro modelo, propomos uma modelagem hierárquica para os tamanhos populacionais locais a fim de obter a distribuição preditiva da abundância populacional para regiões não visitadas pela amostragem. Nesta etapa, uma estrutura para dados zero-inflacionados foi adotada para acomodar situações quando realizam-se amostragens em locais sem a presença da espécie. O segundo modelo proposto leva em consideração o deslocamento dos animais entre os diferentes locais de amostragem, generalizando o primeiro modelo no qual consideramos a permanência dos animais em um mesmo local. Neste caso, tornou-se possível estimar o tamanho da área de vida (movimentação) da espécie além de prever locais com maiores abundâncias de animais. Em ambos modelos, propomos uma abordagem bayesiana para o processo inferencial e derivamos algoritmos de simples implementação computacional, a partir do uso de técnicas de dados aumentados. As propriedades frequentistas dos estimadores bayesianos foram avaliadas por meio de estudos de simulação e, por fim, estas propostas de modelagem foram aplicadas a três conjuntos de dados reais de aracnídeos.

Palavras-chave: Modelo espacial de captura-recaptura, População aberta, Análise bayesiana.

ABSTRACT

PEZZOTT, G. L. M. **Spatial capture-recapture models for open populations**. 2019. 129 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, 2019.

In this thesis we propose two spatial capture-recapture models for estimation of population abundance in the open population. The proposed statistical models conform to data obtained through individual tag capture-recapture sampling performed in different areas within the habitat, taking into account the rates of births and deaths during the study period and the geographical locations of the catches. In the first model, we propose a hierarchical modeling for local population sizes in order to obtain the predictive distribution of population abundance for regions not visited by sampling. In this step, a structure for zero-inflated data was adopted to accommodate situations when sampling is performed in areas without the presence of the species. The second proposed model takes into account the movement of the animals among the different sampling areas, generalizing the first model in which we consider the permanence of the animals in the same area. In this case, it became possible to estimate the size of the area of movement of the species and to predict areas with higher abundances of animals. In both models, we propose a Bayesian approach to the inferential process and derive algorithms from simple computational implementation, from the use of augmented data techniques. The frequentist properties of the Bayesian estimators were evaluated by simulation studies and, finally, these modeling proposals were applied to three real data sets of arachnids.

Keywords: Spatial capture-recapture model, Open population, Bayesian analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 – Mapa em planta baixa da região de estudo. Fonte: Monte (2017).	20
Figura 2 – Locais de marcação: (a) dezenas, (b) unidades, (c) centenas e (d) ocasião da marcação. Fonte: Monte (2017).	21
Figura 3 – Representação gráfica das probabilidades condicionais.	28
Figura 4 – Aproximações das funções de probabilidades a posteriori para os tamanhos populacionais em cada ocasião de amostragem. Linhas verticais azuis indicam o verdadeiro valor do parâmetro.	49
Figura 5 – Valores médios estimados a posteriori para os tamanhos populacionais em cada ocasião e ‘zona’ de amostragem.	50
Figura 6 – Aproximações das funções de probabilidades predita do tamanho da superpopulação em uma região não amostrada.	53
Figura 7 – Viés Relativo (%) do número estimado de animais vivos em cada ocasião de captura.	58
Figura 8 – EAM do número estimado de animais vivos em cada ocasião de captura. . .	59
Figura 9 – Número total de animais distintos capturados em cada região da espécie <i>Iandumoema uai</i>	65
Figura 10 – Abundâncias populacionais estimadas por ocasião.	69
Figura 11 – Abundâncias populacionais estimadas por região.	69
Figura 12 – Densidade média estimada de machos para cada fêmea por região.	70
Figura 13 – Abundâncias populacionais estimadas por região e ocasião.	71
Figura 14 – Distribuição utilizada no estudo para geração dos locais de centros de vida. .	88
Figura 15 – Aproximações das funções de probabilidades a posteriori para os tamanhos populacionais em cada ocasião de amostragem. Linhas verticais azuis indicam o verdadeiro valor do parâmetro.	90
Figura 16 – Valores médios estimados a posteriori para os tamanhos populacionais em cada ocasião e região de amostragem para uma amostra gerada sob o cenário 11.	91
Figura 17 – Valores médios estimados a posteriori para os tamanhos populacionais em cada ocasião e região de amostragem para uma amostra gerada o cenário 17. .	92
Figura 18 – Distribuição a posteriori marginal aproximada do parâmetro N em uma amostra gerada sob o cenário 1.	93
Figura 19 – Viés Relativo (%) e EAM do número estimado de animais vivos em cada ocasião de captura.	97

Figura 20 – Viés Relativo (%) dos estimadores bayesianos pontuais para o vetor de parâmetros π	101
Figura 21 – EAM dos estimadores bayesianos pontuais para o vetor de parâmetros π	102
Figura 22 – Número de animais capturados em cada região de amostragem.	105
Figura 23 – Abundâncias populacionais estimadas por ocasião.	109
Figura 24 – Região de movimentos estimada para um animal com centro de vida localizado em 250 m.	111
Figura 25 – Distribuição espacial estimada dos centros de vida.	111
Figura 26 – Distribuição espaço-temporal estimada dos animais na caverna ao longo do tempo.	112
Figura 27 – Valores gerados para os parâmetros na primeira aplicação real.	126
Figura 28 – Valores gerados para os parâmetros na primeira aplicação real.	127
Figura 29 – Valores gerados para os parâmetros na segunda aplicação real.	128
Figura 30 – Valores gerados para os parâmetros na segunda aplicação real.	129

LISTA DE TABELAS

Tabela 1 – Cenários utilizados para o primeiro estudo de simulação.	47
Tabela 2 – Resumos a posteriori para os tamanhos populacionais em cada ocasião de amostragem para uma amostra gerada do Cenário 8.	49
Tabela 3 – Resumos a posteriori para os hiperparâmetros (μ, σ) para uma amostra gerada do Cenário 8.	51
Tabela 4 – Resumos a posteriori para os hiperparâmetros (μ, σ) para uma amostra gerada do Cenário 12.	51
Tabela 5 – Resumos descritivos dos valores preditos para as abundâncias \tilde{N}_t	53
Tabela 6 – Medidas de desempenho dos estimadores bayesianos para o parâmetro N	54
Tabela 7 – Medidas de desempenho dos estimadores bayesianos para o parâmetro θ	55
Tabela 8 – Medidas de desempenho dos estimadores bayesianos para o vetor de parâmetro β	56
Tabela 9 – Medidas de desempenho dos estimadores bayesianos para o vetor de parâmetro ϕ	57
Tabela 10 – Medidas de desempenho do estimador intervalar bayesiano para os parâmetros N_t	60
Tabela 11 – Medidas de desempenho dos estimadores bayesianos para o parâmetro μ	61
Tabela 12 – Medidas de desempenho dos estimadores bayesianos para o parâmetro σ	61
Tabela 13 – Medidas de desempenho dos estimadores bayesianos para o parâmetro ω	62
Tabela 14 – Frequências de capturas de machos e fêmeas da espécie <i>Iandumoema uai</i>	64
Tabela 15 – Valores do Critério DIC para comparação de modelos.	67
Tabela 16 – Resumos a posteriori para os tamanhos da superpopulação para a espécie <i>Iandumoema uai</i>	67
Tabela 17 – Resumos a posteriori para os parâmetros de nascimento, sobrevivência e captura.	68
Tabela 18 – Resumos a posteriori para as abundâncias populacionais em cada ocasião de captura.	68
Tabela 19 – Resumos a posteriori para os hiperparâmetros da priori hierárquica da abundância populacional.	71
Tabela 20 – Resumos descritivos dos valores preditos para as abundâncias \tilde{N}_t para a espécie <i>Iandumoema uai</i>	72
Tabela 21 – Cenários utilizados no segundo estudo de simulação.	87

Tabela 22 – Resumos a posteriori para os tamanhos populacionais em cada ocasião de amostragem para uma amostra gerada do Cenário 11.	90
Tabela 23 – Medidas de desempenho dos estimadores bayesianos do parâmetro N	92
Tabela 24 – Medidas de desempenho dos estimadores bayesianos para o parâmetro θ	94
Tabela 25 – Medidas de desempenho dos estimadores bayesianos do vetor de parâmetro β	95
Tabela 26 – Medidas de desempenho dos estimadores bayesianos do vetor de parâmetro ϕ	96
Tabela 27 – Medidas de desempenho do estimador intervalar bayesiano para os parâmetros N_t	98
Tabela 28 – Medidas de desempenho dos estimadores bayesianos do parâmetro σ	99
Tabela 29 – Comparação das medidas de desempenho dos estimadores bayesianos para o parâmetro σ em diferentes partições da região de amostragem.	100
Tabela 30 – Frequências de capturas para as espécies <i>Eusarcus cavernicola</i> e <i>Eusarcus aduncus</i>	104
Tabela 31 – Totais de recapturas entre diferentes regiões.	104
Tabela 32 – Valores do Critério DIC para comparação dos modelos.	108
Tabela 33 – Resumos a posteriori para as abundâncias populacionais.	108
Tabela 34 – Resumos a posteriori para as abundâncias populacionais em cada ocasião de captura.	109
Tabela 35 – Resumos a posteriori para os parâmetros de nascimento, sobrevivência, captura e movimentação	110

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Motivação e objetivos da tese	20
2	MODELO DE CR PARA POPULAÇÃO ABERTA	25
2.1	Modelo estatístico de captura-recaptura	25
2.1.1	<i>Processo de nascimento e morte</i>	26
2.1.2	<i>Processo de captura</i>	29
2.1.3	<i>Função de verossimilhança</i>	31
3	MODELO DE CR HIERÁRQUICO PARA POPULAÇÃO ABERTA	33
3.1	Modelo estatístico	33
3.2	Abordagem Bayesiana	35
3.2.1	<i>Aumento de dados</i>	38
3.2.2	<i>Algoritmo MCMC</i>	41
3.3	Estudo de simulação	45
3.3.1	<i>Medidas de desempenho</i>	46
3.3.2	<i>Geração dos dados</i>	46
3.3.3	<i>Resultados numéricos</i>	48
3.3.4	<i>Exemplo</i>	48
3.3.5	<i>Qualidade das estimativas</i>	53
4	APLICAÇÃO: ESPÉCIE <i>IANDUMOEMA UAI</i>	63
4.1	Análise descritiva dos dados	63
4.2	Análise Inferencial	64
5	MODELO DE CRE PARA POPULAÇÃO ABERTA	73
5.1	Introdução	73
5.2	Modelo estatístico de captura-recaptura	76
5.2.1	<i>Distribuição espacial e movimentação dos animais</i>	77
5.2.2	<i>Processo de captura</i>	79
5.2.3	<i>Função de verossimilhança</i>	81
5.3	Abordagem Bayesiana	81
5.3.1	<i>Função de verossimilhança aumentada</i>	82

5.3.2	<i>Algoritmo MCMC</i>	84
5.4	Estudo de simulação	86
5.4.1	<i>Geração dos dados</i>	86
5.4.2	<i>Resultados numéricos</i>	89
5.4.3	<i>Exemplo</i>	89
5.4.4	<i>Qualidade das estimativas</i>	91
6	APLICAÇÃO: ESPÉCIES <i>EUSARCUS CAVERNICOLA</i> E <i>EUSARCUS ADUNCUS</i>	103
6.1	Análise descritiva dos dados	103
6.2	Análise Inferencial	106
7	CONCLUSÕES	113
	REFERÊNCIAS	115
ANEXO A	GRÁFICOS DE CONVERGÊNCIA	125

INTRODUÇÃO

Inúmeros estudos em ecologia visam o conhecimento acerca da dinâmica populacional de uma determinada espécie ([TOWNSEND *et al.*, 2003](#)). A avaliação de quais regiões geográficas possuem maior concentração de animais e quantificadores para as taxas de nascimentos e sobrevivência são aspectos de suma importância para entender a interação de uma espécie com uma dada região e, até mesmo, no direcionamento de recursos para monitoramento de animais e conservação de seu habitat. Contudo, na maioria das vezes é inviável observar todos os animais da população, seja por conta da dificuldade operacional ocasionada por obstáculos físicos do território, animais ariscos ou por exigir tempo excessivo de execução e alto custo financeiro. Nestes casos, procedimentos estatísticos inferenciais são necessários para obter estimativas para os parâmetros populacionais e, neste contexto, o método de captura-recaptura (CR) mostra-se útil e robusto.

Resumidamente, o método de CR consiste em seleções de amostras aleatórias de animais da população em pelo menos duas ocasiões de amostragem ao longo do tempo. Em cada ocasião, os animais capturados são marcados e, em seguida, devolvidos à população o que possibilita a identificação, em ocasiões posteriores, de animais previamente capturados. A ideia deste delineamento amostral é buscar informações sobre o tamanho populacional a partir dos registros de animais marcados e não marcados em cada ocasião. Com efeito, as capturas de animais marcados (recapturas) informam sobre a taxa de sobrevivência no período enquanto que informações sobre o tamanho populacional serão evidenciadas por meio das capturas de animais não marcados (intuitivamente, a abundância está relacionada com a expressão “quanto maior a proporção de animais não marcados na amostra, maior deve ser a população”). Assim, o processo inferencial basear-se-á nas frequências de animais capturados e recapturados em cada ocasião de amostragem.

O uso do método de CR é antigo. Historicamente, [Petersen \(1896\)](#) aplicou-o em um estudo sobre o fluxo migratório de peixes no mar Báltico e [Lincoln \(1930\)](#) em uma pesquisa

sobre o número de patos selvagens da América do Norte. Ambos utilizaram independentemente o método com duas ocasiões de amostragem e apresentaram a bem conhecida estimativa de Lincoln-Petersen para o tamanho populacional. Contudo, Laplace (1783) havia proposto anteriormente o mesmo procedimento para estimar o tamanho populacional da França (SEBER, 1982). Desde então, mais especificamente a partir da segunda metade do século XX, é notável a gama de trabalhos desenvolvidos na literatura sobre o tema com modelos estatísticos mais gerais e complexos para a estimação do tamanho populacional face às diversas aplicações desta metodologia, como pode ser observado nas revisões da literatura apresentadas por Seber (1986), Seber (1992), Schwarz e Seber (1999), Royle e Dorazio (2008), Thomson, Cooch e Conroy (2008), Amstrup, McDonald e Manly (2010), Royle *et al.* (2013) e McCrea e Morgan (2014).

Para a formulação de um modelo estatístico de CR diversos aspectos do procedimento amostral devem ser considerados, dentre os quais o fator temporal é um dos mais importantes. Em estudos ecológicos nos quais é relativamente curto o tempo para realização de todas as ocasiões de amostragem, é razoável supor que o número de animais da população não se altera no período, isto é, não há nascimentos ou mortes nem imigrações ou emigrações de animais com relação à região investigada. Em uma percepção menos rigorosa, de fato presume-se que esta alteração seja ínfima e desprezível em uma análise estatística. Neste caso, a população é definida como (demográfica e geograficamente) ‘fechada’ (OTIS *et al.*, 1978; CHAO, 2001). Entretanto, a especificação do modelo com base em uma premissa errônea de população fechada pode acarretar perdas significativas na precisão das estimativas para a abundância populacional (SEBER, 1982; KENDALL, 1999; DAIL; MADSEN, 2011). Como alternativa, definimos a população como ‘aberta’ e incorporamos no modelo as possíveis variações do tamanho populacional entre as ocasiões de amostragem.

Pioneiro na modelagem de população aberta, Darroch (1959) determinou estimadores para o tamanho populacional considerando ocorrências de mortes ou imigrações de animais durante o estudo, mas não incluiu o cenário quando ambos ocorrem simultaneamente. Um modelo mais geral foi apresentado posteriormente pelos trabalhos independentes de Jolly (1965) e Seber (1965), no sentido de realizar a estimação conjunta dos números totais de recrutamentos (entradas) e de perdas (saídas) de animais da população entre consecutivas ocasiões de amostragem. Consequentemente, tornou-se possível estimar o número total de animais vivos em cada ocasião a partir da combinação destas duas quantidades ao longo do tempo. É importante ressaltar que o termo ‘recrutamento’ é comumente utilizado nesta abordagem para referenciar de uma maneira unificada os nascimentos e as imigrações de animais pois, em geral, não é possível distinguir entre estes dois casos a partir do protocolo de amostragem. Analogamente, as mortes e emigrações permanentes são relacionadas às ‘perdas’. Esta modelagem firmou-se na literatura como modelo Jolly-Seber (JS) para populações abertas (SEBER, 1992; SCHWARZ, 2001).

Após a formulação do modelo JS, diversas pesquisas foram desenvolvidas sobre o tema a partir de modificações da sua função de verossimilhança (BUCKLAND, 1980; BURNHAM,

1987; CORMACK, 1989; SCHWARZ; ARNASON, 1996; LINK; BARKER, 2005; PRADEL, 1996; BURNHAM, 1991). Por exemplo, propostas de modelagem para inclusão de covariáveis (SCHOFIELD; BARKER, 2011; WU; HOLAN, 2017) ou de classes não observadas sob a perspectiva de modelos de mistura (PLEDGER; POLLOCK; NORRIS, 2010; MATECHOU *et al.*, 2013), bem como estudos considerando a possível perda da marcação do animal (ARNASON; MILLS, 1981; MCDONALD; AMSTRUP; MANLY, 2003; COWEN; SCHWARZ, 2006) e o desenvolvimento de programas de computador para estimação dos parâmetros (WHITE; BURNHAM, 1999; SCHWARZ; ARNASON, 2006; LAAKE; REXSTAD, 2008) são encontrados na literatura.

Por outro lado, é natural pensarmos em modelos estatísticos que contemplem as informações espaciais do processo de CR como, por exemplo, a região definida para amostragem e os locais onde ocorreram as capturas na região. Efetivamente, estas informações adicionais auxiliarão na inferência sobre importantes características da dinâmica territorial da espécie, como quais regiões apresentam maior concentração de animais e sua movimentação na região pesquisada. Assim, uma modelagem estatística que contemple estas características espaciais conjuntamente com a estimação das taxas de recrutamento e sobrevivência torna-se de grande importância para o pesquisador, pois possibilitará uma descrição mais detalhada acerca do comportamento demográfico e geográfico da população em análise.

Sob tais aspectos, esta tese tem como objetivo a formulação, estudo e aplicação de modelos estatísticos capazes de acomodar dados de CR com registro das informações espaciais oriundas do procedimento amostral de uma população aberta. Para efeito de aplicação, direcionamos nossos processos inferenciais para populações residentes em cavernas. Em particular, o habitat destas populações destaca-se por ser restrito e bem definido se comparado, por exemplo, aos habitats de tigres (AZIZ *et al.*, 2017; ROYLE *et al.*, 2017), aves (BORCHERS; EFFORD, 2008; LIEURY *et al.*, 2017) e mamíferos aquáticos (MARQUES *et al.*, 2009; BERTULLI *et al.*, 2018) que, em geral, não apresentam barreiras físicas. Uma vantagem desta aplicação é que os esforços amostrais de captura podem ser convenientemente direcionados para avaliar, se não todo, boa parte do habitat da população. Além disto, efeitos de emigração e imigração são, de certa forma, facilmente distinguíveis dos respectivos casos de nascimentos e mortes neste cenário, apresentando uma maior clareza nas interpretações das estimativas no modelo JS. Enfatizamos a importância de pesquisas nesta direção para a área ecológica, evidenciada pelo expressivo número de trabalhos com o uso do método de CR para estas populações como, por exemplo, espécies de opiliões (WILLEMART; GNASPINI, 2004; GNASPINO, 1996; MONTE, 2017), morcegos (TRAJANO, 1996; O'SHEA; ELLISON; STANLEY, 2004; RIVERS; BUTLIN; ALTRINGHAM, 2006), crustáceos (COOPER; COOPER, 2009; VENARSKY; HURYN; BENS-TEAD, 2012; JUGOVIC *et al.*, 2015) e grilos (MOLDOVAN; FEJÉR, 2013; BERNARDINI *et al.*, 1996; CARCHINI; RAMPINI; SBORDONI, 1994).

Com efeito, a formulação do modelo estatístico de CR será baseada no protocolo de

amostragem adotado e das informações espaciais por ele produzidas e, assim, espera-se que diferentes protocolos produzam específicos dados amostrais que precisarão de um tratamento estatístico particular. Com o intuito de propiciar uma melhor compreensão sobre o assunto e apresentar a natureza dos dados amostrais espaciais que iremos utilizar em nossa modelagem estatística, definimos na próxima seção o problema real que motivou as nossas propostas de pesquisa nesta direção.

1.1 Motivação e objetivos da tese

Nossa pesquisa foi motivada pelo estudo conduzido por Monte (2017) sobre quatro populações de aracnídeos, três das quais serão foco dos processos inferenciais desenvolvidos neste trabalho: *Iandumoema uai* Pinto-Da-Rocha (1996), *Eusarcus cavernicola* Hara e Pinto-Da-Rocha (2010) e *Eusarcus aduncus* Mello-Leitão (1942). O estudo destas espécies foi realizado no interior da caverna da Gruta Olhos d'Água, norte do estado de Minas Gerais. A caverna, inserida nos limites do Parque Nacional Cavernas do Peruaçu, consiste em um longo e sinuoso conduto principal com aproximadamente 9.100 metros (m) de comprimento a partir da entrada. Contudo, para a amostragem destes animais, foi utilizada uma região de estudo delimitada pelos 1.500 m iniciais. A fim de obter informações acerca da distribuição espacial das populações estudadas, esta região foi particionada em 50 regiões de 30 m cada. A Figura 1 exhibe o mapa em planta baixa da região de estudo, onde os pontos pretos delimitam as 50 regiões de busca. O protocolo amostral foi delineado em manter dois pesquisadores por 20 minutos em cada região realizando a captura manual e simultaneamente dos aracnídeos dessas três espécies.

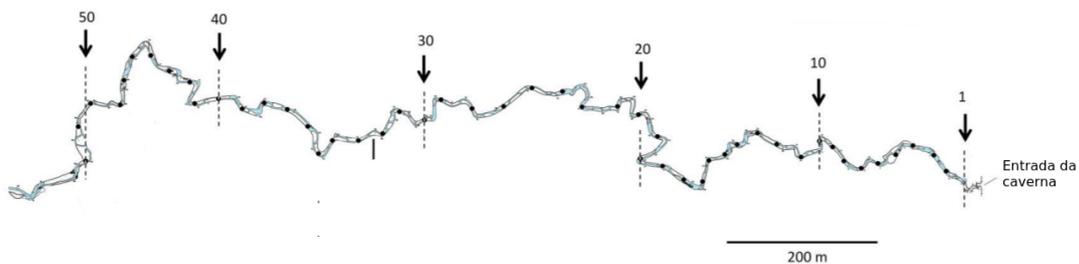


Figura 1 – Mapa em planta baixa da região de estudo. Fonte: Monte (2017).

O processo de CR foi aplicado em seis ocasiões de amostragem definidas, respectivamente, em junho, agosto e outubro do ano de 2014 e junho, setembro e outubro do ano seguinte, 2015. No que tange a marcação, cada aracnídeo capturado foi marcado uma única vez, sendo esta marca sua ‘identidade’ para identificá-lo em futuras recapturas. Para que cada aracnídeo recebesse uma identificação única e diferente dos demais aracnídeos, nove cores enumeradas de 1 a 9 e quatro locais específicos (apresentados na Figura 2) do seu corpo foram utilizados. Para isto, Monte (2017) utiliza uma marcação com tinta lacanitrocelulose e descreve a vantagem do seu uso pois ela é de rápida secagem, sendo adequada para ambientes com alta umidade relativa

do ar. Consequente, foi atribuído a cada aracnídeo um número específico, sendo que a dezena e unidade deste número são marcadas no escudo dorsal (respectivos locais 'a' e 'b') enquanto que a centena é marcada no fêmur IV direito ('c'). O número da ocasião em que o aracnídeo foi capturado pela primeira vez, isto é, a ocasião em que ele foi marcado, é identificada no fêmur IV esquerdo ('d') com a cor correspondente à ocasião. Note que é possível marcar um total de $9 \times 9 \times 9 = 729$ aracnídeos distintos de cada espécie em cada ocasião de amostragem.

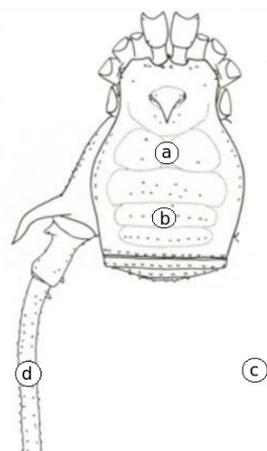


Figura 2 – Locais de marcação: (a) dezenas, (b) unidades, (c) centenas e (d) ocasião da marcação. Fonte: Monte (2017).

Por exemplo, na primeira ocasião, o primeiro aracnídeo marcado recebe a cor referente ao número 1 no fêmur IV esquerdo e no escudo dorsal. O segundo aracnídeo também é marcado com a tinta referente ao número 1 no fêmur IV esquerdo mas com a tinta número 2 no escudo dorsal e assim por diante. Na segunda ocasião de amostragem, repete-se o mesmo procedimento para marcação de todos os aracnídeos não marcados, mas utilizando-se a cor referente ao número 2 no fêmur IV esquerdo. O processo se repete de maneira análoga para as demais ocasiões. Vale ressaltar que não foi necessário identificar com marcação do aracnídeo a respectiva região de sua captura, restando ao pesquisador apenas anotar em seus registros a região em que o respectivo aracnídeo foi marcado. Assim, no momento de sua recaptura é realizada a leitura da sua marcação, constatada sua identidade e anotadas a ocasião e a região atual da recaptura do aracnídeo. No final do estudo, dados amostrais de CR foram obtidos para as espécies de modo que, para cada aracnídeo capturado pelo menos uma vez, foi(ram) registrado(s) a(s) ocasião(ões) e respetiva(s) região(ões) da(s) sua(s) captura(s).

Em sua tese, Monte (2017) apresenta estimativas para os tamanhos populacionais de cada espécie e faz uma análise sobre distribuição espacial dos aracnídeos dentro da caverna. No entanto, estes dois procedimentos foram realizados separadamente em seu trabalho. Desta forma, para realizar uma análise estatística inferencial completa para as três populações consideradas, propomos nesta tese modelos estatísticos que se ajustam a dados obtidos via amostragem de CR com marcação individual realizada em diferentes localidades dentro do habitat, levando-se em consideração as taxas de nascimentos e mortes durante o período de estudo e as localizações geo-

gráficas das capturas. Entretanto, os dados observados evidenciaram comportamentos territoriais diferentes entre as populações. Neste contexto, destacamos a seguir as principais características das espécies, conjuntamente com os objetivos específicos de modelagem estatística para cada uma delas.

Objetivo 1: Constatou-se que o habitat da espécie *Iandumoema uai* é maior do que aquele utilizado para amostragem, isto é, é possível a existência de aracnídeos além da quinquagésima região, uma vez que animais desta população foram capturados ao longo de quase todas as regiões, inclusive na última. Diante disto, no primeiro modelo propomos uma modelagem hierárquica para os tamanhos populacionais em cada região, com o objetivo de obter a distribuição preditiva da abundância populacional para locais não visitados. Uma estrutura para dados zero-inflacionados foi adotada para acomodar situações quando realizam-se amostragens em regiões sem a presença da espécie. Com tal modelagem tornou-se possível obter uma estimativa do tamanho total populacional das regiões mais interiores da caverna e, conseqüentemente, uma estimativa da abundância populacional total da espécie. Para alcançar este objetivo, supomos que os aracnídeos não se movimentam entre as regiões, isto é, permanecem em uma mesma região durante o período de estudo. Esta suposição apresentou-se razoável para esta espécie pois apenas 0,3429% dos aracnídeos (5 de um total de 1458 distintos capturados) foram recapturados em uma região diferente daquela observada em sua primeira captura. Assim, a aplicação do modelo estatístico para os dados de CR da espécie *Iandumoema uai* foi realizada ignorando-se os registros de captura destes cinco aracnídeos.

Objetivo 2: Para as espécies *Eusarcus cavernicola* e *Eusarcus aduncus* propomos um segundo modelo que permite uma movimentação dos aracnídeos entre as regiões de amostragem, generalizando nossa primeira modelagem na qual consideramos a permanência dos aracnídeos em uma única região. Em particular, observamos uma taxa de movimentação mais elevada entre diferentes regiões para estas espécies, algo em torno de 10% dos aracnídeos capturados. Sob esta perspectiva, o modelo proposto possibilita estimar tanto a distribuição espacial dos aracnídeos dentro da caverna como o tamanho da área em que se movimentam durante o período de estudo. Observamos que nenhum aracnídeo destas espécies foi capturado a partir da décima nona região e por esse motivo fixamos na aplicação um habitat dentro das 18 primeiras regiões, que correspondem aos 540 metros iniciais da caverna.

Diante destes objetivos a tese encontra-se organizada da seguinte maneira. Primeiramente, apresentamos no Capítulo 2 a modelagem básica para dados de captura-recaptura (não espacial), levando em consideração as taxas de nascimentos e mortes ao longo das ocasiões de amostragem. Definições, suposições e os parâmetros deste modelo são introduzidos e a explícita forma da

função de verossimilhança apresentada. Este modelo estatístico será então estendido para o contexto espacial nos capítulos seguintes.

No Capítulo 3 propomos o modelo de CR espacial com base nos objetivos traçados para a espécie *Iandumoema uai*. Para o processo inferencial adotamos a abordagem bayesiana e, com o uso de técnicas de dados aumentados, derivamos um algoritmo Monte Carlo em Cadeia de Markov (MCMC) de simples implementação computacional. Obtivemos alguns resultados numéricos para este modelo, analisamos as qualidades das estimativas obtidas a partir de um estudo de simulação com dados gerados artificialmente sob diferentes cenários. No Capítulo 4 aplicamos esta modelagem aos dados reais para a espécie *Iandumoema uai*.

No Capítulo 5 apresentamos o modelo de CR espacial para o ajuste dos dados das espécies *Eusarcus cavernicola* e *Eusarcus aduncus*. Introduzimos o assunto com um levantamento bibliográfico de trabalhos nesta direção e propomos, em seguida, o modelo estatístico para o problema. Adotamos novamente procedimentos bayesianos para as inferências dos parâmetros populacionais por meio de simulações MCMC com o uso de técnicas de dados aumentados para a implementação do algoritmo. No final do capítulo, um estudo de simulação foi realizado para analisar as propriedades frequentistas dos estimadores bayesianos propostos. No Capítulo 6 é realizada uma aplicação com dados das duas espécies em questão.

Finalmente, apresentamos no Capítulo 7 as considerações finais e as conclusões obtidas com o presente trabalho. Alguns comentários sobre projetos de pesquisas futuras sobre o assunto também são apresentados.

MODELO DE CR PARA POPULAÇÃO ABERTA

Resumo: Neste capítulo discutimos os principais pressupostos para a formulação de um modelo estatístico de captura-recaptura para uma população aberta. Introduzimos os dois processos que governam o modelo: o primeiro, que caracteriza as taxas de recrutamentos e sobreviventes ao longo das ocasiões de amostragem, e o segundo que controla as taxas de animais capturados. No final do capítulo, apresentamos as estatísticas amostrais observadas pelo procedimento de marcação adotado e explicitamos a função de verossimilhança para os parâmetros populacionais.

2.1 Modelo estatístico de captura-recaptura

Suponhamos que uma população animal é submetida a um processo de captura-recaptura com T ocasiões de amostragem, $T \in \{2, 3, \dots\}$ e que cada animal, quando capturado pela primeira vez, receba uma marca individual antes de ser devolvido à população, de modo que seja possível identificar futuramente qual(is) ocasião(ões) ele foi previamente capturado.

No que segue, adotaremos as seguintes suposições:

- (i) um animal capturado não perde sua marca até o final do estudo,
- (ii) não há erro na identificação da marca de um animal recapturado,
- (iii) os animais são capturados de maneira independente uns dos outros,
- (iv) a captura de um animal é independente de sua captura em ocasiões anteriores e

- (v) não há recrutamentos (nascimentos ou imigrações) nem saídas (mortes ou emigrações permanentes) de animais durante o decorrer de uma específica ocasião de amostragem, mas estas podem ocorrer entre o final de uma ocasião e o início da próxima.

Pela suposição de população aberta em (v), definimos as seguintes quantidades:

- N_t : número total de animais presentes na população – vivos e dentro dos limites da região de amostragem – na t -ésima ocasião de captura, $t \in \{1, 2, \dots, T\}$,
- R_t : número de recrutamentos na t -ésima ocasião, isto é, número de animais presentes na população na ocasião t e que não estavam em nenhuma anterior, $t \in \{1, 2, \dots, T\}$, e
- D_t : número de perdas de animais da população entre as ocasiões t e $t + 1$, isto é, número de animais presentes na t -ésima ocasião e não presentes na ocasião seguinte, $t \in \{1, 2, \dots, T - 1\}$.

De fato, temos as seguintes relações:

$$N_1 = R_1 \quad \text{e} \quad N_t = N_{t-1} - D_{t-1} + R_t, \quad t \in \{2, 3, \dots, T\}.$$

Para as propostas de modelagem estatística desta tese, consideramos o parâmetro N definido por $N = R_1 + R_2 + \dots + R_T$ e comumente referenciado como o tamanho da superpopulação (SCHWARZ; ARNASON, 1996; DUPUIS; SCHWARZ, 2007; CROSBIE; MANLY, 1985). Este parâmetro corresponde ao número total de animais presentes na população em pelo menos uma ocasião de amostragem. Dado $N \in \{1, 2, \dots\}$, definimos a seguir em quais ocasiões cada animal está presente na população para que possamos calcular as quantidades acima.

2.1.1 Processo de nascimento e morte

Primeiramente, sejam $\mathcal{N} = \{1, 2, \dots, N\}$ e $\mathcal{T} = \{1, 2, \dots, T\}$ os conjuntos de índices referentes aos animais da população e às ocasiões de amostragem, respectivamente. Para cada animal $i \in \mathcal{N}$, consideremos um vetor de variáveis aleatórias categóricas, $\mathbf{Z}_i = (Z_{i,1}, \dots, Z_{i,T})$, definido por

$$Z_{i,t} = \begin{cases} \star & \text{se o animal } i \text{ não foi recrutado até o início da } t\text{-ésima ocasião,} \\ A & \text{se o animal } i \text{ está presente na população no início da } t\text{-ésima ocasião,} \\ \dagger & \text{se o animal } i \text{ deixou a população antes da } t\text{-ésima ocasião,} \end{cases}$$

para $t \in \mathcal{T}$. Por exemplo, para $T = 6$ ocasiões de amostragem, lê-se a configuração $\mathbf{Z}_i = (\star, \star, A, A, A, \dagger)$ da seguinte maneira: o animal i foi recrutado (nasceu ou imigrou para a região de amostragem) após o término da segunda e antes do início da terceira ocasião de amostragem e deixou a população (por morte ou emigração permanente) em algum momento após o término da

quinta e antes do início da última ocasião de amostragem. Naturalmente, somente nas ocasiões 3, 4 e 5 o animal estava suscetível de ser capturado.

Para cada $i \in \mathcal{N}$, supomos que a distribuição de probabilidades da variável aleatória $Z_{i,t+1}$, condicionada às anteriores $(Z_{i,1}, \dots, Z_{i,t})$, só depende da variável $Z_{i,t}$, para $t \in \{1, 2, \dots, T-1\}$. Assim, a distribuição de probabilidades de \mathbf{Z}_i é dada por

$$\begin{aligned} P(\mathbf{Z}_i = \mathbf{z} | \boldsymbol{\gamma}, \boldsymbol{\phi}) &= P(Z_{i,1} = z_1 | \boldsymbol{\gamma}, \boldsymbol{\phi}) \cdot \prod_{t=1}^{T-1} P(Z_{i,t+1} = z_{t+1} | Z_{i,1} = z_1, \dots, Z_{i,t} = z_t, \boldsymbol{\gamma}, \boldsymbol{\phi}) \\ &= P(Z_{i,1} = z_1 | \gamma_1) \cdot \prod_{t=1}^{T-1} P(Z_{i,t+1} = z_{t+1} | Z_{i,t} = z_t, \gamma_{t+1}, \phi_t), \end{aligned} \quad (2.1)$$

para $\mathbf{z} = (z_1, \dots, z_T)$, com $z_t \in \{\star, A, \dagger\}$, e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_T)$ e $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{T-1})$ os vetores que controlam as taxas de recrutamentos e sobreviventes, respectivamente. O parâmetro $\gamma_t \in [0, 1]$ é a probabilidade condicional do animal ser recrutado exatamente na ocasião t , dado que ele não foi em ocasiões anteriores, enquanto $\phi_t \in [0, 1]$ é a probabilidade do animal, presente na população durante a ocasião t , não deixá-la até o início da ocasião $t+1$.

Para a primeira ocasião de amostragem temos

$$P(Z_{i,1} = z | \gamma_1) = \begin{cases} 1 - \gamma_1, & \text{se } z = \star, \\ \gamma_1, & \text{se } z = A, \\ 0, & \text{se } z = \dagger. \end{cases}$$

Notemos que a probabilidade nula acima é adotada para não incluirmos os animais que deixaram a população antes do estudo. Doravante temos três casos para as probabilidades condicionais relativas ao vetor \mathbf{Z}_i :

- para $Z_{i,t} = \star$, temos

$$P(Z_{i,t+1} = z | Z_{i,t} = \star, \gamma_{t+1}, \phi_t) = \begin{cases} 1 - \gamma_{t+1}, & \text{se } z = \star, \\ \gamma_{t+1}, & \text{se } z = A, \\ 0, & \text{se } z = \dagger; \end{cases}$$

- para $Z_{i,t} = A$, temos

$$P(Z_{i,t+1} = z | Z_{i,t} = A, \gamma_{t+1}, \phi_t) = \begin{cases} 0, & \text{se } z = \star, \\ \phi_t, & \text{se } z = A, \\ 1 - \phi_t, & \text{se } z = \dagger; \end{cases}$$

- para $Z_{i,t} = \dagger$, temos

$$P(Z_{i,t+1} = z | Z_{i,t} = \dagger, \gamma_{t+1}, \phi_t) = \begin{cases} 0, & \text{se } z \in \{\star, A\}, \\ 1, & \text{se } z = \dagger. \end{cases}$$

A igualdade $P(Z_{i,t+1} = \dagger | Z_{i,t} = \star, \gamma_{t+1}, \phi_t) = 0$ implica que não incluiremos nas inferências aqueles animais que foram recrutados em algum instante durante o período de estudo mas deixaram a população sem antes estarem presentes em pelo menos uma ocasião de amostragem. Neste mesmo contexto impomos a restrição $\gamma_T = 1$, pois caso o i -ésimo animal da população não tenha sido recrutado até a ocasião $T - 1$, por definição, ele precisa necessariamente estar presente na última. Além disto, se o animal deixou a população em alguma ocasião, ele não será recrutado novamente em ocasiões futuras, isto é, $P(Z_{i,t+1} = \dagger | Z_{i,t} = \dagger, \gamma_{t+1}, \phi_t) = 1$, o que corresponde aos casos de morte ou imigração permanente. Para uma melhor compreensão deste processo e de seus parâmetros, ilustramos na Figura 3 a seguir uma representação gráfica destas probabilidades condicionais.

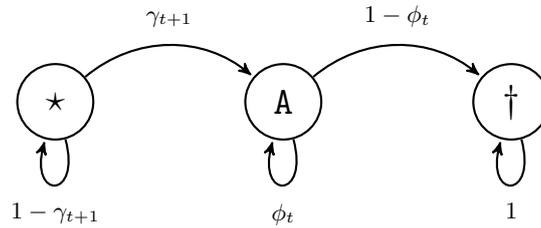


Figura 3 – Representação gráfica das probabilidades condicionais.

Contudo, como o animal é recrutado apenas uma vez durante o estudo, temos que a informação fornecida por $\mathbf{Z}_i = (Z_{i,1}, \dots, Z_{i,T})$ pode ser resumida em apenas duas variáveis aleatórias definidas por

- U_i : primeira ocasião em que o animal $i \in \mathcal{N}$ esteve presente na população, isto é,

$$U_i = \min_{t \in \mathcal{T}} \{t : Z_{i,t} = \mathbf{A}\},$$

- V_i : última ocasião em que o animal $i \in \mathcal{N}$ esteve presente na população, isto é,

$$V_i = \max_{t \in \mathcal{T}} \{t : Z_{i,t} = \mathbf{A}\}.$$

Para simplificar a notação e a implementação computacional, vamos utilizar o par (U_i, V_i) ao invés de \mathbf{Z}_i para descrever o processo de nascimento e morte da população. Consideremos o vetor das probabilidades de recrutamento, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_T)$, cujo t -ésimo elemento, β_t , é a probabilidade de que a primeira ocasião em que o animal esteja presente na população seja a t -ésima ocasião. Logo,

$$\beta_1 = \gamma_1 \quad \text{e} \quad \beta_t = \prod_{l=1}^{t-1} (1 - \gamma_l) \cdot \gamma_t, \quad \text{para } t \in \{2, 3, \dots, T\}, \quad (2.2)$$

com $\beta_t \in [0, 1]$ e $\sum_{t \in \mathcal{T}} \beta_t = 1$. Então, por (2.1) e (2.2), a função de probabilidades conjunta de (U_i, V_i) é dada por

$$P(U_i = u, V_i = v | \boldsymbol{\beta}, \boldsymbol{\phi}) = \beta_u \cdot \left\{ \prod_{t=u}^{v-1} \phi_t \right\}^{I(u < v)} (1 - \phi_v)^{I(v < T)} \cdot I(u \leq v), \quad (2.3)$$

para quaisquer $u \in \mathcal{T}$ e $v \in \mathcal{T}$, onde $I(C)$ é uma função indicadora que assume valor 1 se a condição C for verdadeira ou assume valor 0 caso contrário.

Observe que, dado $N \in \{1, 2, \dots\}$, temos

$$N_t = \sum_{i \in \mathcal{N}} I(U_i \leq t) I(V_i \geq t).$$

Contudo, a observação de cada par (U_i, V_i) é aleatória. Por exemplo, suponha que em $T = 7$ ocasiões de amostragem, um animal seja capturado nas ocasiões 3, 5 e 6, e não capturado nas demais. Neste caso, a menos da quarta ocasião, há uma incerteza se o animal realmente estava presente nas ocasiões em que ele não foi capturado, isto é, se nascera na primeira ou segunda ocasião e se ainda estava presente na última. De fato, o vetor (U_i, V_i) será observado somente se o animal for capturado na primeira e na última ocasiões de amostragem. Caso contrário, este vetor será parcial ou totalmente desconhecido, ou seja, um par de variáveis latentes no estudo. A observação total de (U_i, V_i) dependerá essencialmente da eficiência do processo amostral de captura dos animais, a qual é descrita a seguir.

2.1.2 Processo de captura

Consideremos o vetor de variáveis aleatórias binárias $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,T})$, denominado “histórico de captura” do animal $i \in \mathcal{N}$, com elementos definidos por

$$X_{i,t} = \begin{cases} 1, & \text{se o animal } i \text{ for capturado na } t\text{-ésima ocasião de amostragem,} \\ 0, & \text{caso contrário,} \end{cases}$$

para $t \in \mathcal{T}$. A partir da suposição de independência das capturas entre as ocasiões, definimos as probabilidades marginais por

$$P(X_{i,t} = 1 | U_i = u, V_i = v, \boldsymbol{\theta}_t) = \begin{cases} \theta_t, & \text{para } t \in \{u, u+1, \dots, v\}, \\ 0, & \text{para } t \notin \{u, u+1, \dots, v\}. \end{cases}$$

Pela definição acima, o parâmetro $\theta_t \in (0, 1)$ é a probabilidade de capturar o animal na t -ésima ocasião, dado que ele esteja vivo nesta ocasião, enquanto que este evento será impossível de ocorrer (probabilidade zero) quando ele não estiver.

Contudo, se nesta formulação do modelo JS as probabilidades de captura são funções do tempo, então pelo menos dois parâmetros em $\{\theta_1, \dots, \theta_T\}$ não são identificáveis (SCHWARZ, 2001; SCHWARZ; ARNASON, 2006) e restrições sobre o espaço paramétrico destes parâmetros

precisam ser adotadas. Segundo os autores citados, a identificabilidade se manifesta por diversos meios como, por exemplo, ao fixar $\theta_1 = 1$ (ou $\theta_T = 1$) ou adotar o desenho robusto de Pollock (1982). No entanto, observamos que no primeiro caso perde-se a interpretação de estimativa do número total de animais vivos na primeira (ou na última) ocasião, uma vez que adotamos (provavelmente de maneira errônea) probabilidade certa de capturar todos os animais na respectiva ocasião. Por outro lado, o desenho robusto adota períodos entre duas ou mais ocasiões de amostragem sem recrutamentos ou saídas de animais da população, isto é, população fechada, o que pode ser novamente uma suposição questionável dependendo do intervalo de tempo adotado entre as amostragens.

Uma alternativa é supor uma mesma probabilidade de captura em todas as ocasiões de amostragem, isto é, uma restrição do tipo $\theta_1 = \theta_2 = \dots = \theta_T = \theta$. Esta restrição é comumente utilizada quando esforços amostrais de captura são aplicados igualmente em todas as ocasiões (ROYLE; DORAZIO, 2008; LINK; BARKER, 2005). Nesta tese adotamos esta equiprobabilidade para garantir a identificabilidade do modelo JS, embora a modelagem apresentada a seguir pode ser facilmente modificada para as outras restrições.

Definindo $0^0 = 1$, a função de probabilidades para o histórico de captura é dada por

$$P(\mathbf{X}_i = \mathbf{x} | U_i = u, V_i = v, \theta) = \prod_{t \in \mathcal{T}} \left\{ \left[\theta_t(u, v) \right]^{x_t} \left[1 - \theta_t(u, v) \right]^{1-x_t} \right\} \quad (2.4)$$

para $\mathbf{x} = (x_1, \dots, x_T) \in \{0, 1\}^T$ com

$$\theta_t(u, v) = \theta \cdot I(u \leq t \leq v), \quad \theta \in (0, 1).$$

Pelas relações (2.3) e (2.4), temos

$$\begin{aligned} P(\mathbf{X}_i = \mathbf{x}, U_i = u, V_i = v | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) &= \\ &= P(\mathbf{X}_i = \mathbf{x} | U_i = u, V_i = v, \theta) \cdot P(U_i = u, V_i = v | \boldsymbol{\beta}, \boldsymbol{\phi}) \quad (2.5) \\ &= \prod_{t \in \mathcal{T}} \left\{ \left[\theta_t(u, v) \right]^{x_t} \left[1 - \theta_t(u, v) \right]^{1-x_t} \right\} \cdot \beta_u \left\{ \prod_{t=u}^{v-1} \phi_t \right\}^{I(u < v)} (1 - \phi_v)^{I(v < T)} \cdot I(u \leq v), \end{aligned}$$

para $\mathbf{x} = (x_1, \dots, x_T) \in \{0, 1\}^T$ e $t, u, v \in \mathcal{T}$.

Definimos as funções

$$f(\mathbf{x}) = \begin{cases} \min\{t : x_t \neq 0\}, & \text{se } \mathbf{x} \neq \mathbf{0}, \\ T, & \text{se } \mathbf{x} = \mathbf{0}, \end{cases} \quad (2.6)$$

e

$$l(\mathbf{x}, u) = \begin{cases} \max\{t : x_t \neq 0\}, & \text{se } \mathbf{x} \neq \mathbf{0}, \\ u, & \text{se } \mathbf{x} = \mathbf{0}. \end{cases} \quad (2.7)$$

Logo, somando a função de probabilidades em (2.5) com relação às variáveis latentes (U_i, V_i) , obtemos a função de probabilidades marginal de \mathbf{X}_i dada por,

$$P(\mathbf{X}_i = \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) = \sum_{u=1}^{f(\mathbf{x})} \sum_{v=l(\mathbf{x}, u)}^T P(\mathbf{X}_i = \mathbf{x}, U_i = u, V_i = v | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) \quad (2.8)$$

para todo $\mathbf{x} = (x_1, \dots, x_T) \in \{0, 1\}^T$. A relação (2.8) determina as probabilidades do vetor aleatório gerador dos histórico de captura para os animais da população. Pela suposição de independência condicional entre as capturas dos animais, dados $(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)$, os históricos serão independentes e identicamente distribuídos com a mesma lei de probabilidade (2.8).

2.1.3 Função de verossimilhança

Seja \mathcal{X} o conjunto de todos os possíveis históricos de captura para um determinado animal da população, isto é,

$$\mathcal{X} = \{\mathbf{x} = (x_1, \dots, x_T) : x_t \in \{0, 1\}\} = \{0, 1\}^T.$$

Dado $N \in \{1, 2, \dots\}$, sejam $\mathbf{x}_1, \dots, \mathbf{x}_N$ os históricos de captura de todos os animais da população, com $\mathbf{x}_i \in \mathcal{X}$, $i \in \mathcal{N}$. Para a construção da função de verossimilhança, definimos as frequências

- $n_{[\mathbf{x}]} = \sum_{i \in \mathcal{N}} I(\mathbf{x}_i = \mathbf{x})$ o número de animais da população com histórico de captura \mathbf{x} , $\mathbf{x} \in \mathcal{X}$

e

- $n = \sum_{\mathbf{x} \in \mathcal{X}_0} n_{[\mathbf{x}]}$ o número total de animais distintos capturados desta população,

com $\mathcal{X}_0 = \mathcal{X} \setminus \{\mathbf{0}\}$. No final da etapa de amostragem, a amostra observada será dada pelo conjunto das frequências $\mathbf{n} = \{n_{[\mathbf{x}]}; \mathbf{x} \in \mathcal{X}_0\}$ uma vez que a frequência $n_{[\mathbf{0}]} = N - n$ referente ao número de animais não capturados não é observada.

Denotemos por $\mathcal{C}(\mathbf{n})$ o conjunto de todos os possíveis históricos de captura dos animais que resultam em uma amostra observada \mathbf{n} , isto é,

$$\mathcal{C}(\mathbf{n}) = \left\{ (\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathcal{X}^N : \sum_{i \in \mathcal{N}} I(\mathbf{x}_i = \mathbf{x}) = n_{[\mathbf{x}]}, \text{ para todo } \mathbf{x} \in \mathcal{X}_0 \right\}.$$

Pela independência condicional entre as capturas dos animais, dados $(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)$, e por (2.8), temos que a função de probabilidades de \mathbf{n} é dada por,

$$\begin{aligned}
p(\mathbf{n}|N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) &= P \left[\bigcup_{(\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathcal{C}(\mathbf{n})} (\mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_N = \mathbf{x}_N | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) \right] \\
&= \sum_{(\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathcal{C}(\mathbf{n})} P(\mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_N = \mathbf{x}_N | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) \\
&= \sum_{(\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathcal{C}(\mathbf{n})} \left\{ \prod_{i \in \mathcal{N}} P(\mathbf{X}_i = \mathbf{x}_i | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) \right\} \\
&= \sum_{(\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathcal{C}(\mathbf{n})} \left\{ \prod_{\mathbf{x} \in \mathcal{X}} P(\mathbf{X}_1 = \mathbf{x} | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)^{n_{[\mathbf{x}]}} \right\} \\
&= \frac{N!}{\prod_{\mathbf{x} \in \mathcal{X}} n_{[\mathbf{x}]}} \left\{ \prod_{\mathbf{x} \in \mathcal{X}} P(\mathbf{X}_1 = \mathbf{x} | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)^{n_{[\mathbf{x}]}} \right\}. \tag{2.9}
\end{aligned}$$

O resultado acima segue pois as variáveis $\mathbf{X}_1, \dots, \mathbf{X}_N$ são independentes e identicamente distribuídas. Logo, \mathbf{n} é uma realização de uma distribuição multinomial com N ensaios e probabilidade de cada evento $\mathbf{x} \in \mathcal{X}$ determinado por \mathbf{X}_1 e a função de verossimilhança para os parâmetros $(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)$ é definida por,

$$\begin{aligned}
L(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta; \mathbf{n}) &= p(\mathbf{n}|N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) \\
&= \frac{N!}{(N-n)! \prod_{\mathbf{x} \in \mathcal{X}_0} n_{[\mathbf{x}]}} P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)^{N-n} \prod_{\mathbf{x} \in \mathcal{X}_0} P(\mathbf{X}_1 = \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)^{n_{[\mathbf{x}]}} \tag{2.10}
\end{aligned}$$

para $N \in \{n, n+1, \dots\}$, $\theta \in (0, 1)$, $\boldsymbol{\beta} \in \Theta_{\boldsymbol{\beta}}$ e $\boldsymbol{\phi} \in \Theta_{\boldsymbol{\phi}}$, com

$$\Theta_{\boldsymbol{\beta}} = \left\{ (a_1, \dots, a_T) : a_t \in [0, 1], \sum_{t \in \mathcal{T}} a_t = 1 \right\}$$

e

$$\Theta_{\boldsymbol{\phi}} = \left\{ (a_1, \dots, a_{T-1}) : a_t \in [0, 1] \right\}.$$

Processos inferenciais para os parâmetros populacionais $(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)$ poderão ser baseados na função de verossimilhança acima. Observamos que esta função não considera qualquer informação espacial que pode ser obtida durante o procedimento amostral. Assim, apresentamos nos próximos capítulos duas propostas de modificações desta função para o contexto espacial de CR.

MODELO DE CR HIERÁRQUICO PARA POPULAÇÃO ABERTA

Resumo: Neste capítulo apresentamos um modelo estatístico para estimação da abundância populacional para população aberta localizada em uma determinada região geográfica. A partir de dados obtidos via amostragem de captura-marcação-recaptura realizada em diferentes localidades desta região, propomos uma modelagem bayesiana hierárquica para os tamanhos populacionais locais a fim de obter a distribuição preditiva da abundância populacional para regiões não visitadas pela amostragem. No final do capítulo analisamos as propriedades frequentistas dos estimadores bayesianos para os parâmetros populacionais através de um estudo de simulação.

3.1 Modelo estatístico

Para uma determinada população animal, consideremos a realização de um processo de captura-recaptura com $T \in \{2, 3, \dots\}$ ocasiões de amostragem de acordo com o protocolo de marcação apresentado no Capítulo 2 e assumindo válidas as suposições (i), (ii), (iii), (iv) e (v), página 26. Suponhamos que a amostragem seja realizada em uma área particionada em R regiões (locais) disjuntos tal que

- (a) cada animal estará presente em uma mesma região durante a realização do estudo, isto é, supomos que não ocorrem emigrações ou imigrações de animais entre as regiões,
- (b) os esforços amostrais de captura são realizados de maneira independente entre as regiões e
- (c) cada ocasião é definida como a realização da amostragem em todas as regiões simultaneamente ou, em um curto período de tempo, de forma sequencial.

Definindo por $\mathcal{R} = \{1, 2, \dots, R\}$ o conjunto de índices referentes às regiões de amostragem, consideramos neste capítulo os seguintes parâmetros populacionais de interesse:

- N_r : número total de animais da população residentes na r -ésima região durante o período de estudo, isto é, número total de animais desta região vivos em pelo menos uma ocasião de amostragem, $r \in \mathcal{R}$,
- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_T)$: vetor de probabilidades de nascimento da população,
- $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{T-1})$: vetor de probabilidades de sobrevivência da população e
- θ : probabilidade de captura,

com $(\boldsymbol{\beta}, \boldsymbol{\phi}, \theta)$ comum para todas as regiões.

Dado $N_r \in \{1, 2, \dots\}$, seja $\mathbf{x}_i^r = (x_{i,1}^r, \dots, x_{i,T}^r)$ o histórico de captura do i -ésimo animal residente na região r , $r \in \mathcal{R}$, $i \in \mathcal{N}_r = \{1, 2, \dots, N_r\}$, com

$$\mathbf{x}_i^r \in \mathcal{X} = \{\mathbf{x} = (x_1, \dots, x_T) : x_t \in \{0, 1\}\}.$$

Assumimos que \mathbf{x}_i^r é uma observação de um vetor aleatório $\mathbf{X} = (X_1, \dots, X_T)$ distribuído segundo a função de probabilidades dada por (2.8), onde o evento $(X_t = 1)$ indica que o respectivo animal foi capturado na ocasião $t \in \mathcal{T}$ e $(X_t = 0)$ indica não captura.

A partir da suposição (a) acima, podemos definir as frequências amostrais de captura para cada região $r \in \mathcal{R}$ separadamente, isto é,

- $n_{r, [\mathbf{x}]} = \sum_{i \in \mathcal{N}_r} I(\mathbf{x}_i^r = \mathbf{x})$ o número de animais da região $r \in \mathcal{R}$ observados com histórico de captura $\mathbf{x} \in \mathcal{X}$ e
- $n_r = \sum_{\mathbf{x} \in \mathcal{X}_0} n_{r, [\mathbf{x}]}$ o número de animais distintos capturados na região r ,

onde $\mathcal{X}_0 = \mathcal{X} \setminus \{\mathbf{0}\}$. Assim, a amostra observada pelo procedimento amostral sobre todas as regiões será denotada por $\mathbf{n} = (\mathbf{n}_1, \dots, \mathbf{n}_R)$, onde $\mathbf{n}_r = \{n_{r, [\mathbf{x}]}; \mathbf{x} \in \mathcal{X}_0\}$ é o conjunto das frequências observadas de captura para a região r . Dados $(N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)$ e pelas suposições adotadas, temos que cada observação \mathbf{n}_r é uma realização independente de uma distribuição com função de probabilidades dada por (2.9). Neste capítulo permitiremos regiões com $N_r = 0$ e, neste caso, assumimos que $\mathbf{n}_r = \mathbf{0}$ com probabilidade 1.

Logo, a função de verossimilhança de $(N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)$ associada à amostra observada \mathbf{n} é definida por,

$$\begin{aligned}
\mathcal{L}(N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta; \mathbf{n}) &= \\
&= p(\mathbf{n}_1, \dots, \mathbf{n}_R | N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) \\
&= \prod_{r \in \mathcal{R}} p(\mathbf{n}_r | N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta) \\
&= \prod_{r \in \mathcal{R}} \left\{ \frac{N_r!}{(N_r - n_r)! \prod_{\mathbf{x} \in \mathcal{X}_0} n_{r, [\mathbf{x}]!}} P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)^{N_r - n_r} \prod_{\mathbf{x} \in \mathcal{X}_0} P(\mathbf{X}_1 = \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)^{n_{r, [\mathbf{x}]}} \right\} \\
&\propto \left\{ \prod_{r \in \mathcal{R}} \frac{N_r!}{(N_r - n_r)!} \right\} P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)^{\sum_{r \in \mathcal{R}} (N_r - n_r)} \prod_{\mathbf{x} \in \mathcal{X}_0} P(\mathbf{X}_1 = \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta)^{\sum_{r \in \mathcal{R}} n_{r, [\mathbf{x}]}} , \quad (3.1)
\end{aligned}$$

para $N_r \in \{n_r, n_r + 1, \dots\}$, $r \in \mathcal{R}$, $\theta \in (0, 1)$, $\boldsymbol{\beta} \in \Theta_\beta$ e $\boldsymbol{\phi} \in \Theta_\phi$. Observamos que a expressão acima é igual ao produto de R funções de verossimilhanças da forma como apresentada em (2.10).

3.2 Abordagem Bayesiana

Nesta tese adotamos o paradigma bayesiano para o processo inferencial dos parâmetros populacionais (CASTLEDINE, 1981; GELMAN *et al.*, 1995). Nesta abordagem a função de verossimilhança, que traz informações amostrais sobre os parâmetros, é combinada com o conhecimento prévio do especialista sobre os parâmetros expresso através de distribuições de probabilidades, ditas a priori. A combinação destas duas fontes de informações resulta em uma distribuição, denominada posteriori, para os parâmetros e é baseado nela que serão realizadas as inferências.

Supondo independência a priori entre os parâmetros, definimos as seguintes distribuições a priori marginais:

$$\begin{aligned}
\boldsymbol{\beta} &\sim \text{Dirichlet}(a_1, \dots, a_T), \\
\phi_t &\sim \text{Beta}(b_{1,t}, b_{2,t}), \quad t \in \{1, \dots, T - 1\}, \\
\theta &\sim \text{Beta}(c_1, c_2),
\end{aligned} \quad (3.2)$$

com os valores dos hiperparâmetros a_i 's, $b_{i,t}$'s e c_i 's positivos e pré-determinados pelo pesquisador.

Por outro lado, embora os processos de CR sejam realizados independentemente entre as regiões, a suposição de independência dos tamanhos populacionais pode ser inadequada tratando-se de amostragens espaciais de um mesmo habitat. Além disto, caso haja regiões não amostradas com possível presença de animais da espécie, processos preditivos de abundâncias nestes locais podem ser requeridos e, assim, uma estrutura de dependência entre as regiões

deve ser adotada. Uma maneira relativamente simples de incorporar dependência entre variáveis aleatórias é por meio de distribuições condicionais a efeitos aleatórios. Por exemplo, [Royle e Dorazio \(2006\)](#) apresentam uma modelagem na forma

$$\begin{aligned} N_r | \lambda_r &\overset{ind}{\sim} \text{Poisson}(\lambda_r), \\ \log(\lambda_r) | \mu, \sigma &\overset{ind}{\sim} \text{Normal}(\mu, \sigma^2) \end{aligned} \quad (3.3)$$

onde a distribuição marginal de N_r é denominada Poisson-LogNormal ([BULMER, 1974](#)) com parâmetros (μ, σ) . De fato, a dependência entre os tamanhos populacionais surge a partir do desconhecimento dos parâmetros da distribuição dos λ_r 's. No contexto bayesiano, por exemplo, podemos adotar um nível adicional de hierarquia assumindo

$$\begin{aligned} \mu &\sim \text{Normal}(\mu_0, \xi_0^2), \\ \sigma^{-2} &\sim \text{Gama}(a, b), \end{aligned}$$

com os valores de μ_0, ξ_0, a, b determinados de acordo com a crença a priori sobre estas quantidades. Desta forma, apesar da distribuição Poisson apresentar valor esperado igual à variância, isto é, $\mathbb{E}(N_r | \lambda_r) = \lambda_r = \text{Var}(N_r | \lambda_r)$, a estrutura hierárquica adotada tornará as distribuições a priori marginais dos tamanhos populacionais mais flexíveis no sentido de acomodar possível variabilidade extra-Poisson. Especificamente, na distribuição Poisson-LogNormal a esperança e variância a priori podem ser obtidas, respectivamente, por

$$\mathbb{E}(N_r) = \mathbb{E}[\mathbb{E}(N_r | \lambda_r)] = \mathbb{E}(\lambda_r) = e^{\mu + \frac{\sigma^2}{2}} \quad (3.4)$$

e

$$\begin{aligned} \text{Var}(N_r) &= \mathbb{E}[\text{Var}(N_r | \lambda_r)] + \text{Var}[\mathbb{E}(N_r | \lambda_r)] \\ &= \mathbb{E}(\lambda_r) + \text{Var}(\lambda_r) \\ &= e^{\mu + \frac{\sigma^2}{2}} + e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \\ &> \mathbb{E}(N_r). \end{aligned}$$

Assim, suponhamos que uma região não amostrada tem abundância populacional seguindo a estrutura hierárquica dada por (3.3), então torna-se possível prever seu valor médio por (3.4). Para um confiável processo de predição para abundâncias em locais não visitados, estimativas robustas são necessárias para os parâmetros (μ, σ) . Por outro lado, amostragens em locais onde não há presença de animais da espécie podem interferir negativamente na precisão da predição. Neste caso, o conjunto de dados de CR apresentará um número excessivo de regiões sem animais capturados e, provavelmente, estimativas para abundâncias nestes locais serão nulas. Se o número de zeros for maior que o esperado pela distribuição, estes dados são considerados zero-inflacionados. Dados zero-inflacionados é um caso especial de variabilidade extra-Poisson

e ignorar esta inflação leva, em geral, ao decréscimo do valor médio estimado da distribuição (μ) e a um valor sobrestimado para sua variabilidade, σ .

A fim de modelarmos dados zero-inflacionados, consideraremos a seguinte distribuição de probabilidades alternativa para o parâmetro N_r :

$$N_r | \lambda_r, \omega \stackrel{ind}{\sim} \text{ZIPoisson}(\lambda_r, \omega), \quad (3.5)$$

com $\lambda_r > 0$ e $\omega \in [0, 1]$, onde $\text{ZIPoisson}(\lambda_r, \omega)$ denota a distribuição de Poisson Zero-Inflacionada, cuja função de probabilidades é dada por

$$P(N_r = m | \lambda_r, \omega) = (1 - \omega)I(m = 0) + \omega \frac{e^{-\lambda_r} \lambda_r^m}{m!} I(m \in \mathbb{N}_0), \quad (3.6)$$

com $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Assim, se $\omega \in (0, 1)$ há uma maior probabilidade da variável aleatória assumir o valor zero se comparada à distribuição de Poisson caracterizando, assim, uma inflação de observações zero. Se $\omega = 0$, N_r segue uma distribuição degenerada em zero, enquanto que se $\omega = 1$, N_r tem uma distribuição Poisson de parâmetro λ_r . A probabilidade da região r não ser habitada pela espécie durante o período de estudo é

$$P(N_r = 0 | \omega, \mu, \sigma) = (1 - \omega) + \omega \int_0^{+\infty} e^{-\lambda_r} f(\lambda_r | \mu, \sigma) d\lambda_r,$$

onde f é a função densidade de probabilidades da distribuição Log-Normal.

Finalmente, propomos a seguinte estrutura:

$$\begin{aligned} N_r | \lambda_r, \omega &\stackrel{ind}{\sim} \text{ZIPoisson}(\lambda_r, \omega), \\ \log(\lambda_r) | \mu, \sigma &\stackrel{ind}{\sim} \text{Normal}(\mu, \sigma^2), \\ \mu &\sim \text{Normal}(\mu_0, \xi_0^2), \\ \sigma^{-2} &\sim \text{Gama}(a, b), \\ \omega &\sim \text{Beta}(g_1, g_2). \end{aligned} \quad (3.7)$$

Na literatura, [Joseph et al. \(2009\)](#) propuseram uma modelagem por distribuições Binomial Negativa Zero-Inflacionada, que consiste em assumir $\lambda_r \sim \text{Gama}(a, b)$ em (3.7), que é naturalmente a extensão da modelagem por Binomial Negativa proposta inicialmente por [Royle \(2004\)](#) neste contexto. Contudo, vale ressaltar que ambos os trabalhos abordam o processo inferencial por máxima verossimilhança e, assim como [Royle e Dorazio \(2006\)](#), consideraram dados sem marcação animal, isto é, quando se registra apenas o número de capturados em cada ocasião e região de amostragem. Tanto quanto é do nosso conhecimento, não há na literatura estatística trabalhos que propusessem esta modelagem hierárquica no contexto bayesiano para populações abertas e marcadas.

A distribuição a posteriori para os parâmetros populacionais é escrita proporcional ao produto da função de verossimilhança e a estrutura a priori especificada acima, isto é,

$$f(\mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \sigma^2, \boldsymbol{\omega} | \mathbf{n}) \propto \mathcal{L}(N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{n}) \times \\ \times \prod_{r \in \mathcal{R}} f(N_r | \lambda_r, \boldsymbol{\omega}) f(\boldsymbol{\beta}) f(\boldsymbol{\phi}) f(\boldsymbol{\theta}) f(\boldsymbol{\lambda} | \boldsymbol{\mu}, \boldsymbol{\sigma}) f(\boldsymbol{\mu}) f(\sigma^2) f(\boldsymbol{\omega}),$$

com $\mathbf{N} = (N_1, \dots, N_R)$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_R)$ e f representa genericamente a função de probabilidades (ou função densidade de probabilidades) dos argumentos especificados.

Contudo, processos inferenciais baseados nas funções de verossimilhança acima terão custos computacionais elevados uma vez que tais funções envolvem produto de somatórios. Geralmente esta característica da função de verossimilhança é devido ao fato do processo amostral não obter a informação completa sobre os parâmetros. Em nosso caso, a complexidade da função de verossimilhança é decorrente da impossibilidade de se afirmar em quais épocas os animais estavam vivos em alguns históricos de captura observados. Por outro lado, é possível contornar tal dificuldade considerando técnicas de aumento de dados (TANNER; WONG, 1987; DYK; MENG, 2001), incluindo informações latentes durante o processo amostral aos dados observados obtendo assim uma função de verossimilhança aumentada que é apropriada computacionalmente para manipulações numéricas.

3.2.1 Aumento de dados

Primeiramente, consideraremos o aumento de dados para a distribuição das abundâncias populacionais locais, N_r 's. Notemos que a função de probabilidades (3.5) pode ser vista como uma mistura de uma distribuição degenerada em zero e uma distribuição de Poisson com parâmetro λ_r , com pesos proporcionais à $(1 - \omega)$ e ω , respectivamente. Para modelar distribuições de mistura no contexto bayesiano, vamos utilizar as técnicas apresentadas por Diebolt e Robert (1994) e por Rodrigues (2003) especificamente para a distribuição de Poisson Zero-Inflacionada. A ideia do aumento de dados neste contexto é especificar, por meio de uma variável latente, sob qual das duas distribuições (degenerada em zero ou Poisson) o valor de N_r foi gerado. Assim, propomos incluir no modelo a seguinte variável aleatória (v.a.):

$$O_r = \begin{cases} 1, & \text{se } N_r \text{ é uma v.a. com distribuição Poisson}(\lambda_r), \text{ ou} \\ 0, & \text{se } N_r \text{ é uma v.a. com distribuição degenerada em 0.} \end{cases}$$

Então,

$$O_r \sim \text{Bernoulli}(\omega) \tag{3.8}$$

e a distribuição de probabilidades a priori conjunta de (N_r, O_r) é dada por,

$$\begin{aligned} P(N_r = m, O_r = o | \lambda_r, \omega) &= P(N_r = m | O_r = o, \lambda_r) P(O_r = o | \omega) \\ &= \left[\frac{\lambda_r^m e^{-\lambda_r}}{m!} I(m \in \mathbb{N}_0) \right]^o \left[I(m = 0) \right]^{1-o} \omega^o (1 - \omega)^{1-o} I(o \in \{0, 1\}). \end{aligned} \quad (3.9)$$

Observemos que (3.6) é obtido somando (3.9) com relação à variável latente O_r e, assim, temos um aumento de dados com relação à distribuição definida em (3.5).

No que segue, vamos considerar o aumento de dados com relação à função de verossimilhança. Para cada região $r \in \mathcal{R}$, suponhamos um valor de $N_r \in \{1, 2, \dots\}$ fixado. Sejam (u_i^r, v_i^r) o par que representa a primeira e última ocasião em que um animal $i \in \mathcal{N}_r$ esteve presente (vivo) na região $r \in \mathcal{R}$. Definimos por $\tilde{n}_{r, [\mathbf{x}]}^{(u, v)}$ o número de animais da região r vivos exclusivamente durante as ocasiões $u, u + 1, \dots, v$ e capturados com histórico \mathbf{x} , para todo $\mathbf{x} \in \mathcal{X}$ e $1 \leq u \leq f(\mathbf{x})$ e $l(\mathbf{x}, u) \leq v \leq T$, isto é,

$$\tilde{n}_{r, [\mathbf{x}]}^{(u, v)} = \sum_{i \in \mathcal{N}_r} I(\mathbf{x}_i^r = \mathbf{x}, u_i^r = u, v_i^r = v),$$

onde as funções f e l são definidas em (2.6) e (2.7), respectivamente. Como estas frequências latentes satisfazem

$$\sum_{u=1}^{f(\mathbf{x})} \sum_{v=l(\mathbf{x}, u)}^T \tilde{n}_{r, [\mathbf{x}]}^{(u, v)} = n_{r, [\mathbf{x}]}, \quad \mathbf{x} \in \mathcal{X}, \quad r \in \mathcal{R}, \quad (3.10)$$

então, dado N_r , consideramos o conjunto de frequências

$$\tilde{\mathbf{n}}_r = \left\{ \tilde{n}_{r, [\mathbf{x}]}^{(u, v)}; \mathbf{x} \in \mathcal{X}, 1 \leq u \leq f(\mathbf{x}), l(\mathbf{x}, u) \leq v \leq T \right\}$$

como o aumento de dados para o conjunto $\mathbf{n}_r \cup \{n_{r, [\emptyset]}\}$.

Seja $\tilde{\mathbf{n}} = (\tilde{\mathbf{n}}_1, \dots, \tilde{\mathbf{n}}_R)$. Portanto, a função de verossimilhança aumentada para os parâmetros $(N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})$ associada à informação completa $(\mathbf{n}, \tilde{\mathbf{n}})$ é definida por

$$\begin{aligned} \tilde{\mathcal{L}}(N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{n}, \tilde{\mathbf{n}}) &= \\ &= p(\mathbf{n}, \tilde{\mathbf{n}} | N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) \\ &= \prod_{r \in \mathcal{R}} p(\mathbf{n}_r, \tilde{\mathbf{n}}_r | N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) \\ &= \prod_{r \in \mathcal{R}} p(\mathbf{n}_r | \tilde{\mathbf{n}}_r, N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) p(\tilde{\mathbf{n}}_r | N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}), \end{aligned} \quad (3.11)$$

onde

$$p(\mathbf{n}_r | \tilde{\mathbf{n}}_r, N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) = \begin{cases} 1, & \text{se } n_{r, [\mathbf{x}]} = \sum_{u=1}^{f(\mathbf{x})} \sum_{v=l(\mathbf{x}, u)}^T \tilde{n}_{r, [\mathbf{x}]}^{(u, v)} \text{ para todo } \mathbf{x} \in \mathcal{X}, \\ 0, & \text{caso contrário,} \end{cases}$$

e $p(\tilde{\mathbf{n}}_r|N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})$ é a função de probabilidades de uma distribuição multinomial com N_r ensaios e probabilidades multinomiais dadas pela equação (2.5). Logo, a função de verossimilhança aumentada é não nula se e somente se a restrição (3.10) for satisfeita e, neste caso, é dada por

$$\begin{aligned} \tilde{\mathcal{L}}(N_1, \dots, N_R, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{n}, \tilde{\mathbf{n}}) &= \\ &= \prod_{r \in \mathcal{R}} p(\tilde{\mathbf{n}}_r|N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) \\ &= \prod_{r \in \mathcal{R}} \frac{N_r!}{\prod_{\mathbf{x} \in \mathcal{X}} \prod_{u=1}^{f(\mathbf{x})} \prod_{v=l(\mathbf{x},u)}^T \tilde{n}_{r,[\mathbf{x}]}^{(u,v)}!} \prod_{\mathbf{x} \in \mathcal{X}} \prod_{u=1}^{f(\mathbf{x})} \prod_{v=l(\mathbf{x},u)}^T \left\{ P(\mathbf{X}_1 = \mathbf{x}, U_1 = u, V_1 = u | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) \right\}^{\tilde{n}_{r,[\mathbf{x}]}^{(u,v)}} \end{aligned} \quad (3.12)$$

$$\begin{aligned} &= \prod_{r \in \mathcal{R}} \left\{ \frac{N_r!}{\prod_{\mathbf{x} \in \mathcal{X}} \prod_{u=1}^{f(\mathbf{x})} \prod_{v=l(\mathbf{x},u)}^T \tilde{n}_{r,[\mathbf{x}]}^{(u,v)}!} \right\} \times \prod_{\mathbf{x} \in \mathcal{X}} \prod_{u=1}^{f(\mathbf{x})} \prod_{v=l(\mathbf{x},u)}^T \left\{ \prod_{t=u}^v \theta^{x_t} [1 - \theta]^{1-x_t} \beta_u \times \right. \\ &\quad \left. \times \left[\prod_{t=u}^{v-1} \phi_t \right]^{I(u < v)} (1 - \phi_v)^{I(v < T)} \right\}^{\sum_{r \in \mathcal{R}} \tilde{n}_{r,[\mathbf{x}]}^{(u,v)}} \\ &\propto \prod_{r \in \mathcal{R}} \left\{ \frac{N_r!}{\prod_{u=1}^T \prod_{v=u}^T \tilde{n}_{r,[\mathbf{0}]}^{(u,v)}!} \right\} \theta^{\bar{n}} (1 - \theta)^{\bar{N} - \bar{n}} \left\{ \prod_{t \in \mathcal{T}} \beta_t^{R_t} \right\} \left\{ \prod_{t=1}^{T-1} \phi_t^{W_t} (1 - \phi_t)^{N_t - W_t} \right\}, \end{aligned}$$

com

- $\bar{n} = \sum_{r \in \mathcal{R}} \sum_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{t \in \mathcal{T}} x_t \right\} n_{r,[\mathbf{x}]}$ o número total de capturas e recapturas,
- $\bar{N} = \sum_{t \in \mathcal{T}} N_t$, onde $N_t = \sum_{r \in \mathcal{R}} \sum_{\mathbf{x} \in \mathcal{X}} \sum_{u=1}^t \sum_{v=t}^T \tilde{n}_{r,[\mathbf{x}]}^{(u,v)}$ o número de animais vivos na ocasião t ,
- $R_t = \sum_{r \in \mathcal{R}} \sum_{\mathbf{x} \in \mathcal{X}} \sum_{v=t}^T \tilde{n}_{r,[\mathbf{x}]}^{(t,v)}$ o número de animais que nasceram na ocasião t ,
- $W_t = \sum_{r \in \mathcal{R}} \sum_{\mathbf{x} \in \mathcal{X}} \sum_{u=1}^t \sum_{v=t+1}^T \tilde{n}_{r,[\mathbf{x}]}^{(u,v)}$ o número de animais que sobreviveram entre as ocasiões t e $t+1$,

e válida a restrição $N_r = n_r + \sum_{u=1}^T \sum_{v=u}^T \tilde{n}_{r,[\mathbf{0}]}^{(u,v)}$. Observe que o número de perdas entre as ocasiões t e $t+1$ é dada por $D_t = N_t - W_t$. Ressaltamos que, dado uma configuração fixada para o conjunto de dados observados \mathbf{n} , se (3.10) for válida, então a função de verossimilhança (3.1) pode ser

obtida somando-se com relação a todos os possíveis valores das variáveis latentes $\tilde{\mathbf{n}}$ na função (3.12).

Portanto, a distribuição a posteriori conjunta das quantidades desconhecidas (parâmetros e variáveis latentes) é proporcional ao produto da função de verossimilhança aumentada e das distribuições a priori especificadas:

$$f(\mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \mu, \boldsymbol{\lambda}, \sigma^2, \omega, \mathbf{O}, \tilde{\mathbf{n}}|\mathbf{n}) \propto \tilde{\mathcal{L}}(\mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta; \mathbf{n}, \tilde{\mathbf{n}}) \times \prod_{r \in \mathcal{R}} f(N_r | \lambda_r, O_r) f(O_r | \omega) f(\boldsymbol{\beta}) f(\boldsymbol{\phi}) f(\theta) f(\boldsymbol{\lambda} | \mu, \sigma) f(\mu) f(\sigma^2) f(\omega), \quad (3.13)$$

com $\mathbf{O} = (O_1, \dots, O_R)$.

3.2.2 Algoritmo MCMC

Para realização das inferências, utilizaremos o bem-conhecido algoritmo de Monte Carlo via Cadeia de Markov (MCMC). Especificamente, o algoritmo baseia-se em amostragem por Gibbs-sampling com Metropolis-Hasting (CASELLA; GEORGE, 1992; GILKS; RICHARDSON; SPIEGELHALTER, 1995), que demanda a especificação das distribuições a posteriori condicionais das quantidades desconhecidas. Notemos que, além dos parâmetros desconhecidos, precisamos atualizar os valores das frequências latentes, \mathbf{O} e $\tilde{\mathbf{n}}$, dada a observação \mathbf{n} . Para tanto, definimos o algoritmo MCMC considerando quatro blocos de atualizações:

- Bloco I: atualiza os valores de $(\mathbf{N}, \mathbf{O}, \tilde{\mathbf{n}})$ a partir da distribuição condicional dados valores fixos de $(\mathbf{n}, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \boldsymbol{\lambda}, \mu, \sigma^2, \omega)$,
- Bloco II: atualiza os valores de $(\boldsymbol{\beta}, \boldsymbol{\phi}, \theta)$ a partir da distribuição condicional dados valores fixos de $(\mathbf{N}, \mathbf{O}, \tilde{\mathbf{n}}, \mathbf{n}, \boldsymbol{\lambda}, \mu, \sigma^2, \omega)$,
- Bloco III: atualiza $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_R)$ a partir da distribuição condicional dados valores fixos de $(\mathbf{N}, \mathbf{O}, \tilde{\mathbf{n}}, \mathbf{n}, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \mu, \sigma^2, \omega)$, e
- Bloco IV: atualiza (μ, σ, ω) a partir da distribuição condicional dados valores fixos de $(\mathbf{N}, \mathbf{O}, \tilde{\mathbf{n}}, \mathbf{n}, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \boldsymbol{\lambda})$.

As distribuições condicionais para a implementação do algoritmo são apresentadas a seguir.

Bloco I: atualização de $(\mathbf{N}, \mathbf{O}, \tilde{\mathbf{n}})$

Para a realização do Bloco I, notamos que a distribuição condicional a posteriori de $(N_r, O_r, \tilde{\mathbf{n}}_r)$, dados $(\boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \lambda_r, \mathbf{n}_r, \omega)$, é independente das demais $(N_s, O_s, \tilde{\mathbf{n}}_s)$, para $s \neq r$, uma vez que as distribuições a priori (3.7) e a função de verossimilhança aumentada (3.11) são fatoradas por região e, portanto, podem ser atualizadas separadamente. Além disto, pela regra

do produto de probabilidades, podemos escrever a função de probabilidades condicional de $(N_r, O_r, \tilde{\mathbf{n}}_r)$ como

$$\begin{aligned} p(N_r, O_r, \tilde{\mathbf{n}}_r | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r, \omega) &= \\ &= p(O_r | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r, \omega) \cdot p(N_r | O_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r) \cdot p(\tilde{\mathbf{n}}_r | N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r). \end{aligned} \quad (3.14)$$

Logo, é possível dividir a geração de valores para $(N_r, O_r, \tilde{\mathbf{n}}_r)$ em três etapas. Primeiramente, atualizamos O_r por meio da distribuição condicional à esquerda de (3.14). Observamos que, nesta etapa, $n_r > 0$ implica que $N_r > 0$ e, conseqüentemente, $O_r = 1$ com probabilidade 1 de acordo com a distribuição de probabilidade definida em (3.9). Por outro lado, o desconhecimento do valor para a variável latente O_r dá-se quando $n_r = 0$, isto é, não há captura na região r . Neste caso, geramos O_r a partir da seguinte distribuição a posteriori condicional:

$$\begin{aligned} P(O_r = 1 | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r, \omega) &= \sum_{m=0}^{\infty} P(N_r = m, O_r = 1 | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r, \omega) \\ &= \sum_{m=0}^{\infty} P(N_r = m | O_r = 1, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r) P(O_r = 1 | \omega) \\ &\propto \sum_{m=0}^{\infty} p(\mathbf{n}_r | m, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) P(N_r = m | \lambda_r, O_r = 1) P(O_r = 1 | \omega) \\ &\propto \sum_{m=0}^{\infty} P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})^m \frac{\lambda_r^m e^{-\lambda_r}}{m!} \omega \\ &= \omega e^{-\lambda_r [1 - P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})]} \end{aligned}$$

e

$$\begin{aligned} P(O_r = 0 | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r, \omega) &= \sum_{m=0}^{\infty} P(N_r = m, O_r = 0 | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r, \omega) \\ &= \sum_{m=0}^{\infty} P(N_r = m | O_r = 0, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r) P(O_r = 0 | \omega) \\ &\propto \sum_{m=0}^{\infty} p(\mathbf{n}_r | m, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) P(N_r = m | \lambda_r, O_r = 0) P(O_r = 0 | \omega) \\ &\propto \sum_{m=0}^{\infty} P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})^m I(m = 0) (1 - \omega) \\ &= (1 - \omega). \end{aligned}$$

Assim, quando $n_r = 0$, geramos o valor da variável latente O_r segundo

$$O_r | \bullet \sim \text{Bernoulli} \left(\frac{\omega e^{-\lambda_r [1 - P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})]}}{(1 - \omega) + \omega e^{-\lambda_r [1 - P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})]}} \right). \quad (3.15)$$

Para atualizar N_r notamos que, dado $O_r = 0$ gerado pela distribuição acima, $N_r = 0$ com probabilidade 1. Por outro lado, se $O_r = 1$, temos a seguinte distribuição de probabilidades

condicional

$$p(N_r | O_r = 1, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r, \lambda_r) \propto p(\mathbf{n}_r | N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) p(N_r | \lambda_r, O_r = 1) \quad (3.16)$$

$$\begin{aligned} &\propto \mathcal{L}(N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{n}_r) p(N_r | \lambda_r) \\ &\propto \frac{N_r!}{(N_r - n_r)!} P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})^{N_r} \frac{[\lambda_r]^{N_r}}{N_r!} \\ &\propto \frac{[P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) \cdot \lambda_r]^{N_r}}{(N_r - n_r)!}, \end{aligned} \quad (3.17)$$

para $N_r \in \{n_r, n_r + 1, \dots\}$, onde \mathcal{L} é a função de verossimilhança para os dados observados em uma região r apresentada em (2.10) e $P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})$ é dado em (2.8). Logo,

$$(N_r - n_r) | \bullet \sim \text{Poisson} \left\{ P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) \cdot \lambda_r \right\}, \quad (3.18)$$

e a atualização do valor do parâmetro N_r é dada por $z + n_r$, onde z é uma observação de uma variável aleatória Z com distribuição Poisson de parâmetro igual a $P(\mathbf{X}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) \lambda_r$.

Finalmente, dado um valor de N_r atualizado por (3.18), a geração das variáveis latentes $\tilde{\mathbf{n}}_r$ é realizada segundo a distribuição condicional

$$\begin{aligned} p(\tilde{\mathbf{n}}_r | N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{n}_r) &= \frac{p(\mathbf{n}_r, \tilde{\mathbf{n}}_r | N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})}{p(\mathbf{n}_r | N_r, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})} \\ &= \prod_{\mathbf{x} \in \mathcal{X}} \left\{ \frac{n_{r, [\mathbf{x}]}!}{\prod_{u=1}^{f(\mathbf{x})} \prod_{v=l(\mathbf{x}, u)}^T \tilde{n}_{r, [\mathbf{x}]}^{(u, v)}!} \prod_{u=1}^{f(\mathbf{x})} \prod_{v=l(\mathbf{x}, u)}^T \left[\frac{P(\mathbf{X} = \mathbf{x}, U = u, V = u | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})}{P(\mathbf{X} = \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})} \right]^{\tilde{n}_{r, [\mathbf{x}]}^{(u, v)}} \right\}, \end{aligned} \quad (3.19)$$

com a restrição (3.10) imposta sobre a geração das variáveis latentes. Através de algumas manipulações algébricas para obter a expressão acima identificamos que, para um dado histórico $\mathbf{x} \in \mathcal{X}$ com $n_{r, [\mathbf{x}]} > 0$, a distribuição condicional que gera as frequências latentes $\tilde{n}_{r, [\mathbf{x}]}^{(u, v)}$, para todo $1 \leq u \leq f(\mathbf{x})$, $l(\mathbf{x}, u) \leq v \leq T$, é uma distribuição Multinomial com $n_{r, [\mathbf{x}]}$ ensaios e probabilidade de cada evento correspondente proporcional à (2.5). Para $n_{r, [\mathbf{x}]} = 0$, as frequências $\tilde{n}_{r, [\mathbf{x}]}^{(u, v)}$ serão todas iguais à 0, considerando a restrição (3.10). Além disso, as frequências latentes são geradas de maneira independente para cada histórico $\mathbf{x} \in \mathcal{X}$ decorrente do produto das funções de probabilidades multinomiais observadas em (3.19).

Bloco II: atualização de $(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})$

Para o Bloco II, notamos que as distribuições a priori para $(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})$ apresentadas em (3.2) são conjugadas com relação à função de verossimilhança aumentada dada por (3.12) para valores fixos de $(\mathbf{N}, \tilde{\mathbf{n}})$. Assim, a mesma distribuição de probabilidades é obtida a posteriori para cada parâmetro, restando apenas a atualização de seus hiperparâmetros. Além disto, a função de

verossimilhança aumentada é fatorada em funções independentes para $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ e $\boldsymbol{\theta}$ resultando as seguintes distribuições condicionais completas independentes:

$$\phi_t | \bullet \sim \text{Beta}(W_t + b_{1,t}, N_t - W_t + b_{2,t}), \quad t \in \{1, \dots, T-1\}, \quad (3.20)$$

$$\boldsymbol{\beta} | \bullet \sim \text{Dirichlet}(R_1 + a_1, \dots, R_T + a_T), \quad (3.21)$$

$$\boldsymbol{\theta} | \bullet \sim \text{Beta}(\bar{n} + c_1, \bar{N} - \bar{n} + c_2). \quad (3.22)$$

Bloco III: atualização de λ

Para o Bloco III do algoritmo proposto, atualizamos separadamente cada λ_r devido a suposição de independência em (3.7). Logo, a distribuição condicional de λ_r é dada por

$$\begin{aligned} f(\lambda_r | N_r = m, O_r = o, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\omega}) &= c \cdot \text{P}(N_r = m | O_r = o, \lambda_r) \cdot f(\lambda_r | \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\omega}) \\ &= c \left[\frac{e^{-\lambda_r} \lambda_r^m}{m!} \right]^o \frac{1}{\lambda_r \sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} [\log(\lambda_r) - \mu]^2 \right\}, \end{aligned} \quad (3.23)$$

onde c é uma constante que não depende de λ_r . Para atualização de λ_r , sejam λ_r^* o valor atual de λ_r e $\tilde{\lambda}_r$ um valor (proposto) com $\tilde{\lambda}_r = \exp(x)$ onde x é gerado de uma distribuição normal com média $\log(\lambda_r^*)$ e variância conhecida b^2 . Assim, o valor de λ_r é atualizado de λ_r^* para $\tilde{\lambda}_r$ com uma probabilidade de aceitação Metropolis-Hasting dada por

$$\alpha(\lambda_r^*, \tilde{\lambda}_r) = \min \left\{ 1, \frac{f(\tilde{\lambda}_r | N_r = m, O_r = o, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\omega})}{f(\lambda_r^* | N_r = m, O_r = o, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\omega})} \right\}. \quad (3.24)$$

Bloco IV: atualização de $(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\omega})$

Consideremos $\varepsilon_r = \log(\lambda_r)$ com $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_R)$ e $\boldsymbol{\psi} = \boldsymbol{\sigma}^{-2}$. Para atualizar $(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\omega} | \boldsymbol{\varepsilon}, \mathbf{O})$, seja

$$\begin{aligned} f(\boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\omega} | \boldsymbol{\varepsilon}, \mathbf{O}) &\propto \prod_{r \in \mathcal{R}} f(\varepsilon_r | \boldsymbol{\mu}, \boldsymbol{\sigma}) \text{P}(O_r = o_r | \boldsymbol{\omega}) f(\boldsymbol{\mu}) f(\boldsymbol{\sigma}) f(\boldsymbol{\omega}) \\ &\propto \boldsymbol{\psi}^{R/2} \exp \left\{ -\frac{\boldsymbol{\psi}}{2} \sum_{r \in \mathcal{R}} [\varepsilon_r - \boldsymbol{\mu}]^2 \right\} \exp \left\{ -\frac{1}{2\xi_0^2} [\boldsymbol{\mu} - \boldsymbol{\mu}_0]^2 \right\} \boldsymbol{\psi}^{a-1} e^{-b\boldsymbol{\psi}} \times \\ &\quad \times \boldsymbol{\omega}^{\tilde{o}_r} (1 - \boldsymbol{\omega})^{R - \tilde{o}_r} \boldsymbol{\omega}^{g_1 - 1} (1 - \boldsymbol{\omega})^{g_2 - 1} \\ &\propto \exp \left\{ -\frac{1}{2} \left[-2\boldsymbol{\mu} \left(\boldsymbol{\psi} \sum_{r \in \mathcal{R}} \varepsilon_r + \frac{\boldsymbol{\mu}_0}{\xi_0^2} \right) + \boldsymbol{\mu}^2 \left(R\boldsymbol{\psi} + \frac{1}{\xi_0^2} \right) \right] \right\} \boldsymbol{\psi}^{\frac{R}{2} + a - 1} e^{-b\boldsymbol{\psi}} \times \\ &\quad \times \boldsymbol{\omega}^{\tilde{o} + g_1 - 1} (1 - \boldsymbol{\omega})^{R - \tilde{o} + g_2 - 1} \end{aligned}$$

onde $\tilde{o} = \sum_{r \in \mathcal{R}} o_r$. Portanto, as seguintes distribuições a posteriori condicionais são

$$\mu|\bullet \sim \text{Normal}(\kappa\tau, \tau), \quad (3.25)$$

$$\psi|\bullet \sim \text{Gama}\left(\frac{R}{2} + a, \frac{1}{2} \sum_{r \in \mathcal{R}} [\varepsilon_r - \mu]^2 + b\right), \quad (3.26)$$

$$\omega|\bullet \sim \text{Beta}(\tilde{o} + g_1, R - \tilde{o} + g_2), \quad (3.27)$$

$$\text{onde } \kappa = \psi \sum_{r \in \mathcal{R}} \varepsilon_r + \frac{\mu_0}{\xi_0^2} \text{ e } \tau = \left(R\psi + \frac{1}{\xi_0^2}\right)^{-1}.$$

Logo, o algoritmo MCMC pode ser implementado através de um processo iterativo utilizando as distribuições condicionais dadas por (3.15), (3.18), (3.19), (3.20), (3.21), (3.22), (3.23), (3.25), (3.26) e (3.27). Neste trabalho as rotinas foram implementadas no *software* R (R Development Core Team, 2014). Então, a partir do algoritmo MCMC proposto, suponhamos B valores, ξ_1, \dots, ξ_B , gerados da distribuição posteriori de um determinado parâmetro ξ . A esperança (média) a posteriori para este parâmetro pode ser aproximada por

$$\widehat{\mathbb{E}(\xi)} = \frac{1}{B} \sum_{b=1}^B \xi_b, \quad (3.28)$$

enquanto que a estimativa da mediana a posteriori pode ser obtida por ordenamento dos valores, $\xi_{(1)}, \dots, \xi_{(B)}$, com $\xi_{(b)} \leq \xi_{(b+1)}$, e fazendo

$$\widehat{\mathbb{M}(\xi)} = \begin{cases} \xi_{(B+1)/2}, & \text{se } B \text{ for ímpar,} \\ \frac{\xi_{(B/2)} + \xi_{(B/2+1)}}{2}, & \text{se } B \text{ for par.} \end{cases}$$

Estes valores são as estimativas para o parâmetro ξ utilizadas neste trabalho. Elas minimizam as perdas quadrática e absoluta esperadas a posteriori, respectivamente. Adicionalmente, uma estimativa intervalar para ξ pode ser obtida determinando-se os valores de a e b tais que $a \leq b$ e $P(a \leq \xi \leq b | \text{Dados}) \geq 1 - \alpha$, para $\alpha \in (0, 1)$ fixo. A partir dos valores gerados, utilizamos a aproximação para esta probabilidade por

$$\frac{1}{B} \sum_{b=1}^B I(a \leq \xi_b \leq b) \geq 1 - \alpha. \quad (3.29)$$

Dizemos que (a, b) é um intervalo de credibilidade com $100(1 - \alpha)\%$ para ξ . Nesta tese, adotamos intervalos de credibilidade HPD (sigla para *highest posterior density*, em inglês) (CHEN; SHAO, 1999; COLVIN, 2013), encontrando valores de a e b que satisfaçam conjuntamente (3.29) e que minimizem a amplitude do intervalo, $(b - a)$. Para isto, utilizamos a função `hdi` do pacote `HDInterval` (MEREDITH; KRUSCHKE, 2017).

3.3 Estudo de simulação

Nesta seção realizaremos um estudo computacional para verificar a qualidade das estimativas bayesianas para os parâmetros populacionais, a partir da modelagem e do processo

inferencial propostos. O estudo refere-se a dados artificiais de captura-recaptura gerados sob diferentes cenários e à avaliação das propriedades frequentistas dos estimadores de acordo com algumas medidas de desempenho, definidas na sequência.

3.3.1 Medidas de desempenho

Para as medidas de desempenho dos estimadores pontuais consideramos os vieses absoluto (Viés) e relativo (ViésR) e os erros médios absoluto (EAM) e quadrático (EQM). Para os estimadores intervalares consideramos a probabilidade de cobertura (PC) e a amplitude média (AM) do intervalo. Neste contexto, métodos de simulação Monte Carlo podem ser utilizados para aproximar tais quantidades. Por exemplo, suponhamos que o experimento aleatório seja repetido M vezes de maneira independente fornecendo, assim, M amostras. Sejam $\hat{\xi}_1, \dots, \hat{\xi}_M$ as respectivas estimativas pontuais para o parâmetro ξ relativas a um determinado estimador. Primeiramente, se todas as amostras foram geradas a partir de um mesmo valor de ξ , digamos ξ^* , então os vieses e os erros médios podem ser aproximados, respectivamente, por

$$\begin{aligned} \text{Viés} &= \frac{1}{M} \sum_{m=1}^M \{ \hat{\xi}_m - \xi^* \}, & \text{ViésR} &= \frac{1}{M} \sum_{m=1}^M \left\{ \frac{\hat{\xi}_m - \xi^*}{\xi^*} \right\}, & \text{EQM} &= \frac{1}{M} \sum_{m=1}^M \{ \hat{\xi}_m - \xi^* \}^2 \quad \text{e} \\ \text{EAM} &= \frac{1}{M} \sum_{m=1}^M | \hat{\xi}_m - \xi^* |. \end{aligned} \quad (3.30)$$

Valores próximos de zero para estas medidas indicam boa precisão das estimativas produzidas pelo estimador. Além disto, se um estimador intervalar for adotado para o parâmetro ξ , então podemos determinar sua probabilidade de cobertura e amplitude média do intervalo por

$$\text{PC} = \frac{1}{M} \sum_{m=1}^M I(\hat{\xi}_m^L \leq \xi^* \leq \hat{\xi}_m^U) \cdot 100\% \quad \text{e} \quad \text{AM} = \frac{1}{M} \sum_{m=1}^M (\hat{\xi}_m^U - \hat{\xi}_m^L), \quad (3.31)$$

respectivamente, onde $(\hat{\xi}_m^L, \hat{\xi}_m^U)$ é a estimativa intervalar para ξ obtida da m -ésima amostra. Se os intervalos forem obtidos utilizando $100(1 - \alpha)\%$ de credibilidade, então espera-se que o valor de PC esteja próximo a esta porcentagem. Adicionalmente, espera-se que a amplitude média do intervalo seja a menor possível dada uma probabilidade de cobertura próxima ao nível de credibilidade adotado. Se cada amostra m for gerada a partir de um valor diferente de ξ , dito ξ_m^* , então substituiremos o valor de ξ^* pelo específico valor verdadeiro do parâmetro, ξ_m^* , nas fórmulas acima para obter uma aproximação destas medidas de desempenho neste caso.

3.3.2 Geração dos dados

Para a geração dos dados consideramos 18 diferentes cenários apresentados na Tabela 1. Estes cenários foram determinados levando-se em conta diferentes valores para o número esperado da superpopulação global ($\mathbb{E}[N] = \mathbb{E}[N_1 + \dots + N_R]$), para a proporção de inflação de zero ($1 - \omega$) na distribuição Poisson Zero-Inflacionada e para a probabilidade de captura

(θ). Especificamente, verificamos o efeito do aumento do número de animais na população e da probabilidade de captura nas medidas de desempenho adotadas, bem como o impacto da proporção de regiões zero-inflacionadas nas inferências a posteriori dos parâmetros. Ademais, fixamos para todos os cenários $R = 50$ regiões de amostragem, $T = 6$ ocasiões de captura e $\sigma = 1$ na estrutura hierárquica dos tamanhos populacionais.

Tabela 1 – Cenários utilizados para o primeiro estudo de simulação.

Cenário	$\mathbb{E}(N)$	ω	θ	Cenário	$\mathbb{E}(N)$	ω	θ	Cenário	$\mathbb{E}(N)$	ω	θ
1	1000	0,6	0,20	7	2000	0,6	0,20	13	3000	0,6	0,20
2	1000	0,6	0,30	8	2000	0,6	0,30	14	3000	0,6	0,30
3	1000	0,8	0,20	9	2000	0,8	0,20	15	3000	0,8	0,20
4	1000	0,8	0,30	10	2000	0,8	0,30	16	3000	0,8	0,30
5	1000	1,0	0,20	11	2000	1,0	0,20	17	3000	1,0	0,20
6	1000	1,0	0,30	12	2000	1,0	0,30	18	3000	1,0	0,30

Para obter as medidas de desempenho, $M = 200$ amostras foram geradas de cada cenário. A fim de propiciar um maior número de cenários para estudo, as probabilidades de nascimento e sobrevivência foram simuladas aleatoriamente em cada uma das $M = 200$ amostras. Especificamente, em cada amostra $m = 1, \dots, 200$, geramos $\phi = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)$ por

$$\begin{aligned} \phi_1 &\sim \text{Beta}(146, 3; 62, 7), & \phi_2 &\sim \text{Beta}(139, 875; 46, 625), & \phi_3 &\sim \text{Beta}(95, 6; 143, 4), \\ \phi_4 &\sim \text{Beta}(80, 1; 8, 9) & \text{e} & & \phi_5 &\sim \text{Beta}(127, 2; 31, 8), \end{aligned}$$

resultando valores esperados $\mathbb{E}(\phi_1) = 0,7$, $\mathbb{E}(\phi_2) = 0,75$, $\mathbb{E}(\phi_3) = 0,4$, $\mathbb{E}(\phi_4) = 0,90$ e $\mathbb{E}(\phi_5) = 0,8$, com variância comum igual a 0,001, respectivamente. Estes valores esperados foram escolhidos com a finalidade de verificar o comportamento do estimador em períodos de baixa e alta taxas de sobrevivência. Da mesma forma, geramos para cada amostra um diferente vetor de probabilidades de recrutamento definido por

$$\beta \sim \text{Dirichlet}(40, 15, 10, 20, 10, 5),$$

condizente com os valores esperados $\mathbb{E}(\beta_1) = 0,4$, $\mathbb{E}(\beta_2) = 0,15$, $\mathbb{E}(\beta_3) = 0,1$, $\mathbb{E}(\beta_4) = 0,2$, $\mathbb{E}(\beta_5) = 0,1$ e $\mathbb{E}(\beta_6) = 0,05$.

Para obtenção de uma amostra de CR, consideramos os passos definidos a seguir.

1. Para cada $r \in \mathcal{R}$, gere $o_r \sim \text{Bernoulli}(\omega)$;
2. Obtenha $\mu = \log \{ \mathbb{E}(N) / \sum_{r \in \mathcal{R}} o_r \} - \sigma^2 / 2$;
3. Para cada $r \in \mathcal{R}$:

a) Gere

$$x \sim \text{Normal}(\mu, \sigma^2);$$

$$N_r \sim \text{Poisson}(\exp(x) \cdot o_r);$$

b) Se $N_r = 0$, faça: $n_{r,[\mathbf{x}]} = 0; \forall \mathbf{x} \in \mathcal{X}$;

c) Se $N_r > 0$ faça:

i. Para cada $i \in \mathcal{N}_r = \{1, \dots, N_r\}$, gere $\mathbf{x}_i \sim (2.8)$;

ii. Para cada $\mathbf{x} \in \mathcal{X}$, faça: $n_{r,[\mathbf{x}]} = \sum_{i \in \mathcal{N}_r} (\mathbf{x}_i = \mathbf{x})$;

iii. Obtenha $\mathbf{n}_r = \{n_{r,[\mathbf{x}]}; \mathbf{x} \in \mathcal{X}_0\}$;

4. Obtenha $\mathbf{n} = \{\mathbf{n}_r; r \in \mathcal{R}\}$.

3.3.3 Resultados numéricos

Para obter as inferências bayesianas consideramos as seguintes distribuições a priori:

$$\phi_t \sim \text{Beta}(1, 1), \quad t \in \{1, 2, \dots, T-1\},$$

$$\theta \sim \text{Beta}(1, 1),$$

$$\boldsymbol{\beta} \sim \text{Dirichlet}(1, 1, 1, 1, 1, 1),$$

$$\mu \sim \text{Normal}(0, 10^2),$$

$$\sigma^{-2} \sim \text{Gama}(0, 1; 0, 1),$$

$$\omega \sim \text{Beta}(1, 1).$$

Tais valores dos hiperparâmetros foram escolhidos para que as distribuições a priori sejam pouco informativas. Para cada cenário aplicamos o algoritmo MCMC com 100.000 iterações, descartamos as primeiras 50.000 iterações como ‘aquecimento’ da cadeia e, com base na amostra restante, utilizamos saltos de tamanho 10 resultando uma amostra final de 5.000 valores gerados para cada parâmetro. A partir das simulações a posteriori determinamos as estimativas pontuais e intervalares para os parâmetro em cada conjunto de dados. As estimativas intervalares foram determinadas adotando um nível de credibilidade de 95%. Utilizamos o valor de $b^2 = 0,5$ na geração de candidatos das taxas médias (λ_r 's) para o cálculo da probabilidade de aceitação definida em (3.24), o que apresentou taxas de aceitação variando de 35% a 50% para $\mathbb{E}(N) = 1000$, de 30% a 40% para $\mathbb{E}(N) = 2000$ e de 25% a 40% para $\mathbb{E}(N) = 3000$.

3.3.4 Exemplo

Para fins ilustrativos, descrevemos a seguir algumas inferências a posteriori obtidas com base em uma amostra gerada segundo o Cenário 8. Primeiramente, apresentamos na Tabela 2 os valores médios, medianos, desvios-padrões (d.p.) e intervalos HPD com 95% de credibilidade a

posteriori para os números de animais vivos nas ocasiões de amostragem, N_1, N_2, \dots, N_6 . Estas medidas foram calculadas utilizando os valores gerados para os dados aumentados, $\tilde{\mathbf{n}}$, a partir da relação

$$N_t = \sum_{r \in \mathcal{R}} \sum_{\mathbf{x} \in \mathcal{X}} \sum_{u=1}^t \sum_{v=t}^T \tilde{n}_{r, [\mathbf{x}]}^{(u,v)},$$

como previamente definida na página 40. Observamos que as estimativas pontuais para estas abundâncias estão próximas dos números verdadeiros de animais vivos nas ocasiões, exibidos na coluna N_t^* , e que os intervalos de credibilidade contém os verdadeiros valores. A Figura 4 exhibe as aproximações das funções de probabilidades a posteriori destas quantidades utilizando os valores gerados da amostra MCMC onde podemos observar distribuições aproximadamente simétricas, resultando em valores da média e mediana a posteriori próximos entre si.

Tabela 2 – Resumos a posteriori para os tamanhos populacionais em cada ocasião de amostragem para uma amostra gerada do Cenário 8.

Abundância	N_t^*	média	mediana	d.p	$IC_b(95\%)$
N_1	550	542,34	538	56,27	[436, 653]
N_2	688	619,78	616	52,67	[518, 719]
N_3	772	781,59	778	65,94	[663, 917]
N_4	638	591,44	588	50,78	[494, 691]
N_5	673	667,36	667	50,28	[564, 758]
N_6	622	607,71	602	60,32	[500, 729]

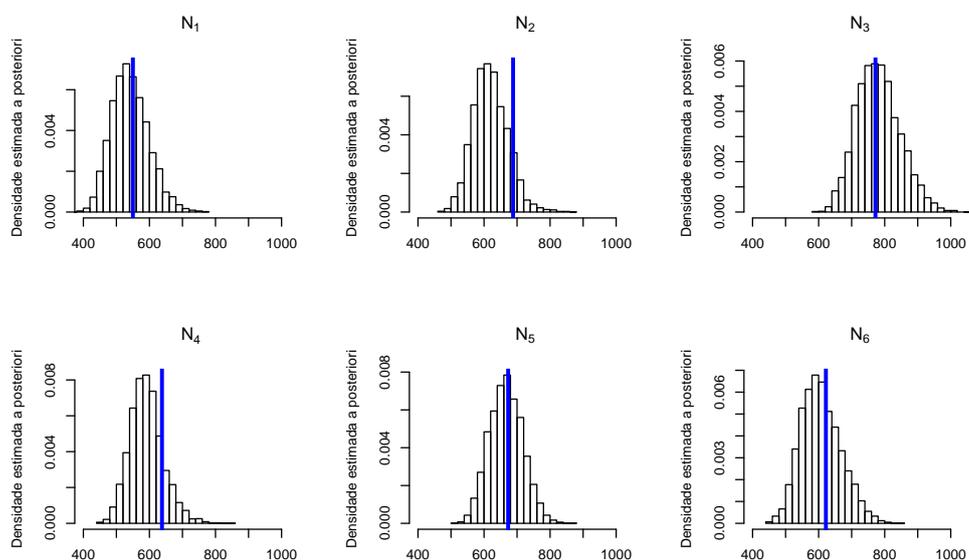


Figura 4 – Aproximações das funções de probabilidades a posteriori para os tamanhos populacionais em cada ocasião de amostragem. Linhas verticais azuis indicam o verdadeiro valor do parâmetro.

Exibimos na Figura 5 os números médios estimados de animais vivos em cada ocasião

por região, definidos por

$$N_{t,r} = \sum_{\mathbf{x} \in \mathcal{X}} \sum_{u=1}^t \sum_{v=t}^T \tilde{n}_{r,[\mathbf{x}]}^{(u,v)},$$

para $r \in \mathcal{R}$ e $t \in \mathcal{T}$, onde a primeira imagem representa os números verdadeiros destas quantidades e a segunda as estimativas utilizando a modelagem proposta neste capítulo. Para comparação recalculamos estas estimativas a partir de uma estrutura hierárquica para as superpopulações locais (N_r 's) utilizando uma distribuição Poisson sem inflação de zeros, isto é, $\omega = 1$ fixo. Para simplicidade na exposição dos resultados, apresentamos 10 ‘zonas’ territoriais, cada uma representando uma união de 5 regiões de amostragem. Notamos que as duas propostas de priori produziram estimativas a posteriori próximas da verdadeira distribuição espaço-temporal dos animais desta população simulada. Por exemplo, a maior abundância foi observada na sexta ‘zona’ na terceira ocasião de amostragem, com valor verdadeiro 171 animais, e tendo as respectivas estimativas de 167,57 e 166,69. Observamos que as duas estruturas a priori apresentaram resultados a posteriori semelhantes, embora as estimativas quando não levamos em consideração a inflação de zero foram, em geral, inferiores quando comparadas às estimativas do outro modelo.

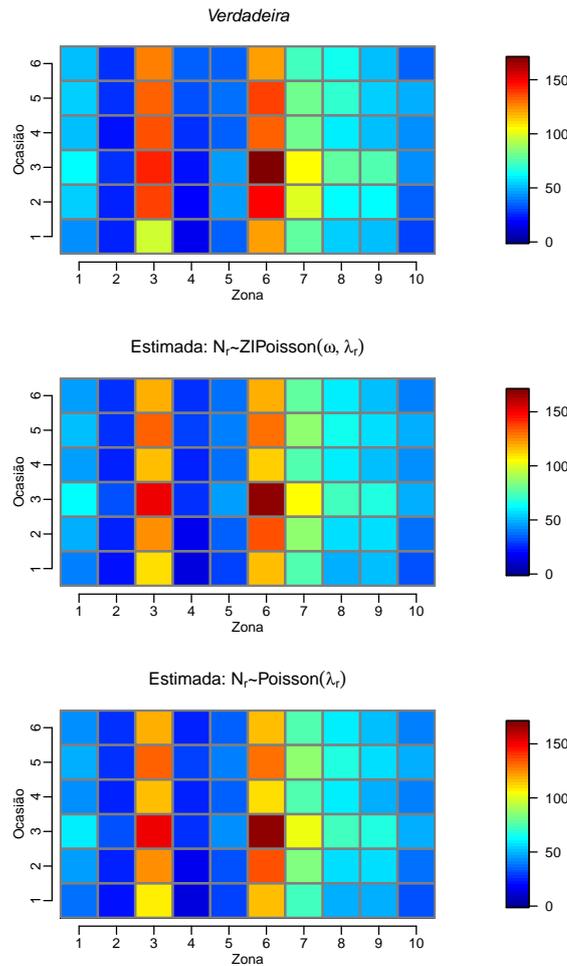


Figura 5 – Valores médios estimados a posteriori para os tamanhos populacionais em cada ocasião e ‘zona’ de amostragem.

Apesar das duas estruturas hierárquicas não apresentarem diferenças significativas nas inferências considerando as regiões de amostragem, o mesmo não ocorre no processo preditivo de abundâncias para regiões não amostradas. Com efeito, o método de predição depende dos hiperparâmetros (μ, σ) , cujas estimativas são apresentadas na Tabela 3 para os dois casos considerados. Observamos valores médios estimados para μ de 1,1830 considerando o modelo sem inflação e 3,8426 para o outro caso, sendo que o valor verdadeiro deste parâmetro é 3,8051. Para o parâmetro σ , fixado em 1 para a geração do dados, obtivemos estimativas médias a posteriori 3,1928 e 0,6495, respectivamente. Como esperado, a inclusão do número excessivo de zeros na distribuição de Poisson não inflacionada fez com que seu valor esperado (em escala logarítmica) apresentasse média subestimada e uma alta variabilidade. Por outro lado, considerando a distribuição Poisson Zero-Inflacionada, estimativas satisfatórias foram observadas para estes parâmetros.

Tabela 3 – Resumos a posteriori para os hiperparâmetros (μ, σ) para uma amostra gerada do Cenário 8.

Priori	Parâmetro	Média	Mediana	d.p	IC _b (95%)
ZIPoisson	μ	3,8426	3,8418	0,1358	(3,5644; 4,0960)
	σ	0,6495	0,6375	0,1080	(0,4701; 0,8688)
	ω	0,5398	0,5399	0,0686	(0,4096; 0,6763)
Poisson	μ	1,1830	1,2181	0,5578	(0,0313; 2,2156)
	σ	3,1928	3,1242	0,5352	(2,2220; 4,2077)

Para comparação, calculamos as estimativas destes parâmetros utilizando uma amostra gerada a partir do Cenário 12, isto é, dados sem inflação de zeros. Nestes cenários, no entanto, poucas regiões apresentam nenhuma captura de animais ($n_r = 0$) o que, de fato, não caracteriza dados inflacionados de zeros. Como exemplo, apresentamos na Tabela 4 as estimativas a posteriori para uma amostra deste cenário contendo uma região com $n_r = 0$ e o valor verdadeiro de μ igual a 3,1889. Observamos que as estimativas a posteriori sob as duas estruturas a priori consideradas foram próximas entre elas e dos valores verdadeiros destes hiperparâmetros. Assim, consideramos que a distribuição Poisson Zero-Inflacionada pode ser ajustada sem perda significativa da eficiência na estimação mesmo em situações com dados sem inflação de zeros, embora seja natural que ela seja adotada apenas em situações com excessos de regiões sem animais capturados.

Tabela 4 – Resumos a posteriori para os hiperparâmetros (μ, σ) para uma amostra gerada do Cenário 12.

Priori	Parâmetro	Média	Mediana	d.p	IC _b (95%)
ZIPoisson	μ	3,2834	3,2832	0,1225	(3,0523; 3,5354)
	σ	0,7633	0,7551	0,0982	(0,5793; 0,9530)
	ω	0,9643	0,9704	0,0263	(0,9131; 0,9999)
Poisson	μ	3,2436	3,2481	0,1315	(2,9881; 3,4979)
	σ	0,8108	0,8039	0,1017	(0,6272; 1,0183)

Consideramos na sequência um problema de predição de abundância populacional. Seja \tilde{N}_t o número de animais durante a t -ésima ocasião de captura em uma região não amostrada. Dada a observação da amostra $\mathbf{n} = \{\mathbf{n}_r; r \in \mathcal{R}\}$, temos interesse em determinar a distribuição a posteriori preditiva de \tilde{N}_t , dado \mathbf{n} . Para isto, suponhamos que o tamanho da superpopulação desta região, denotado por \tilde{N} , tem distribuição predita marginal obtida pela marginalização da seguinte estrutura hierárquica:

$$\begin{aligned}\mu, \sigma | \mathbf{n} &\sim f(\mu, \sigma | \mathbf{n}), \\ \log[\tilde{\lambda}] | \mu, \sigma &\sim \text{Normal}(\mu, \sigma^2), \\ \tilde{N} | \tilde{\lambda} &\sim \text{Poisson}(\tilde{\lambda}),\end{aligned}\tag{3.32}$$

sendo condicionalmente independente das demais regiões, dado (μ, σ) .

Dado \tilde{N} , consideramos que os números de recrutamentos desta região têm a seguinte relação:

$$\begin{aligned}\boldsymbol{\beta} | \mathbf{n} &\sim f(\boldsymbol{\beta} | \mathbf{n}), \\ (\tilde{R}_1, \dots, \tilde{R}_T) | \tilde{N}, \boldsymbol{\beta} &\sim \text{Multinomial}(\tilde{N}, \boldsymbol{\beta}).\end{aligned}\tag{3.33}$$

Por fim, as quantidades $\tilde{N}_1, \dots, \tilde{N}_T$ são definidas por

$$\begin{aligned}\tilde{N}_1 &= \tilde{R}_1, \\ \boldsymbol{\phi} | \mathbf{n} &\sim f(\boldsymbol{\phi} | \mathbf{n}), \\ \tilde{D}_t | \boldsymbol{\phi}_t, \tilde{N}_t &\sim \text{Binomial}(\tilde{N}_t, \boldsymbol{\phi}_t), \\ \tilde{N}_{t+1} &= \tilde{R}_{t+1} + \tilde{N}_t - \tilde{D}_t, \quad t \in \{1, 2, \dots, T-1\}.\end{aligned}\tag{3.34}$$

Logo, como a geração de valores em (3.32), (3.33) e (3.34) pode ser feita utilizando o algoritmo MCMC desenvolvido na Seção 3.2.2, torna-se possível obter uma amostra da distribuição preditiva de \tilde{N}_t , dado \mathbf{n} , incorporando gerações estocásticas das demais distribuições condicionais acima em cada iteração do algoritmo MCMC. Utilizando o procedimento de predição descrito, geramos os valores para \tilde{N}_t em uma região não amostrada. Para comparação, determinamos os valores esperados para esta região dado por $\mathbb{E}(N_t)$, onde

$$\begin{aligned}\mathbb{E}(N_1) &= \mathbb{E}(N) \cdot \beta_1, \\ \mathbb{E}(N_t) &= \mathbb{E}(N_{t-1}) \phi_{t-1} + \mathbb{E}(N) \cdot \beta_t, \quad t \in \{2, \dots, T\}\end{aligned}$$

com $\mathbb{E}(N) = \mathbb{E}[\mathbb{E}(N | \lambda)] = \mathbb{E}(\lambda) = \exp\{\mu + \sigma^2/2\}$. Os resumos descritivos destas predições são apresentadas na Tabela 5. Consideramos os valores preditos satisfatórios a partir do uso da estrutura a priori via distribuição zero-inflacionada. Por outro lado, o processo de predição para a estrutura sem inflação de zeros acarretou em estimativas questionáveis, no sentido de alta variabilidade e amplitude do intervalo de credibilidade. Apesar dos valores de N_t 's tenham sido

subestimados em todas as ocasiões, acreditamos que este fato tenha ocorrido para esta amostra em específico e não decorrente do procedimento de predição adotado.

Esta diferença entre os resultados obtidos pelos dois processos de predição pode ser observada através das distribuições preditivas de \tilde{N} , apresentadas na Figura 6. A distribuição de \tilde{N} via dados zero-inflacionados tem comportamento regular, enquanto que a distribuição para o outro modelo apresentou alta variabilidade e valores gerados muito elevados. Nesta figura delimitamos a amplitude máxima do eixo horizontal em 800, embora tenhamos observados valores para \tilde{N} na imagem à direita de até 44.843.635.

Tabela 5 – Resumos descritivos dos valores preditos para as abundâncias \tilde{N}_t .

Parâmetro	$\mathbb{E}(N_t)$	ZIPoisson			Poisson		
		média	mediana	$IC_b(95\%)$	média	mediana	$IC_b(95\%)$
\tilde{N}_1	25,95	20,95	17	[1, 51]	274,10	1	[0, 218]
\tilde{N}_2	33,30	23,87	19	[1, 58]	318,08	1	[0, 238]
\tilde{N}_3	35,77	30,13	24	[1, 74]	408,05	2	[0, 315]
\tilde{N}_4	29,05	22,83	18	[0, 56]	314,84	1	[0, 248]
\tilde{N}_5	30,65	25,70	20	[1, 63]	353,81	2	[0, 276]
\tilde{N}_6	29,26	23,33	18	[1, 57]	308,45	1	[0, 237]

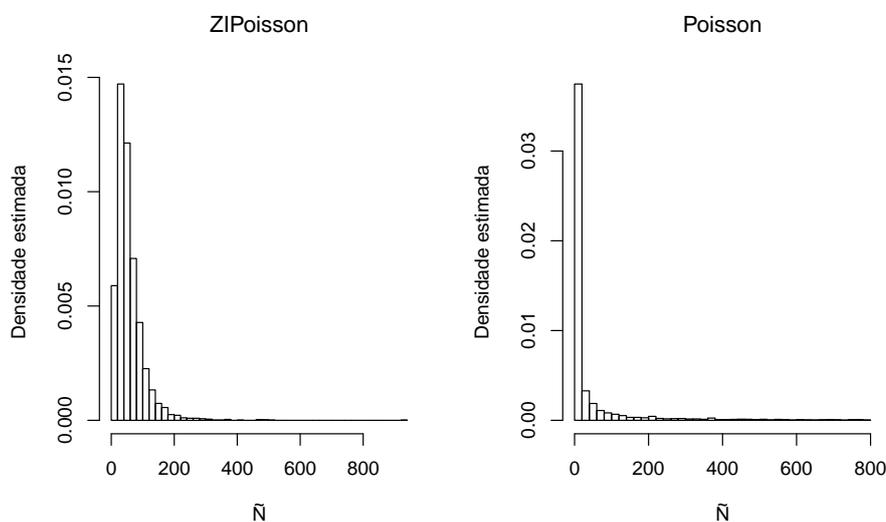


Figura 6 – Aproximações das funções de probabilidades predita do tamanho da superpopulação em uma região não amostrada.

3.3.5 Qualidade das estimativas

Em sequência, determinamos as medidas de desempenho apresentadas na Seção 3.3.1 para cada cenário, dadas as estimativas de cada parâmetro com base em $M = 200$ gerações de dados de captura-recaptura. Primeiramente, exibimos na Tabela 6 as medidas de desempenho

para os estimadores pontuais, média ($\mathbb{E}(N)$) e mediana ($\mathbb{M}(N)$), e para o estimador intervalar do parâmetro N . Como os resultados mostram, os viesés e erros médios foram maiores quando $\theta = 0,2$ em comparação com $\theta = 0,3$ para ambos os estimadores. Isto decorre do fato que, com maior probabilidade de captura, mais dados amostrais (animais capturados) são obtidos o que torna o processo inferencial mais robusto. Para ambos estimadores pontuais os viesés relativos foram próximos de zero, especificamente menores que 1% em valores absolutos, indicando boa precisão das estimativas produzidas. Ademais, houve uma redução de aproximadamente 40% na amplitude média dos intervalos de credibilidade com o aumento do valor da probabilidade de captura. Além disto, os valores para a probabilidade de cobertura (PC) foram próximos ao nível de credibilidade de 95%, principalmente para maiores valores verdadeiros do tamanho populacional. Aparentemente, a proporção de inflação de zeros ($1 - \omega$) não exerceu influência significativa no desempenho dos estimadores para a abundância populacional.

Tabela 6 – Medidas de desempenho dos estimadores bayesianos para o parâmetro N .

Cenário	Viés		ViésR(%)		EAM		$\sqrt{\text{EQM}}$		AM	PC
	$\mathbb{E}(N)$	$\mathbb{M}(N)$	$\mathbb{E}(N)$	$\mathbb{M}(N)$	$\mathbb{E}(N)$	$\mathbb{M}(N)$	$\mathbb{E}(N)$	$\mathbb{M}(N)$		
1	7,26	0,91	0,75	0,07	72,55	71,73	98,35	96,74	356,12	94,00
2	-1,19	-4,17	-0,10	-0,42	47,41	47,35	63,01	62,61	219,02	94,00
3	7,75	1,24	0,76	0,09	71,46	70,31	94,28	92,76	357,98	93,50
4	-0,92	-3,81	-0,07	-0,37	46,80	46,91	60,29	60,12	220,65	94,50
5	7,45	0,82	0,86	0,17	75,78	74,74	100,29	98,44	358,94	93,50
6	0,01	-2,91	-0,06	-0,36	46,38	46,59	63,52	63,37	220,94	92,50
7	16,17	9,46	0,72	0,37	102,06	101,56	137,13	135,44	503,11	93,00
8	1,26	-1,51	0,07	-0,07	66,62	66,44	87,38	87,15	310,13	93,50
9	11,19	4,10	0,51	0,14	102,80	102,11	134,26	132,47	506,67	94,00
10	-0,73	-3,48	-0,03	-0,18	66,64	66,89	87,03	87,18	310,80	92,00
11	10,07	3,72	0,46	0,13	101,84	101,50	132,61	131,66	503,45	92,50
12	-1,26	-4,49	-0,10	-0,26	65,28	65,34	86,35	86,03	313,23	92,00
13	7,82	1,53	0,21	-0,01	109,81	109,52	142,54	141,50	615,55	96,50
14	3,48	0,29	0,14	0,03	82,77	82,73	103,08	102,90	378,74	94,00
15	11,80	6,33	0,31	0,12	113,24	113,70	151,23	151,32	613,80	94,50
16	6,78	3,90	0,19	0,09	78,04	77,82	101,75	101,31	383,13	95,00
17	3,96	-2,09	0,07	-0,14	107,50	107,66	144,75	144,90	613,35	95,50
18	-0,70	-3,88	-0,02	-0,13	79,24	79,16	100,42	100,30	383,56	95,00

Para os demais parâmetros, consideramos a média a posteriori como estimador pontual na apresentação das medidas de desempenho, e ressaltamos que o uso da mediana como estimador apresentou resultados similares aos discutidos a seguir.

Na Tabela 7 são exibidas as medidas de desempenho dos estimadores para a probabilidade de captura. Observamos que o aumento do número esperado da população ($\mathbb{E}(N)$) produziu um ganho considerável na precisão das estimativas, isto é, diminuição dos viesés, erros médios e amplitude média do intervalo de credibilidade. Em todos os cenários, a probabilidade de cobertura do estimador intervalar esteve próximo ao nível nominal adotado de 95%. Novamente, não há indícios de que o valor do parâmetro ω influencia as estimativas da probabilidade de

captura.

Tabela 7 – Medidas de desempenho dos estimadores bayesianos para o parâmetro θ .

Cenário	Viés	ViésR(%)	EAM	EQM	AM	PC
1	0,0107	5,3451	0,0235	0,0009	0,1150	93,5
2	0,0064	2,1228	0,0213	0,0007	0,1112	96,0
3	0,0104	5,1843	0,0238	0,0009	0,1131	92,5
4	0,0066	2,1933	0,0218	0,0007	0,1105	96,5
5	0,0101	5,0439	0,0237	0,0009	0,1138	93,5
6	0,0061	2,0347	0,0217	0,0007	0,1098	96,0
7	0,0036	1,8228	0,0169	0,0005	0,0827	94,0
8	0,0016	0,5469	0,0168	0,0004	0,0813	95,0
9	0,0035	1,7742	0,0166	0,0004	0,0815	94,5
10	0,0018	0,6007	0,0168	0,0004	0,0809	95,0
11	0,0038	1,9016	0,0165	0,0004	0,0822	95,0
12	0,0023	0,7758	0,0168	0,0004	0,0807	95,5
13	0,0017	0,8643	0,0134	0,0003	0,0691	97,0
14	0,0007	0,2187	0,0153	0,0004	0,0671	93,5
15	0,0010	0,5197	0,0135	0,0003	0,0684	94,5
16	0,0002	0,0508	0,0147	0,0004	0,0675	92,0
17	0,0019	0,9674	0,0133	0,0003	0,0687	95,5
18	0,0009	0,3138	0,0150	0,0004	0,0669	94,5

Na Tabela 8 são apresentadas as medidas de desempenho dos estimadores pontual e intervalar para o vetor de probabilidades de recrutamento, β . Para o estimador pontual notamos que vieses e erros absolutos médios mais elevados estiveram presentes nos cenários com valores mais baixos do tamanho populacional e da probabilidade de captura, assim como observado anteriormente em relação aos outros parâmetros. Para o estimador intervalar, suas medidas de desempenho apresentaram comportamento esperado com o aumento da probabilidade de captura e tamanho populacional no sentido de diminuição da amplitude do intervalo, embora a probabilidade de cobertura em alguns cenários ficou abaixo de 90%.

Na Tabela 9 apresentamos as medidas de desempenho dos estimadores para o vetor de probabilidades de sobrevivência, ϕ . Notamos um comportamento de subestimação de ϕ_t quando seu valor esperado na geração dos dados é superior à 0,5 e sobreestimação no caso inferior à 0,5. Este último caso dá-se especificamente para o parâmetro ϕ_3 com $\mathbb{E}(\phi_3) = 0,4$. No entanto, consideramos que estes vieses foram relativamente baixos em todos os cenários. A probabilidade de cobertura para este vetor, em geral, ficou próxima ou acima do nível nominal adotado na construção do intervalo HPD.

Tabela 8 – Medidas de desempenho dos estimadores bayesianos para o vetor de parâmetro β .

Cenário	β_1				β_2				β_3			
	Viés	EAM	AM	PC	Viés	EAM	AM	PC	Viés	EAM	AM	PC
1	-0,0150	0,0419	0,1967	91,50	0,0013	0,0391	0,2014	92,00	0,0158	0,0341	0,1776	95,00
2	-0,0035	0,0332	0,1580	94,50	-0,0010	0,0339	0,1631	91,50	0,0060	0,0303	0,1375	91,00
3	-0,0143	0,0402	0,1934	91,50	-0,0020	0,0397	0,1989	92,50	0,0168	0,0362	0,1763	93,50
4	-0,0045	0,0314	0,1567	96,00	-0,0006	0,0338	0,1641	92,00	0,0066	0,0304	0,1386	90,50
5	-0,0136	0,0425	0,1956	90,00	-0,0012	0,0412	0,1999	92,00	0,0173	0,0370	0,1751	94,50
6	-0,0031	0,0330	0,1571	94,00	-0,0018	0,0338	0,1635	93,00	0,0072	0,0305	0,1385	89,50
7	-0,0018	0,0320	0,1535	92,00	-0,0059	0,0346	0,1622	92,50	0,0075	0,0286	0,1452	90,50
8	0,0032	0,0249	0,1202	93,50	-0,0034	0,0255	0,1245	93,50	0,0011	0,0222	0,1100	91,50
9	-0,0022	0,0328	0,1524	93,00	-0,0055	0,0342	0,1602	90,50	0,0068	0,0297	0,1423	92,50
10	0,0022	0,0253	0,1190	93,00	-0,0019	0,0246	0,1233	93,00	-0,0005	0,0223	0,1080	90,50
11	-0,0026	0,0332	0,1525	92,00	-0,0051	0,0354	0,1602	88,00	0,0070	0,0283	0,1444	93,50
12	0,0013	0,0251	0,1185	92,00	-0,0007	0,0251	0,1221	92,00	-0,0004	0,0221	0,1084	90,50
13	0,0007	0,0289	0,1325	92,50	-0,0050	0,0305	0,1400	93,50	0,0037	0,0248	0,1260	94,50
14	0,0014	0,0218	0,0988	93,50	-0,0002	0,0212	0,1032	93,50	-0,0008	0,0190	0,0926	91,50
15	0,0017	0,0287	0,1320	94,00	-0,0067	0,0308	0,1392	91,50	0,0041	0,0247	0,1257	94,00
16	0,0018	0,0216	0,0995	91,50	-0,0013	0,0209	0,1026	94,00	-0,0002	0,0192	0,0919	93,00
17	-0,0002	0,0281	0,1310	93,50	-0,0055	0,0299	0,1392	93,50	0,0044	0,0257	0,1251	94,50
18	0,0014	0,0212	0,0987	92,00	-0,0013	0,0212	0,1025	93,00	-0,0004	0,0194	0,0916	92,00
β_4												
1	-0,0252	0,0366	0,1776	91,00	0,0071	0,0310	0,1625	95,00	0,0160	0,0274	0,1256	97,50
2	-0,0091	0,0241	0,1341	94,50	0,0009	0,0241	0,1316	95,50	0,0068	0,0212	0,0983	95,00
3	-0,0249	0,0356	0,1751	92,00	0,0092	0,0309	0,1631	95,50	0,0151	0,0263	0,1251	98,00
4	-0,0092	0,0244	0,1326	94,00	0,0022	0,0245	0,1312	95,50	0,0055	0,0214	0,0971	91,50
5	-0,0248	0,0370	0,1771	93,00	0,0079	0,0300	0,1630	96,00	0,0143	0,0259	0,1235	97,50
6	-0,0088	0,0243	0,1332	95,50	0,0004	0,0239	0,1312	94,50	0,0061	0,0207	0,0978	92,00
7	-0,0123	0,0277	0,1358	94,50	0,0034	0,0250	0,1328	96,00	0,0091	0,0228	0,1048	96,50
8	-0,0031	0,0186	0,0992	97,50	-0,0008	0,0207	0,1033	95,50	0,0031	0,0177	0,0813	93,50
9	-0,0121	0,0275	0,1340	94,50	0,0039	0,0250	0,1315	94,00	0,0092	0,0213	0,1048	98,50
10	-0,0034	0,0191	0,0983	95,00	-0,0003	0,0206	0,1030	94,50	0,0038	0,0172	0,0824	95,50
11	-0,0128	0,0271	0,1344	94,00	0,0042	0,0257	0,1323	96,50	0,0093	0,0214	0,1056	97,00
12	-0,0032	0,0182	0,0982	95,50	-0,0009	0,0202	0,1025	93,50	0,0039	0,0169	0,0815	96,00
13	-0,0063	0,0223	0,1138	93,00	0,0022	0,0229	0,1165	94,00	0,0046	0,0184	0,0935	97,50
14	-0,0008	0,0167	0,0813	95,00	-0,0006	0,0181	0,0870	91,00	0,0009	0,0148	0,0705	94,00
15	-0,0060	0,0215	0,1131	95,00	0,0015	0,0217	0,1161	97,00	0,0054	0,0182	0,0931	97,00
16	-0,0004	0,0163	0,0819	96,50	-0,0008	0,0177	0,0875	91,00	0,0009	0,0146	0,0711	95,50
17	-0,0059	0,0216	0,1125	95,50	0,0025	0,0221	0,1152	96,00	0,0048	0,0191	0,0930	98,00
18	-0,0009	0,0165	0,0812	97,00	-0,0001	0,0175	0,0872	92,50	0,0013	0,0150	0,0712	93,0

Tabela 9 – Medidas de desempenho dos estimadores bayesianos para o vetor de parâmetro ϕ .

Cenário	ϕ_1				ϕ_2				ϕ_3			
	Viés	EAM	AM	PC	Viés	EAM	AM	PC	Viés	EAM	AM	PC
1	-0,0039	0,0879	0,4441	92,50	-0,0381	0,0838	0,4173	94,50	0,0591	0,0841	0,3554	93,00
2	0,0056	0,0630	0,3423	93,00	-0,0089	0,0654	0,3232	94,50	0,0259	0,0497	0,2306	95,00
3	0,0014	0,0865	0,4396	93,50	-0,0382	0,0819	0,4148	94,00	0,0599	0,0857	0,3499	92,00
4	0,0032	0,0661	0,3435	93,00	-0,0090	0,0672	0,3211	93,00	0,0272	0,0496	0,2294	94,00
5	-0,0011	0,0913	0,4375	92,50	-0,0380	0,0803	0,4145	95,50	0,0570	0,0817	0,3488	92,50
6	0,0066	0,0670	0,3427	92,00	-0,0101	0,0643	0,3228	93,00	0,0249	0,0477	0,2282	94,50
7	0,0133	0,0794	0,3617	94,50	-0,0185	0,0666	0,3398	94,00	0,0334	0,0565	0,2469	94,50
8	0,0066	0,0545	0,2597	94,00	-0,0003	0,0497	0,2561	95,00	0,0123	0,0305	0,1615	97,50
9	0,0140	0,0766	0,3588	95,00	-0,0172	0,0680	0,3355	92,50	0,0327	0,0529	0,2437	94,50
10	0,0051	0,0527	0,2561	93,00	0,0028	0,0496	0,2522	94,50	0,0120	0,0302	0,1607	97,50
11	0,0101	0,0781	0,3604	92,50	-0,0134	0,0639	0,3388	94,50	0,0325	0,0531	0,2440	95,00
12	0,0033	0,0512	0,2558	95,00	0,0029	0,0504	0,2521	95,50	0,0111	0,0298	0,1599	96,50
13	0,0128	0,0644	0,3170	94,50	-0,0034	0,0550	0,2983	96,00	0,0203	0,0409	0,1991	95,50
14	0,0018	0,0441	0,2151	94,50	0,0050	0,0434	0,2140	93,50	0,0072	0,0248	0,1314	96,00
15	0,0147	0,0653	0,3165	94,00	-0,0033	0,0574	0,2938	94,50	0,0219	0,0427	0,1956	95,50
16	0,0030	0,0435	0,2139	94,50	0,0043	0,0432	0,2130	92,50	0,0078	0,0252	0,1306	95,50
17	0,0131	0,0677	0,3125	93,00	-0,0041	0,0575	0,2947	92,50	0,0194	0,0418	0,1948	96,50
18	0,0035	0,0452	0,2134	94,00	0,0047	0,0444	0,2118	95,50	0,0071	0,0247	0,1304	96,00
ϕ_4												
1	-0,0752	0,0821	0,3431	97,00	-0,0381	0,0745	0,4018	95,50	-	-	-	-
2	-0,0418	0,0548	0,2648	98,50	-0,0138	0,0609	0,3305	98,00	-	-	-	-
3	-0,0739	0,0795	0,3414	97,00	-0,0399	0,0731	0,4039	96,00	-	-	-	-
4	-0,0427	0,0554	0,2630	96,50	-0,0151	0,0604	0,3280	96,50	-	-	-	-
5	-0,0745	0,0817	0,3433	98,00	-0,0347	0,0723	0,3974	96,00	-	-	-	-
6	-0,0407	0,0539	0,2616	96,50	-0,0142	0,0609	0,3303	97,50	-	-	-	-
7	-0,0444	0,0578	0,2701	96,50	-0,0148	0,0676	0,3306	94,00	-	-	-	-
8	-0,0201	0,0387	0,2102	99,50	-0,0014	0,0538	0,2673	93,00	-	-	-	-
9	-0,0425	0,0555	0,2693	96,50	-0,0155	0,0670	0,3282	96,00	-	-	-	-
10	-0,0187	0,0387	0,2079	98,50	-0,0050	0,0548	0,2667	91,50	-	-	-	-
11	-0,0452	0,0572	0,2705	95,50	-0,0123	0,0635	0,3316	96,50	-	-	-	-
12	-0,0197	0,0391	0,2079	96,00	-0,0011	0,0548	0,2646	93,00	-	-	-	-
13	-0,0307	0,0461	0,2407	97,00	-0,0069	0,0556	0,2954	93,50	-	-	-	-
14	-0,0129	0,0369	0,1815	95,00	-0,0008	0,0488	0,2311	91,00	-	-	-	-
15	-0,0296	0,0473	0,2363	96,00	-0,0069	0,0560	0,2943	94,50	-	-	-	-
16	-0,0134	0,0366	0,1809	95,00	0,0002	0,0477	0,2310	91,50	-	-	-	-
17	-0,0301	0,0467	0,2391	97,50	-0,0056	0,0600	0,2920	95,00	-	-	-	-
18	-0,0128	0,0357	0,1831	96,50	-0,0005	0,0483	0,2312	92,50	-	-	-	-
ϕ_5												
1	-0,0752	0,0821	0,3431	97,00	-0,0381	0,0745	0,4018	95,50	-	-	-	-
2	-0,0418	0,0548	0,2648	98,50	-0,0138	0,0609	0,3305	98,00	-	-	-	-
3	-0,0739	0,0795	0,3414	97,00	-0,0399	0,0731	0,4039	96,00	-	-	-	-
4	-0,0427	0,0554	0,2630	96,50	-0,0151	0,0604	0,3280	96,50	-	-	-	-
5	-0,0745	0,0817	0,3433	98,00	-0,0347	0,0723	0,3974	96,00	-	-	-	-
6	-0,0407	0,0539	0,2616	96,50	-0,0142	0,0609	0,3303	97,50	-	-	-	-
7	-0,0444	0,0578	0,2701	96,50	-0,0148	0,0676	0,3306	94,00	-	-	-	-
8	-0,0201	0,0387	0,2102	99,50	-0,0014	0,0538	0,2673	93,00	-	-	-	-
9	-0,0425	0,0555	0,2693	96,50	-0,0155	0,0670	0,3282	96,00	-	-	-	-
10	-0,0187	0,0387	0,2079	98,50	-0,0050	0,0548	0,2667	91,50	-	-	-	-
11	-0,0452	0,0572	0,2705	95,50	-0,0123	0,0635	0,3316	96,50	-	-	-	-
12	-0,0197	0,0391	0,2079	96,00	-0,0011	0,0548	0,2646	93,00	-	-	-	-
13	-0,0307	0,0461	0,2407	97,00	-0,0069	0,0556	0,2954	93,50	-	-	-	-
14	-0,0129	0,0369	0,1815	95,00	-0,0008	0,0488	0,2311	91,00	-	-	-	-
15	-0,0296	0,0473	0,2363	96,00	-0,0069	0,0560	0,2943	94,50	-	-	-	-
16	-0,0134	0,0366	0,1809	95,00	0,0002	0,0477	0,2310	91,50	-	-	-	-
17	-0,0301	0,0467	0,2391	97,50	-0,0056	0,0600	0,2920	95,00	-	-	-	-
18	-0,0128	0,0357	0,1831	96,50	-0,0005	0,0483	0,2312	92,50	-	-	-	-

Na sequência analisamos o comportamento das estimativas do número de animais vivos na t -ésima ocasião de amostragem, N_t , $t \in \mathcal{T}$. Observe que o número esperado de animais vivos em cada ocasião é definido por

$$\mathbb{E}(N_1) = \mathbb{E}(N) \cdot \mathbb{E}(\beta_1)$$

$$\mathbb{E}(N_t) = \mathbb{E}(N_{t-1})\mathbb{E}(\phi_{t-1}) + \mathbb{E}(N) \cdot \mathbb{E}(\beta_t), \quad t \in \{2, \dots, T\}.$$

Por exemplo, no primeiro cenário os tamanhos populacionais esperados nas seis ocasiões são dados respectivamente por 400, 430, 422,5, 369, 432,1 e 395,68 enquanto que, para o cenário 18, estes valores são 1200, 1290, 1267,5, 1107, 1296,3 e 1187,04. Na Figura 7 apresentamos os vieses relativos do estimador pontual para os tamanhos populacionais em cada ocasião de amostragem.

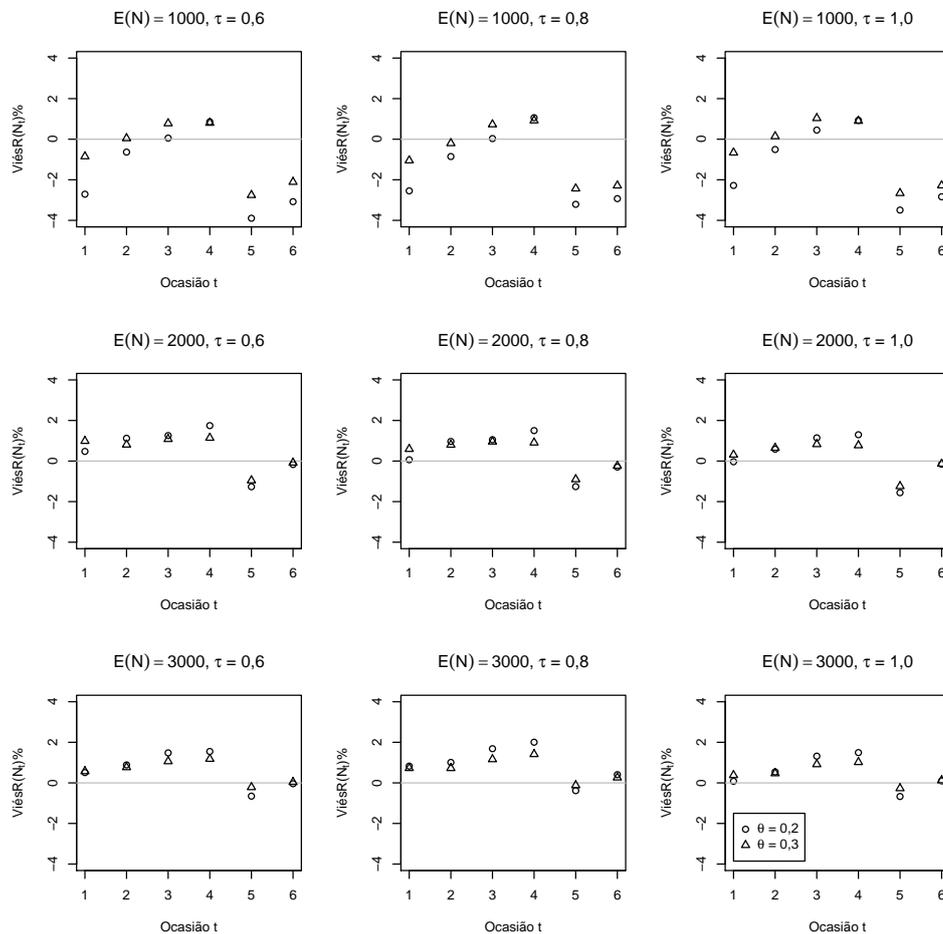


Figura 7 – Viés Relativo (%) do número estimado de animais vivos em cada ocasião de captura.

Primeiramente, destacamos que um aumento da probabilidade de captura diminui, em geral, os vieses observados destas quantidades e que o valor do parâmetro ω não exerce influência significativa neste sentido. Por outro lado, observamos na quinta ocasião de amostragem um viés relativo mais acentuado quando comparado às demais ocasiões, principalmente quando

temos tamanho populacional esperado de 1000 animais. Isto pode ter ocorrido pelo fato de que a probabilidade de sobrevivência entre a quarta e quinta ocasiões de amostragem foi definida por 90% (a mais alta no estudo de simulação), e as estimativas pontuais ficaram pouco abaixo deste valor para este parâmetro (veja Tabela 9). Assim, o modelo estimou, em média, menos animais sobreviventes entre estas duas ocasiões o que, possivelmente, levou ao viés negativo para N_5 . Contudo, destacamos que os vieses relativos não ultrapassaram, em valores absolutos, mais de 4% do verdadeiro valor do tamanho populacional, apresentando um comportamento razoável para estes estimadores. Com efeito, o comportamento do estimador mostrou-se aproximadamente constante em todas as ocasiões com relação ao erro absoluto médio, exibido na Figura 8, com maior precisão (menor EAM) quando $\theta = 0,30$.

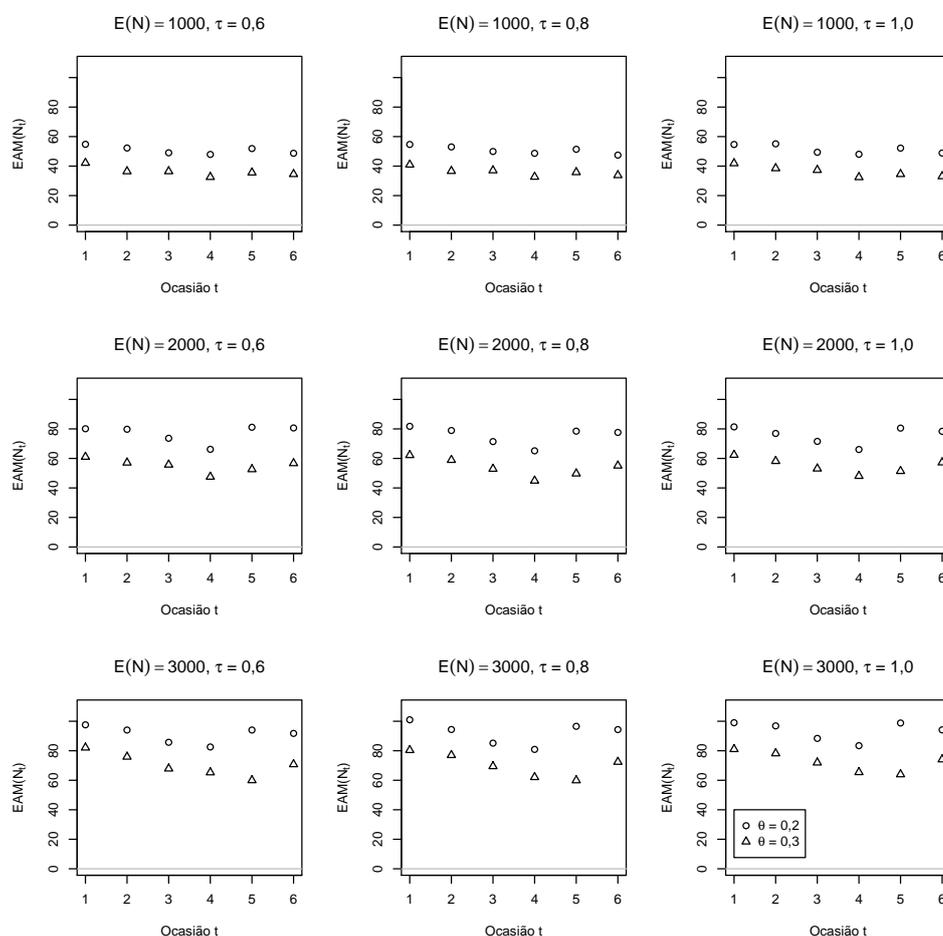


Figura 8 – EAM do número estimado de animais vivos em cada ocasião de captura.

Na Tabela 10 são apresentadas as medidas de desempenho do estimador intervalar para estas quantidades. Em geral, notamos uma diminuição de aproximadamente 30% na amplitude média do intervalo de credibilidade com o aumento da probabilidade de captura. Além disto, a probabilidade de cobertura esteve próximo do verdadeiro valor de credibilidade adotado para o intervalo HDP. Novamente, notamos que o valor de ω não influenciou significativamente as

qualidades das estimativas produzidas.

Tabela 10 – Medidas de desempenho do estimador intervalar bayesiano para os parâmetros N_i .

Cen.	N_1		N_2		N_3		N_4		N_5		N_6	
	AM	PC										
1	255,72	90,0	259,50	93,5	257,96	96,0	232,44	92,0	239,95	92,0	247,00	95,0
2	188,14	93,0	182,89	93,0	180,73	97,0	162,96	95,5	166,62	92,5	180,40	96,5
3	254,34	92,0	258,26	92,0	257,33	96,0	231,46	93,0	240,51	90,5	247,43	93,5
4	188,58	92,5	183,49	92,5	182,01	95,0	163,05	94,0	167,22	93,0	181,10	95,0
5	258,99	91,0	260,16	91,0	260,19	96,5	232,97	92,5	241,20	93,0	248,76	93,5
6	189,76	94,5	185,26	94,5	182,50	96,0	164,05	98,0	166,72	92,0	180,86	96,5
7	385,53	92,0	389,28	94,5	383,58	95,0	343,77	95,5	361,29	91,0	377,64	94,5
8	281,21	93,0	271,85	96,5	267,59	93,5	239,67	96,0	249,28	93,5	273,06	95,0
9	385,00	92,0	387,52	93,5	382,66	97,5	343,04	94,5	361,36	93,5	372,33	94,5
10	280,95	91,5	273,76	93,0	266,76	94,5	240,14	95,5	251,13	94,0	273,46	94,5
11	386,60	91,5	390,45	94,5	387,87	96,5	346,56	94,0	362,10	92,5	381,25	94,5
12	280,87	91,5	273,67	93,5	267,37	94,5	240,36	95,5	250,40	91,5	273,67	94,0
13	487,99	95,0	489,23	97,0	482,49	98,5	433,84	97,5	457,69	95,5	477,22	97,5
14	345,72	94,5	337,48	93,5	328,78	95,0	296,53	95,0	312,19	95,5	340,79	94,0
15	492,56	93,0	489,97	95,0	484,46	97,5	436,22	97,5	458,05	94,0	482,11	96,0
16	352,56	93,5	340,65	93,0	334,86	94,0	300,78	93,5	319,52	94,5	347,98	94,0
17	490,85	94,0	492,17	97,0	484,25	97,5	435,17	95,5	459,67	96,5	478,99	94,5
18	350,40	91,5	338,86	93,0	332,10	94,0	298,85	92,5	316,62	95,0	344,68	94,5

Por fim, exibimos as medidas de desempenho para os parâmetros da estrutura hierárquica para os tamanhos populacionais. Primeiramente, apresentamos nas Tabelas 11 e 12 as medidas para os parâmetros μ e σ , respectivamente. Notamos um viés negativo para μ quando $\omega < 1$ e positivo quando $\omega = 1$, enquanto ocorreu o inverso para o parâmetro σ , isto é, vieses positivos quando $\omega < 1$ e negativos quando $\omega = 1$. Contudo, estes valores foram considerados relativamente baixos com vieses relativos, em termos absolutos, abaixo de 4% para μ e de 8% para σ . Além disto, o aumento de $\mathbb{E}(N)$, θ e ω acarretou, em geral, na diminuição dos vieses, erros médios e amplitude do intervalo de credibilidade. Na maioria dos cenários, a probabilidade de cobertura para μ esteve acima do valor nominal considerado. Para σ , esta medida de desempenho ficou próximo a 95% embora tenha apresentado em alguns cenários valores abaixo do esperado, como 89% e 91%.

Exibimos na Tabela 13 os resultados obtidos para o parâmetro ω . Notamos que vieses baixos foram observados em todos os cenários (com $|\text{Viés}| < 5\%$). De fato, quando $\omega = 1$, valores subestimados foram observados, uma vez que há probabilidade zero de gerar a posteriori valores iguais a 1 para este parâmetro. Da mesma forma, a probabilidade de cobertura deste estimador é zero. Comportamento similar aos estimadores de outros parâmetros também foi observado para ω , no sentido de aumento da precisão a partir de valores mais elevados para $\mathbb{E}(N)$ e θ .

Tabela 11 – Medidas de desempenho dos estimadores bayesianos para o parâmetro μ .

Cen.	Viés	ViésR(%)	EAM	EQM	AM	PC
1	-0,0939	-3,09	0,2183	0,0831	1,1376	96,5
2	-0,0745	-2,45	0,1978	0,0655	0,9967	92,5
3	-0,0423	-1,54	0,1834	0,0515	0,9507	97,0
4	-0,0371	-1,35	0,1655	0,0441	0,8433	96,0
5	0,0868	3,47	0,1515	0,0348	0,7316	94,0
6	0,0587	2,35	0,1376	0,0284	0,6522	94,5
7	-0,0412	-1,09	0,1845	0,0529	0,9097	97,0
8	-0,0441	-1,17	0,1691	0,0461	0,8526	97,0
9	-0,0275	-0,80	0,1498	0,0353	0,7960	96,5
10	-0,0204	-0,60	0,1338	0,0288	0,7317	96,5
11	0,0370	1,16	0,1329	0,0273	0,6458	95,5
12	0,0221	0,69	0,1210	0,0228	0,6061	95,0
13	-0,0324	-0,78	0,1644	0,0426	0,8443	96,0
14	-0,0289	-0,69	0,1544	0,0382	0,8052	95,5
15	-0,0229	-0,60	0,1370	0,0302	0,7348	96,0
16	-0,0163	-0,43	0,1265	0,0257	0,6975	95,5
17	0,0195	0,54	0,1276	0,0252	0,6139	93,0
18	0,0132	0,36	0,1226	0,0226	0,5889	94,5

Tabela 12 – Medidas de desempenho dos estimadores bayesianos para o parâmetro σ .

Cen.	Viés	ViésR(%)	EAM	EQM	AM	PC
1	0,0847	8,46	0,1847	0,0613	0,8793	94,5
2	0,0707	7,07	0,1715	0,0488	0,7985	93,5
3	0,0418	4,17	0,1512	0,0370	0,7190	96,5
4	0,0395	3,95	0,1361	0,0327	0,6649	95,5
5	-0,0596	-5,96	0,1195	0,0221	0,5300	89,0
6	-0,0427	-4,27	0,1084	0,0180	0,5022	91,5
7	0,0546	5,45	0,1490	0,0384	0,7063	94,5
8	0,0539	5,38	0,1429	0,0395	0,6710	93,5
9	0,0406	4,06	0,1208	0,0236	0,6169	97,0
10	0,0337	3,36	0,1154	0,0198	0,5759	97,0
11	-0,0233	-2,32	0,0970	0,0145	0,4780	95,0
12	-0,0143	-1,42	0,0907	0,0126	0,4599	96,0
13	0,0466	4,65	0,1290	0,0296	0,6489	94,0
14	0,0434	4,34	0,1265	0,0290	0,6239	93,5
15	0,0363	3,62	0,1144	0,0213	0,5674	96,5
16	0,0312	3,18	0,1035	0,0175	0,5390	96,0
17	-0,0139	-1,38	0,0921	0,0127	0,4534	94,0
18	-0,0075	-0,75	0,0887	0,0120	0,4409	93,5

Tabela 13 – Medidas de desempenho dos estimadores bayesianos para o parâmetro ω .

Cen.	Viés		ViésR(%)		EAM		EQM		AM	PC
	$E(\omega)$	$M(\omega)$	$E(\omega)$	$M(\omega)$	$E(\omega)$	$M(\omega)$	$E(\omega)$	$M(\omega)$		
1	0,0132	0,0117	2,1981	1,9534	0,0649	0,0655	0,0068	0,0068	0,3038	92,0
2	0,0072	0,0070	1,2022	1,1693	0,0613	0,0616	0,0060	0,0061	0,2845	91,5
3	0,0024	0,0044	0,2990	0,5522	0,0554	0,0567	0,0043	0,0045	0,2631	93,0
4	-0,0001	0,0024	-0,0152	0,3038	0,0491	0,0498	0,0035	0,0036	0,2462	94,5
5	-0,0513	-0,0441	-5,1321	-4,4143	0,0513	0,0441	0,0032	0,0026	0,1210	-
6	-0,0407	-0,0340	-4,0688	-3,3976	0,0407	0,0340	0,0020	0,0015	0,1005	-
7	0,0006	0,0014	0,0917	0,2256	0,0592	0,0598	0,0054	0,0055	0,2690	92,5
8	0,0002	0,0011	0,0250	0,1857	0,0584	0,0592	0,0054	0,0056	0,2642	92,0
9	-0,0042	-0,0010	-0,5270	-0,1306	0,0481	0,0490	0,0034	0,0035	0,2293	96,0
10	-0,0067	-0,0032	-0,8360	-0,4012	0,0474	0,0478	0,0034	0,0034	0,2211	95,0
11	-0,0329	-0,0267	-3,2869	-2,6661	0,0329	0,0267	0,0013	0,0009	0,0837	-
12	-0,0276	-0,0217	-2,7632	-2,1681	0,0276	0,0217	0,0009	0,0006	0,0730	-
13	-0,0018	-0,0007	-0,2935	-0,1097	0,0591	0,0598	0,0054	0,0055	0,2624	93,0
14	-0,0020	-0,0009	-0,3408	-0,1528	0,0585	0,0592	0,0054	0,0055	0,2602	90,5
15	-0,0054	-0,0019	-0,6714	-0,2408	0,0472	0,0476	0,0034	0,0034	0,2189	94,0
16	-0,0068	-0,0032	-0,8521	-0,3964	0,0463	0,0467	0,0033	0,0033	0,2148	93,0
17	-0,0256	-0,0197	-2,5573	-1,9658	0,0256	0,0197	0,0007	0,0005	0,0693	-
18	-0,0233	-0,0175	-2,3284	-1,7492	0,0233	0,0175	0,0006	0,0004	0,0645	-

APLICAÇÃO: ESPÉCIE *IANDUMOEMA UAI*

Neste capítulo aplicamos a metodologia proposta no capítulo anterior para estimação dos parâmetros populacionais utilizando os conjuntos de dados reais de captura da espécie *Iandumoema uai* apresentados por Monte (2017). Como comentado na Seção 1.1 do Capítulo 1, o processo de captura desta espécie foi em seis ocasiões de amostragem dadas, respectivamente, em Junho, Agosto e Outubro de 2014 e Junho, Setembro e Outubro de 2015, no interior da caverna Gruta Olhos d'Água, Minas Gerais. A caverna foi subdividida em 50 regiões de 30 metros cada no sentido horizontal a partir da sua entrada, resultando um comprimento total de 1.500 metros para a realização da amostragem. A partir do processo de marcação adotado, foi possível obter o histórico completo de captura dos aracnídeos, isto é, em quais ocasiões e respectivas regiões cada aracnídeo foi capturado. A seguir, apresentamos uma descrição desse conjunto de dados.

4.1 Análise descritiva dos dados

Primeiramente, apresentamos na Tabela 14 um resumo do total de capturas e recapturas obtidas para esta espécie com relação ao sexo dos aracnídeos, onde n_t representa o número total de aracnídeos capturados (marcados e não marcados) na t -ésima ocasião de amostragem, f_t o número de aracnídeos capturados não marcados na t -ésima ocasião de amostragem e m_t' o número de aracnídeos capturados pela primeira vez na t -ésima ocasião de amostragem e recapturados na ocasião t' , com $t < t'$. Na última linha da tabela, temos os totais (soma) de cada coluna.

Notemos que foram capturados 1.453 aracnídeos distintos da espécie (905 machos e 548 fêmeas), com uma taxa de recaptura de 23,94% (285 de 1190) para os machos e de 25,13% para as fêmeas (184 de 732). Observamos também que os aracnídeos de ambos os sexos apresentaram números proporcionalmente semelhantes de recapturas entre os períodos de amostragem, o que

Tabela 14 – Frequências de capturas de machos e fêmeas da espécie *Iandumoema uai*.

Ocasião (t)	Machos							Fêmeas						
	n_t	f_t	$m_t^{(2)}$	$m_t^{(3)}$	$m_t^{(4)}$	$m_t^{(5)}$	$m_t^{(6)}$	n_t	f_t	$m_t^{(2)}$	$m_t^{(3)}$	$m_t^{(4)}$	$m_t^{(5)}$	$m_t^{(6)}$
1	17	17	6	1	1	0	0	25	25	10	3	1	0	0
2	161	155	-	33	9	11	8	110	100	-	13	4	3	7
3	269	235	-	-	25	27	17	89	73	-	-	6	7	7
4	198	163	-	-	-	51	38	173	162	-	-	-	53	40
5	274	185	-	-	-	-	58	194	131	-	-	-	-	30
6	271	150	-	-	-	-	-	141	57	-	-	-	-	-
Total	1190	905	6	34	35	89	121	732	548	10	16	11	63	84

pode ser indicativo de taxas iguais de sobrevivência. Por outro lado, nota-se baixo número de capturas na primeira ocasião quando comparado às demais. Isto decorre do fato que não houve amostragem a partir da região 19 nesta ocasião, isto é, apenas as 18 primeiras regiões foram visitadas pelos pesquisadores. Assim, o modelo estatístico necessita de uma simples modificação fixando a probabilidade de captura igual à zero nestas regiões durante a primeira ocasião. Para isto redefinimos na equação (2.4) para a r -ésima região uma probabilidade de captura igual à

$$\theta_t(u, v) = \theta \cdot I(u \leq t \leq v) \cdot [I(t > 1) + I(t = 1, r \leq 18)]. \quad (4.1)$$

Na Figura 9 exibimos os números totais de capturas (e recapturas) da espécie ao longo da área de estudo. Primeiramente, destacamos a similaridade da distribuição espacial dos aracnídeos para ambos os sexos e, como foi constatado na Tabela 14, da maior frequência de machos. Nota-se diferenças significativas nos totais de capturas entre regiões com baixos números de capturados (até mesmo sem captura alguma) e outras com números elevados, como a região 20 que apresentou 83 aracnídeos distintos capturados. A baixa frequência de aracnídeos capturados nas primeiras regiões de amostragem indica a preferência da espécie em locais mais internos da caverna, o que torna o início da caverna uma região possivelmente não habitada pela espécie. No mesmo sentido, nota-se que os aracnídeos foram vistos até a última região de amostragem, o que pode indicar um habitat além destas regiões de amostragens.

Vale ressaltar que apenas uma fêmea foi recapturada em regiões diferentes da sua primeira captura (representando apenas 0,1821% do total de fêmeas capturadas), enquanto que este fato ocorreu em apenas 0,5495% do total de machos capturados (5 machos). Desta forma, não consideramos os históricos de captura destes seis animais na análise de dados. Assim, como todos os demais animais foram recapturados na mesma região da sua primeira captura, o processo inferencial para obtenção das estimativas dos tamanhos populacionais pode ser realizado segundo a modelagem proposta no capítulo anterior.

4.2 Análise Inferencial

Nesta aplicação fazemos uma análise inferencial para dois conjuntos de dados referentes a cada um dos sexos da espécie *Iandumoema uai*. Para isto, seja $p = 1$ referente à população de

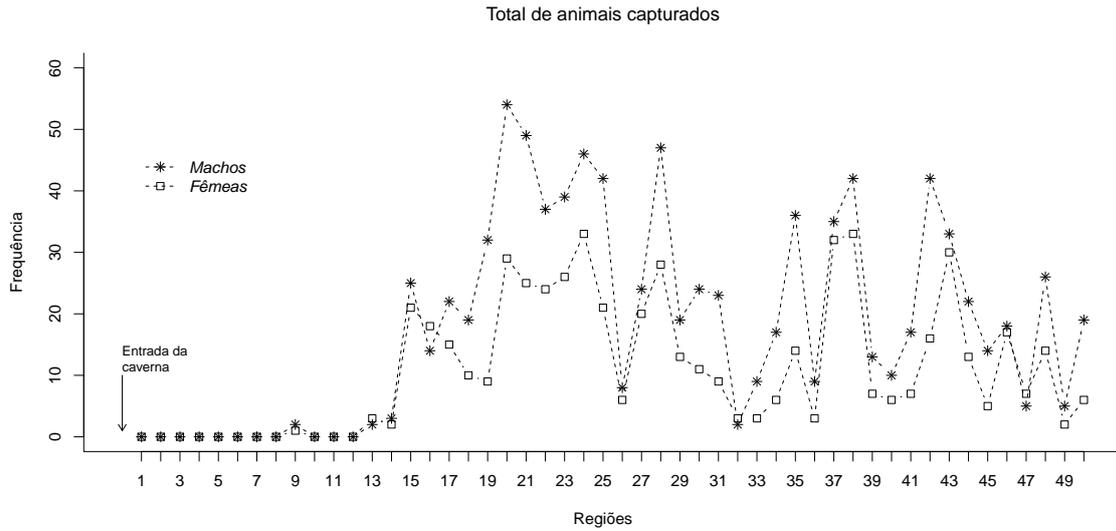


Figura 9 – Número total de animais distintos capturados em cada região da espécie *Iandumoema uai*.

machos e $p = 2$ às fêmeas. No que tange a estrutura hierárquica dos tamanhos populacionais, definimos

$$\begin{aligned}
 N^{(p)} | \lambda_r^{(p)} &\sim \text{ZIPoisson}(\lambda^{(p)}, \omega^{(p)}), r \in \{1, 2, \dots, 50\} \\
 \log(\lambda_r^{(p)}) | \mu^{(p)}, \sigma &\sim \text{Normal}(\mu^{(p)}, \sigma^2), r \in \{1, 2, \dots, 50\} \\
 \mu^{(p)} &\sim \text{Normal}(0, 10^2), \\
 \sigma^{-2} &\sim \text{Gama}(0, 1; 0, 1), \\
 \omega^{(p)} &\sim \text{Beta}(1, 1),
 \end{aligned}$$

para $p = 1, 2$, com o parâmetro σ comum para ambos os sexos. Adicionalmente, adotamos para os demais parâmetros as seguintes distribuições a priori:

$$\begin{aligned}
 \phi_t^{(p)} &\sim \text{Beta}(1, 1), \quad t \in \{1, 2, \dots, T-1\}, \\
 \theta^{(p)} &\sim \text{Beta}(1, 1), \\
 \beta^{(p)} &\sim \text{Dirichlet}(1, 1, 1, 1, 1, 1),
 \end{aligned}$$

para $p = 1, 2$. Tais valores dos hiperparâmetros foram escolhidos para que as distribuições a priori sejam pouco informativas, de acordo com a nossa crença.

Além da determinação das estimativas das abundâncias populacionais para cada sexo, também comparamos modelos para igualdade das probabilidades de nascimento, morte e captura entre eles, uma vez que os dois conjuntos de dados referem-se à amostragem da mesma espécie. Para isto, supomos que as amostras observadas, $\mathbf{n}^{(1)}$ para machos e $\mathbf{n}^{(2)}$ para fêmeas, são realizações independentes de captura no sentido que podemos associar uma função de verossimilhança

conjunta por

$$\begin{aligned} \mathcal{L}_c(\mathbf{N}^{(1)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\phi}^{(1)}, \boldsymbol{\theta}^{(1)}, \mathbf{N}^{(2)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\phi}^{(2)}, \boldsymbol{\theta}^{(2)}; \mathbf{n}^{(1)}, \mathbf{n}^{(2)}) = \\ = \mathcal{L}(\mathbf{N}^{(1)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\phi}^{(1)}, \boldsymbol{\theta}^{(1)}; \mathbf{n}^{(1)}) \mathcal{L}(\mathbf{N}^{(2)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\phi}^{(2)}, \boldsymbol{\theta}^{(2)}; \mathbf{n}^{(2)}) \end{aligned}$$

onde \mathcal{L} é a função de verossimilhança definida em (3.1) considerando a restrição (4.1).

Para verificar a igualdade destes parâmetros, consideramos o critério *Deviance Information Criterion* (DIC) (SPIEGELHALTER *et al.*, 1998). Para comparação entre dois modelos competidores, o que se ajusta melhor a um conjunto de dados é aquele com menor valor de DIC. Para facilidade de comparação, seja DIC_i o valor calculado do critério DIC para o i -ésimo modelo ajustado e defina

$$\Delta_i(\text{DIC}) = \text{DIC}_i - \min_j \{\text{DIC}_j\}.$$

Portanto, valor de $\Delta_i(\text{DIC})$ igual à 0 (zero) favorece o modelo i .

Assim, utilizamos o algoritmo MCMC para obter uma amostra de 150.000 valores gerados para cada parâmetro. Para a obtenção dos resumos a posteriori descartamos as 50.000 primeiras amostras geradas e, do restante, tomamos valores a cada 10 gerados para minimizar o efeito de autocorrelação da cadeia, resultando em uma amostra final de $B = 10.000$ valores. Analisamos a convergência das cadeias através do critério de Gelman e Rubin (1992) (GR) utilizando o pacote CODA (BEST; COWLES; VINES, 1995) do *software* R (R Development Core Team, 2014). Para utilizar este diagnóstico de convergência, geramos duas cadeias em paralelo com valores iniciais para os parâmetros em diferentes pontos do espaço paramétrico e, assim, obtivemos duas amostras a posteriori para cada parâmetro. Se o valor deste critério for aproximadamente 1, usualmente menor que 1,1, temos fortes indícios de convergência da cadeia MCMC para uma distribuição estacionária que, sob certas condições é a distribuição marginal a posteriori do respectivo parâmetro. Além deste critério, monitoramos a convergência da cadeia por meio de análise gráfica dos valores gerados cujos gráficos podem ser encontrados no Anexo A conjuntamente com o valor do critério GR. Os valores do critério e os gráficos forneceram indícios de convergência do algoritmo.

Na Tabela 15 são apresentados os valores do critério $\Delta_i(\text{DIC})$ para os 8 diferentes modelos considerados, onde verificamos que o modelo com melhor ajuste aos dados foi aquele que não diferencia taxas de captura, nascimento e recrutamento entre os sexos. Embora o segundo modelo escolhido pelo critério assuma probabilidades de captura diferentes entre os sexos, notamos que a estimativa média a posteriori para $\theta^{(1)}$ foi igual a 0,3105 e para $\theta^{(2)}$ igual a 0,32024 e as inferências sob este modelo são similares às que são apresentadas na sequência considerando o modelo escolhido pelo critério.

Apresentamos na Tabela 16 as estimativas pontuais e intervalares para o parâmetro referente ao tamanho da superpopulação com relação à região de amostragem. Adicionalmente, temos estimativas aproximadas da densidade populacional utilizando uma largura média da

Tabela 15 – Valores do Critério DIC para comparação de modelos.

Comparação entre os parâmetros de			$\Delta_i(\text{DIC})$
nascimento	sobrevivência	captura	
iguais	iguais	iguais	0
iguais	iguais	diferentes	0,93
iguais	diferentes	diferentes	130,03
diferentes	iguais	iguais	239,36
iguais	diferentes	iguais	242,77
diferentes	iguais	diferentes	247,53
diferentes	diferentes	diferentes	408,38
diferentes	diferentes	iguais	411,42

caverna de 2,5 m, totalizando aproximadamente 3.750 m² de área de habitat. A abundância total média de animais machos (1895,38) foi superior à população de fêmeas (1142,53), produzindo respectivas densidades populacionais de 0,5054 e 0,3047 aracnídeos/m². A estimativa total da espécie (machos e fêmeas) foi de 3037,92 aracnídeos, com densidade populacional estimada em 0,8101 aracnídeos/m².

Tabela 16 – Resumos a posteriori para os tamanhos da superpopulação para a espécie *Iandumoema uai*.

População	Quantidade	Média	Mediana	IC _b (95%)
Machos	Abundância	1895,38	1886	(1678, 2133)
	Densidade (m ²)	0,5054	0,5029	(0,4475; 0,5688)
Fêmeas	Abundância	1142,53	1136	(1008, 1293)
	Densidade (m ²)	0,3047	0,3011	(0,2688; 0,3448)
Total	Abundância	3037,91	3024	(2714, 3423)
	Densidade (m ²)	0,8101	0,8064	(0,7237; 0,9128)

Na Tabela 17 são exibidas as estimativas dos parâmetros β , ϕ e θ . Como previamente discutido, estes parâmetros foram considerados iguais para ambos os sexos segundo o critério DIC (ver Tabela 15). Pelas estimativas de β , observamos que 35% do número total estimado de aracnídeos já faziam presentes na população na primeira ocasião de amostragem e aproximadamente 58% do total sendo recrutados até a última ocasião de 2014. A maior taxa de sobrevivência estimada foi aquela entre as épocas 4 e 5 com mais de 88% de sobreviventes, e a menor sendo para o período entre as épocas 3 e 4, com estimativa média de aproximadamente 35%, corroborando com o maior período de tempo entre duas ocasiões de amostragem (aproximadamente oito meses). A probabilidade de captura estimada em 0,3125 significa que espera-se capturar, em cada ocasião, aproximadamente 31% dos animais vivos.

Em seguida, na Tabela 18, apresentamos as estimativas pontuais e intervalares para os números de aracnídeos vivos em cada ocasião de amostragem, N_t , também exibidos graficamente na Figura 10. Observamos uma tendência de crescimento do tamanho populacional para esta espécie. Para comparação de estimativas totais, adotamos para exemplificação a terceira e sexta ocasiões de amostragem, referentes às últimas de cada ano, 2014 e 2015, respectivamente.

Tabela 17 – Resumos a posteriori para os parâmetros de nascimento, sobrevivência e captura.

Parâm.	Média	Mediana	d.p.	IC _b (95%)
β_1	0,3548	0,3543	0,0387	(0,2772, 0,4294)
β_2	0,0355	0,0298	0,0271	(0,0000, 0,0878)
β_3	0,1991	0,1991	0,0235	(0,1523, 0,2450)
β_4	0,2639	0,2629	0,0259	(0,2148, 0,3153)
β_5	0,1195	0,1201	0,0259	(0,0676, 0,1686)
β_6	0,0273	0,0247	0,0180	(0,0000, 0,0609)
ϕ_1	0,7143	0,7153	0,1116	(0,4991, 0,9344)
ϕ_2	0,6579	0,6566	0,0669	(0,5315, 0,7922)
ϕ_3	0,3587	0,3577	0,0336	(0,2933, 0,4228)
ϕ_4	0,8890	0,8897	0,0542	(0,7986, 0,9999)
ϕ_5	0,8563	0,8575	0,0576	(0,7476, 0,9693)
θ	0,3125	0,3119	0,0203	(0,2713, 0,3507)

Entre este período, houve um crescimento de 12,5% no número médio total de aracnídeos desta espécie, de 1179,22 aracnídeos iniciais para 1326,55. Como observado anteriormente, em todas as ocasiões de captura o número de machos foi superior ao número de fêmeas. Observamos que a maior variabilidade e, conseqüentemente, maior amplitude do intervalo HPD, foi na primeira ocasião de amostragem que teve menos regiões amostradas. Esta variabilidade permaneceu aproximadamente constante para as demais ocasiões.

Tabela 18 – Resumos a posteriori para as abundâncias populacionais em cada ocasião de captura.

População	Parâm.	Média	Mediana	d.p.	IC _b (95%)
Machos	N_1	676,24	669	103,24	(489, 889)
	N_2	547,02	545	47,19	(452, 636)
	N_3	767,42	765	54,93	(662, 873)
	N_4	744,13	742	57,34	(633, 854)
	N_5	890,79	891	57,44	(780, 1004)
	N_6	820,59	818	61,59	(702, 942)
Fêmeas	N_1	406,35	402	62,16	(293, 533)
	N_2	324,52	323	28,76	(267, 378)
	N_3	411,80	410	37,26	(336, 482)
	N_4	479,08	478	34,68	(413, 547)
	N_5	560,47	560	35,85	(491, 629)
	N_6	505,96	505	41,49	(424, 585)
Total	N_1	1082,59	1071	161,90	(791, 1415)
	N_2	871,54	868	71,42	(725, 1004)
	N_3	1179,22	1176	87,24	(1015, 1351)
	N_4	1223,21	1221	88,23	(1051, 1395)
	N_5	1451,26	1451	88,94	(1272, 1619)
	N_6	1326,55	1323	99,38	(1124, 1513)

No contexto territorial, apresentamos na Figura 11 as estimativas médias a posteriori e intervalos HPD dos tamanhos das superpopulações em cada região. No início da caverna, até a região 14, há baixo número estimado de animais para ambos os sexos. Exceto a nona

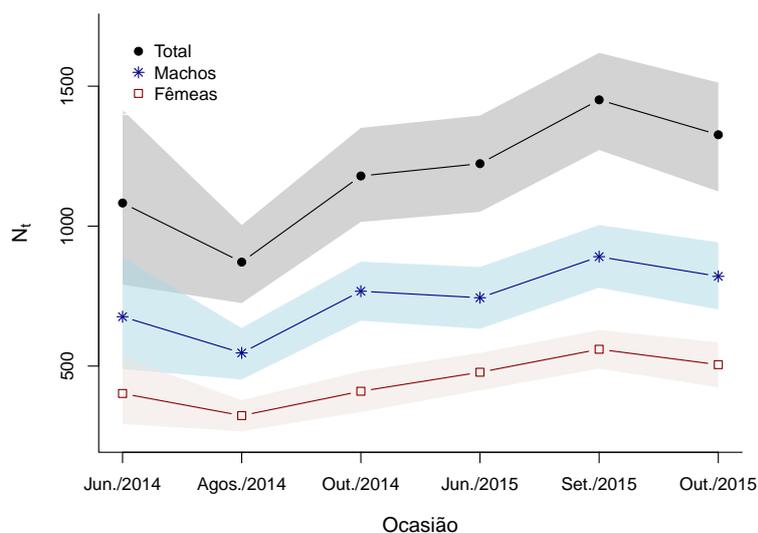


Figura 10 – Abundâncias populacionais estimadas por ocasião.

região de amostragem, em que se observou animais de ambos os sexos, o modelo atribuiu, em média, probabilidade menor que 2% para cada uma das doze primeiras regiões de contar com a presença de aracnídeos machos e aproximadamente 3% para fêmeas desta espécie. Para as demais regiões, obtivemos números mínimos e máximos estimados para as abundâncias locais de 5,76 e 112,92 para machos (regiões 9 e 20, respectivamente) e 3,68 e 68,29 para fêmeas (regiões 9 e 24, respectivamente).

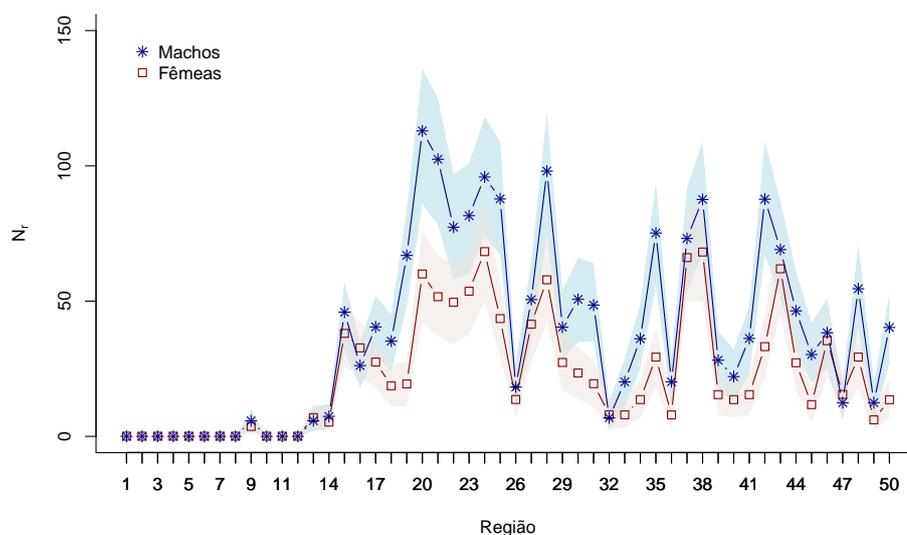


Figura 11 – Abundâncias populacionais estimadas por região.

Para comparação das abundâncias entre os sexos, exibimos na Figura 12 a taxa média

estimada do número de machos para cada fêmea em cada região. Vale ressaltar que não determinamos esta quantidade para as regiões que apresentaram baixa probabilidade de serem povoadas pela espécie. O número de machos foi superior ao número de fêmeas em quase todas as regiões. Em média, observamos um número estimado de aproximadamente dois machos para cada fêmea, sendo que as regiões 19, 33 e 50 apresentaram valores superiores à três machos/fêmea. Embora a maioria das regiões apresentou maior abundância de machos, o número de fêmeas foi moderadamente superior para as regiões 16 e 47. Calculamos densidades médias 0,6480 machos/m² e 0,3906 fêmeas/m² entre as décima terceira e quinquagésima regiões.

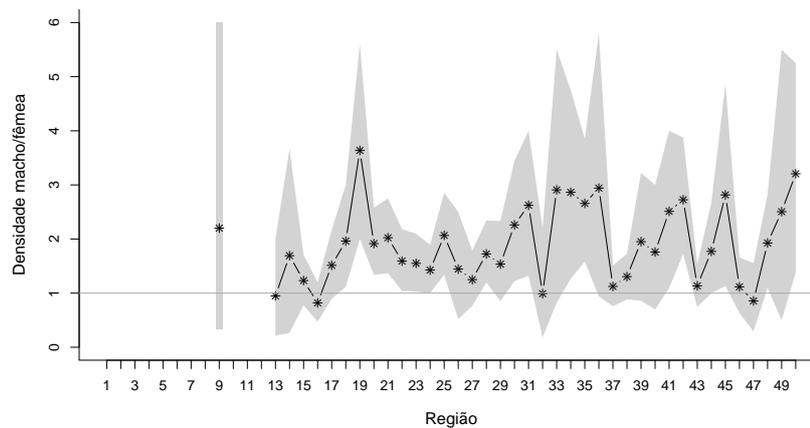


Figura 12 – Densidade média estimada de machos para cada fêmea por região.

Por fim, apresentamos na Figura 13 os números médios estimados de aracnídeos machos e fêmeas vivos em cada região e ocasião de amostragem. Os valores foram calculados em regiões de 60 metros. Como observado nas Figuras 10 e 11 o número de machos foi superior ao total de fêmeas em todas as ocasiões e na maioria das regiões, com maior abundância para a região entre 540 a 840 metros no sentido horizontal da entrada da caverna durante a quinta ocasião de amostragem.

Em seguida, exibimos na Tabela 19 as estimativas dos hiperparâmetros a priori da estrutura hierárquica proposta para os tamanhos populacionais locais. Como esperado, o número médio estimado de aracnídeos (em escala logarítmica) foi superior para a população de machos, com $\hat{\mu}^{(1)} = 3,61$ contra $\hat{\mu}^{(2)} = 3,08$ para as fêmeas e uma variabilidade comum de $\hat{\sigma} = 0,8148$. Além disto, observamos que aproximadamente 23% das regiões de amostragem são consideradas não habitadas pela espécie, o que corresponde especificamente a 11 das 50 regiões pesquisadas sem capturas de aracnídeos.

Com base nos valores gerados para estes parâmetros, realizamos simulações estocásticas para obter a distribuição preditiva da abundância nas regiões não amostradas de acordo com o procedimento descrito no Exemplo da Seção 3.3.3. No total, a caverna tem aproximadamente 9.100 metros no sentido horizontal, sendo que o estudo foi delimitado até 1.500 metros à

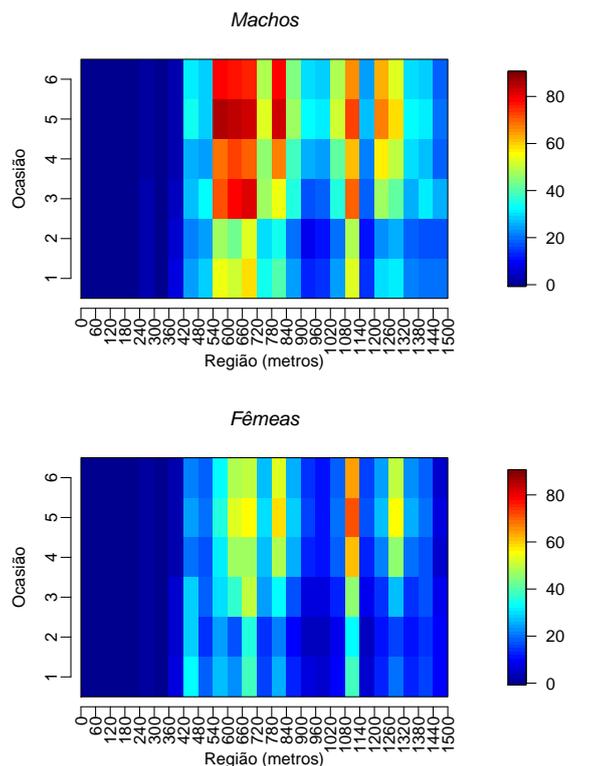


Figura 13 – Abundâncias populacionais estimadas por região e ocasião.

Tabela 19 – Resumos a posteriori para os hiperparâmetros da priori hierárquica da abundância populacional.

Parâm.	Média	Mediana	d.p.	IC _b (95%)
μ (machos)	3,6101	3,6110	0,1493	(3,3359, 3,9117)
μ (fêmeas)	3,0847	3,0855	0,1571	(2,7677, 3,3877)
σ	0,8148	0,8077	0,0875	(0,6542, 0,9913)
ω (machos)	0,7711	0,7745	0,0584	(0,6558, 0,8808)
ω (fêmeas)	0,7762	0,7797	0,0584	(0,6659, 0,8898)

montante da entrada. Desta forma, as predições foram realizadas considerando que existem mais 253 regiões com 30 m cada habitadas no interior da caverna. Na Tabela 20 são apresentadas os resumos obtidos pela predição dos tamanhos populacionais em cada ocasião para toda a caverna. Os resultados predizem um número mínimo de pelo menos 6851 aracnídeos desta espécie (N_2) considerando toda a extensão da caverna, e chegando ao valor de 11403 aracnídeos na quinta ocasião de amostragem. Na última ocasião, estimamos um número médio total de 10424,63 aracnídeos desta espécie, sendo 6536,27 de machos e 3888,37 de fêmeas. Os tamanhos da superpopulação para as regiões não amostradas foram preditos em 23877,69 aracnídeos, sendo 13090,51 machos e 7749,27 fêmeas.

Tabela 20 – Resumos descritivos dos valores preditos para as abundâncias \tilde{N}_i para a espécie *Iandumoema uai*.

	Parâm.	Predição (restante da caverna)			Predição (toda a caverna)		
		Média	Mediana	IC _b (95%)	Média	Mediana	IC _b (95%)
Machos	N_1	4662,61	4535,0	[2782, 6748]	5338,85	5210,0	[3327, 7581]
	N_2	3755,79	3681,0	[2562, 5229]	4302,81	4226,0	[3045, 5797]
	N_3	5080,21	4984,0	[3444, 6958]	5847,63	5755,0	[4203, 7827]
	N_4	5271,93	5160,0	[3530, 7188]	6016,06	5906,0	[4166, 7936]
	N_5	6251,21	6130,0	[4339, 8551]	7142,00	7017,0	[5085, 9385]
	N_6	5715,68	5600,0	[3755, 7733]	6536,27	6422,0	[4544, 8630]
Fêmeas	N_1	2761,61	2686,0	[1639, 4075]	3167,96	3091,0	[1986, 4586]
	N_2	2224,02	2176,0	[1466, 3068]	2548,54	2502,0	[1743, 3407]
	N_3	3006,59	2942,0	[2006, 4141]	3418,39	3356,0	[2368, 4560]
	N_4	3121,19	3048,5	[2094, 4279]	3600,27	3529,0	[2558, 4801]
	N_5	3701,05	3624,5	[2511, 5055]	4261,52	4189,0	[3038, 5642]
	N_6	3382,41	3316,0	[2266, 4658]	3888,37	3823,0	[2719, 5177]
Total	N_1	7424,22	7260,5	[4816, 10479]	8506,81	8340,5	[5581, 11775]
	N_2	5979,81	5889,0	[4281, 7787]	6851,35	6759,0	[5082, 8761]
	N_3	8086,80	7975,5	[5903, 10535]	9266,02	9154,5	[7014, 11843]
	N_4	8393,12	8267,0	[6127, 10831]	9616,34	9492,0	[7277, 12194]
	N_5	9952,26	9818,5	[7377, 12735]	11403,53	11276,0	[8731, 14282]
	N_6	9098,09	8972,0	[6683, 11840]	10424,63	10300,0	[7986, 13385]

MODELO DE CRE PARA POPULAÇÃO ABERTA

Resumo: Neste capítulo propomos um modelo de captura-recaptura espacial para estimação do número de animais de uma população aberta. A partir de dados por amostragem de captura-marcação-recaptura realizada em diferentes localidades de uma região, o modelo proposto leva em consideração a movimentação dos animais entre as regiões de amostragem ao longo das ocasiões de captura. A partir desta modelagem, tornou-se possível estimar o tamanho da área de movimento dos animais da espécie. Além disto, a modelagem proposta leva em consideração as taxas de nascimento e morte no período de estudo e a localização espacial dos animais na região pesquisada, permitindo assim estimar o número total de animais vivos em cada ocasião de captura bem como predizer regiões com maiores concentrações de animais. Por fim, propomos uma abordagem bayesiana para o processo inferencial e analisamos as propriedades frequentistas dos estimadores através de um estudo de simulação.

5.1 Introdução

Dupuis e Schwarz (2007) propõem uma reformulação do modelo JS com objetivo de incorporar os deslocamentos geográficos dos animais entre as regiões de amostragem e estimar do tamanho populacional em cada uma delas. O modelo é definido por um processo de transição markoviano com dependência de primeira ordem para caracterizar as probabilidades de mudança de estados do tipo: “ainda não recrutado”, “região atual de moradia enquanto vivo” e “morte”, para cada animal em cada ocasião de amostragem. Como o estado do animal – especificamente a sua região atual de moradia – é observado apenas nas ocasiões em que ele é capturado, o processo subjacente de captura é visto como um Modelo de Markov Oculto (MACDONALD; ZUCCHINI, 1997; ZUCCHINI; MACDONALD; LANGROCK, 2016) ou HMM (sigla para *Hidden Markov Model*, em inglês). Neste contexto, as inferências para as taxas de recrutamentos, deslocamentos

e saídas dos animais da população são baseadas nas estimativas das probabilidades de transição do HMM.

Esta modelagem generalizou diversos trabalhos iniciais que foram propostos no estudo de deslocamentos geográficos dos animais entre regiões de amostragem como, por exemplo, de Schwarz, Schweigert e Arnason (1993), Brownie *et al.* (1993), Dupuis (1995) e King e Brooks (2002) em que não consideram a estimação da abundância populacional e de Arnason (1973) e Seber (1982), que propuseram a estimação do tamanho populacional levando em consideração apenas a ocorrência de mortes. Recentemente, Wu e Holan (2017) modificam o modelo de Dupuis e Schwarz (2007) para incluir covariáveis dos animais e das regiões de amostragem nas probabilidades de captura e sobrevivência do HMM e para aplicá-lo em estudos com múltiplas populações simultaneamente.

Por outro lado, o número de parâmetros desconhecidos nestes modelos depende do total de regiões e ocasiões de amostragem, o qual pode ser elevado dependendo do estudo. Por exemplo, adotando-se probabilidades de transição não estacionárias, o modelo possui $R^2(T - 1) + 2RT - 1$ parâmetros, onde R é o número de regiões e T é o número de ocasiões de amostragem, além de um parâmetro adicional referente ao tamanho da superpopulação. Dupuis e Schwarz (2007) aplicam sua metodologia em um cenário amostral com apenas duas regiões de amostragem e 27.263 animais distintos capturados, enquanto que Wu e Holan (2017) apresentam um estudo de dados simulados para duas populações com tamanhos somados de 18.000 animais em três regiões de amostragem e uma aplicação real com 1.714 animais distintos capturados em seis regiões. Contudo, estudos que apresentam escassez de dados, isto é, baixa frequência de animais capturados, o uso desta modelagem pode ocasionar perdas de precisão das estimativas e de eficiência computacional como otimização numérica e convergências de algoritmos, principalmente quando há um número elevado de regiões de amostragem.

Como alternativa a estes modelos, estudaremos nesta tese a classe de modelos de captura-recaptura espaciais (CRE). Introduzido por Efford (2004), a hipótese ecológica fundamental na formulação de um modelo de CRE segue que cada animal concentra seus movimentos em torno de um determinado ponto de referência, chamado “centro de vida”. Neste contexto, a formulação do modelo compõe-se basicamente de dois submodelos. O primeiro, governado por um processo pontual, descreve o número de animais e as respectivas localizações de seus centros de vida na região de estudo e o segundo leva em conta as probabilidades de captura e recaptura de cada animal, conforme suas localizações. Nesta modelagem, em geral, a região de estudo é definida como um subconjunto do plano cartesiano e os dados amostrais são representados por pontos coordenados no plano informando os locais onde ocorreram as capturas. Tais pontos podem ser obtidos de diversas maneiras como, por exemplo, a partir das localizações espacialmente referenciadas das armadilhas distribuídas estrategicamente na região de estudo.

Naturalmente, o primeiro tipo de armadilha que vem a mente é a captura física do

animal, onde ele ficará preso até ser devidamente marcado e solto novamente na população (BORCHERS; EFFORD, 2008; EFFORD; BORCHERS; BYROM, 2009). Contudo, com o avanço da tecnologia, é possível utilizar atualmente outros mecanismos para identificação do animal sem seu efetivo aprisionamento. Como exemplos, temos as armadilhas compostas por câmeras em que fotografam o animal e registram os instantes em que ele passa em suas frentes (O'CONNELL; NICHOLS; KARANTH, 2010) identificando-o por meio dos seus padrões corporais, como as manchas em onças-pintadas (BORCHERS *et al.*, 2014; SOLLMANN *et al.*, 2011) e listas em tigres (AZIZ *et al.*, 2017; ROYLE *et al.*, 2017), e as armadilhas do tipo “arame-farpado e isca” que coletam vestígios do animal (sangue, pelo, pele) para análises refinadas de identificação do DNA (GARDNER; ROYLE; WEGAN, 2009; GARDNER *et al.*, 2010; KERY *et al.*, 2011). Em geral, dada a posição da armadilha, a probabilidade de captura é comumente definida por uma função decrescente da distância entre ela e o centro de vida do animal.

Por outro lado, a aplicação dos modelos de CRE pode ser estendida para os casos em que os locais de eventuais capturas não são pré-determinados, como obtidos pelas armadilhas de posições fixas supracitadas. Esta situação ocorre quando é atribuída a um (grupo de) pesquisador(es) a tarefa de realizar buscas dos animais na região de estudo. Neste caso, a probabilidade de captura do animal dependerá do protocolo de buscas adotado. Por exemplo, o trajeto pode ser fixado a priori (ROYLE; KERY; GUELAT, 2011) ou ocasional dependendo de informações sobre regiões com a presença dos animais a ser guiado por cães farejadores (THOMPSON; ROYLE; GARNER, 2012) ou, ainda ser realizado em diversas regiões estratégicas e delimitadas dentro do habitat da espécie (ROYLE; YOUNG, 2008; MARQUES; THOMAS; ROYLE, 2011; RUSSELL *et al.*, 2012). Neste contexto, um procedimento estocástico adicional é adotado para determinar a localização instantânea do animal no momento de sua captura, sendo comumente modelada por uma distribuição normal bivariada utilizando a localização do centro de vida como vetor de médias (EFFORD, 2011; ROYLE; KERY; GUELAT, 2011). A probabilidade de captura é determinada a partir de alguma medida da distância entre a localização instantânea do animal até o caminho de buscas percorrido ou, se a busca for dada por regiões, ela pode ser proporcional ao esforço amostral dentro de cada região, como tempo de execução e o tamanho do trajeto percorrido.

É importante ressaltar que a localização do centro de vida não é diretamente observada mesmo quando o animal é capturado. Por este motivo, ela é vista como uma variável latente na modelagem e eliminada via integração (BORCHERS; EFFORD, 2008; EFFORD, 2011; ROYLE *et al.*, 2012; GLENNIE *et al.*, 2017), embora técnicas de dados aumentados podem ser utilizadas para explicitá-la no modelo (CHANDLER; ROYLE *et al.*, 2013; ROYLE *et al.*, 2007; GARDNER *et al.*, 2010). Assim, para inferir a localização do centro de vida para um animal, utilizam-se as posições espacialmente referenciadas dos locais onde ele foi capturado. A grande vantagem dos modelos de CRE é a possibilidade de obter estimativas para o tamanho da área de movimento de cada animal em torno do centro de vida. Novamente, as inferências serão baseadas

nas respectivas localizações onde um determinado animal foi capturado. Capturas em múltiplos e distantes pontos informam grande movimentação do animal, e vice-versa. Por exemplo, na modelagem a partir de dados por caminho de busca, a movimentação animal estará diretamente associada à variabilidade da distribuição normal bivariada adotada. Logo, a partir das inferências sobre a movimentação dos animais e das localizações dos centros de vida, torna-se possível obter uma descrição detalhada acerca do comportamento territorial da espécie.

Apesar destes aspectos inferenciais, a maioria dos trabalhos sobre modelos de CRE foi desenvolvida para população fechada, como pode ser visto nas revisões de literatura apresentadas por Royle *et al.* (2013) e Borchers, Fewster *et al.* (2016). Algumas propostas iniciais de modelagem para populações abertas foram apresentadas sob o paradigma bayesiano, como os trabalhos de Gardner *et al.* (2010) a partir de uma reformulação do modelo JS, de Raabe, Gardner e Hightower (2013) em um estudo de sobrevivência de peixes adaptando o modelo Cormack-Jolly-Seber (CJS) (CORMACK, 1964; JOLLY, 1965; SEBER, 1965) para o contexto espacial e de Ergon e Gardner (2014) com dados de amostragem por desenho robusto. Recentemente, Glennie *et al.* (2017) apresentaram modificações dos modelos JS e CJS para abordagem de CRE e propuseram inferências via estimação por máxima verossimilhança. No entanto, estes trabalhos modelam dados amostrais obtidos por armadilhas de posição fixa. Sollmann *et al.* (2015) apresenta uma modelagem de população aberta no contexto de CRE para amostragem por caminho de buscas. Porém, os autores levam em consideração na amostragem apenas o registro da distância do animal visto (capturado) até o pesquisador, sem a efetiva marcação (identificação) individual para avaliar futuras recapturas (BUCKLAND *et al.*, 2001).

Tanto quanto é do nosso conhecimento um modelo de CRE para população aberta, a partir de dados amostrais obtidos mediante buscas manuais sobre a região de estudo considerando a marcação individual dos animais capturados ainda não foi apresentada na literatura estatística. Assim, propomos neste capítulo a construção de um modelo estatístico nesta direção.

5.2 Modelo estatístico de captura-recaptura

Suponhamos que um processo de captura-marcação-recaptura com T ocasiões de amostragens, $T \in \{2, 3, \dots\}$ seja aplicado em uma população animal aberta. Consideramos novamente o protocolo amostral e as suposições apresentadas no início do Capítulo 3, página 33. Entretanto, suponhamos que a suposição (a) não seja adequada, isto é, a localização de um determinado animal não é restrita em uma mesma região durante as ocasiões de amostragem, porém vamos supor que cada animal é capturado em, no máximo, uma região por ocasião de amostragem.

Seja N o parâmetro referente ao número total de animais desta população presentes (vivos) em pelo menos uma ocasião de captura em alguma região de amostragem, isto é, o tamanho da superpopulação sobre todas as regiões amostradas, $N \in \{1, 2, \dots\}$. Consideramos os conjuntos de índices $\mathcal{N} = \{1, 2, \dots, N\}$, $\mathcal{T} = \{1, 2, \dots, T\}$ e $\mathcal{R} = \{1, 2, \dots, R\}$ referentes à

identificação dos animais da população, ocasiões e regiões de amostragem, respectivamente. O histórico de captura para o animal $i \in \mathcal{N}$ será representado pelo vetor de variáveis aleatórias $\mathbf{Y}_i = (Y_{i,1}, \dots, Y_{i,T})$ com

$$Y_{i,t} = \begin{cases} r, & \text{se o animal } i \text{ for capturado na região } r \text{ durante a } t\text{-ésima ocasião, ou} \\ 0, & \text{se o animal } i \text{ não for capturado em região alguma durante a } t\text{-ésima ocasião,} \end{cases}$$

para $r \in \mathcal{R}$ e $t \in \mathcal{T}$. Notemos que as informações espaciais (localização) de captura foram incorporadas no histórico do animal. Por exemplo, uma configuração $\mathbf{Y}_i = (0, 0, 4, 0, 3, 0)$ significa que o animal i foi capturado somente nas ocasiões 3 e 5 nas respectivas regiões 4 e 3, em um total de $T = 6$ ocasiões de amostragem. Contudo, observe que mesmo sabendo que o animal esteja vivo na quarta ocasião, não temos a informação da sua localização, uma vez que não foi capturado. Além disso, não observamos por quais regiões o animal passou nem se ele estava vivo durante a primeira, segunda e sexta ocasiões.

Observamos neste momento que estamos lidando com três processos estocásticos para definir um histórico de captura do animal, dois deles previamente discutidos no Capítulo 2. O primeiro é o de nascimento e morte dos animais, que será representado novamente pelas variáveis aleatórias (U_i, V_i) referentes, respectivamente, à primeira e última ocasião em que o animal i esteve presente na população, com função de probabilidades conjunta dada pela equação (2.3). O segundo, definido pelo vetor \mathbf{X}_i no início da Seção 2.1.2, página 29, refere-se às capturas do animal ao longo das ocasiões de amostragem, cuja distribuição de probabilidades (condicional) é dada pela equação (2.4). Vale ressaltar que este último diz respeito apenas à captura ou não do animal durante uma determinada ocasião de amostragem, mas sem a informação da respectiva região de ocorrência do evento. Assim, incluiremos neste capítulo um terceiro processo, dito espacial, que definirá a localização de cada animal (vivo) na região de estudo durante uma respectiva ocasião de amostragem. Por fim, estes três processos serão combinados para determinar a distribuição de probabilidades do histórico de captura do animal, \mathbf{Y}_i .

5.2.1 Distribuição espacial e movimentação dos animais

Para a modelagem espacial partiremos do pressuposto que cada animal concentra seus movimentos em torno de um ponto de referência, denominado centro de vida (EFFORD, 2004). Isto posto, o processo espacial será subdividido em duas partes. Primeiramente, modelaremos para cada animal a localização do seu centro de vida e na sequência definiremos o processo que caracteriza uma movimentação ao longo do tempo e em torno desta localização.

Genericamente, denotemos por A a região do habitat da população representada, por exemplo, por um subconjunto do plano cartesiano (ROYLE *et al.*, 2009; EFFORD; BORCHERS; BYROM, 2009; GLENNIE *et al.*, 2017; ROYLE; FULLER; SUTHERLAND, 2018). Consideremos esta região particionada em R regiões disjuntas denotadas por A_1, A_2, \dots, A_R , isto é,

$A_r \cap A_{r'} = \emptyset$, para $r \neq r'$, com

$$A = \bigcup_{r \in \mathcal{R}} A_r.$$

Consideramos uma variável aleatória S_i que denotará a localização do centro de vida do animal $i \in \mathcal{N}$ em A . Supomos que, dado N , as variáveis aleatórias S_1, \dots, S_N são condicionalmente independentes e identicamente distribuídas por uma distribuição com função densidade de probabilidades (f.d.p) $f_S : A \rightarrow [0, +\infty)$ definida por

$$f_S(s|\boldsymbol{\pi}) = \sum_{r \in \mathcal{R}} \pi_r g_r(s), \quad (5.1)$$

com

$$\boldsymbol{\pi} = (\pi_1, \dots, \pi_R) \in \Theta_{\boldsymbol{\pi}} = \left\{ (x_1, \dots, x_R) : x_r \in (0, 1), \sum_{r \in \mathcal{R}} x_r = 1 \right\}$$

e $g_r(s)$ uma f.d.p. definida sobre A_r que, por sua vez, pode depender de parâmetros desconhecidos. Primeiramente, interpretamos o parâmetro π_r como a probabilidade de um animal alocar seu centro de vida na região $r \in \mathcal{R}$, uma vez que

$$P(S_i \in A_r | \boldsymbol{\pi}) = \int_{A_r} f_S(s|\boldsymbol{\pi}) ds = \int_{A_r} \left\{ \sum_{r' \in \mathcal{R}} \pi_{r'} g_{r'}(s) \right\} ds = \pi_r \int_{A_r} g_r(s) ds = \pi_r.$$

De fato, prever quais regiões possuem maior densidade populacional (número de centros de vida) é um dos aspectos importantes para descrição acerca da distribuição espacial da espécie e isto será evidenciada a partir de estimativas obtidas para o vetor paramétrico $\boldsymbol{\pi}$. Por outro lado, a forma de função g_r definirá como os centros de vida serão distribuídos dentro da r -ésima região de amostragem. Ela poderá ser especificada, por exemplo, de acordo com o conhecimento da relação do animal com a respectiva região, bem como função de covariáveis ambientais medidas no local.

Para nossos propósitos e sem perda de generalidade, suponhamos que A seja um intervalo real, $A = (a_0, a_R]$ com $A_r = (a_{r-1}, a_r]$ para $a_{r-1} < a_r$, $r \in \mathcal{R}$. Adicionalmente, suponhamos que a função f_S é uniforme por intervalos, isto é, definimos

$$g_r(s) = \frac{1}{a_r - a_{r-1}} I(s \in A_r).$$

Apesar desta suposição de uniformidade, esta abordagem é mais geral se comparada aos trabalhos desenvolvidos nesta área, onde a maioria deles supõe distribuição uniforme sobre toda a região de estudo A (ROYLE *et al.*, 2013; BORCHERS; FEWSTER *et al.*, 2016). Observamos que a distribuição uniforme é um caso particular da nossa modelagem se fixarmos $\pi_r = (a_r - a_{r-1}) / (a_R - a_0)$ para todo $r \in \mathcal{R}$.

Consideraremos uma variável aleatória $W_{i,t}$ que define a localização instantânea do animal i na região A durante a t -ésima ocasião de amostragem, dado que ele esteja vivo nesta

ocasião. Dado $(U_i, V_i, S_i) = (u, v, s)$, assumimos, para todo $t \in \{u, u+1, \dots, v\}$,

$$W_{i,t} | (U_i, V_i, S_i) = (u, v, s) \stackrel{ind.}{\sim} \text{TNormal}(s, \sigma^2, (a_0, a_R]), \quad (5.2)$$

onde $\text{TNormal}(s, \sigma^2, (a_0, a_R])$ denota uma distribuição normal com média s e variância σ^2 truncada no intervalo $(a_0, a_R]$ cuja f.d.p. é dada por

$$f_W(w|s, \sigma) = \begin{cases} c \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{w-s}{\sigma}\right)^2\right\}, & \text{para } w \in A = (a_0, a_R], \\ 0, & \text{para } w \notin A = (a_0, a_R], \end{cases} \quad (5.3)$$

com a constante normalizadora c dada por

$$c^{-1} = \int_{a_0}^{a_R} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{z-s}{\sigma}\right)^2\right\} dz.$$

Assim, o parâmetro σ controla quão longe um animal pode movimentar-se em torno do seu centro de vida. Em outras palavras, este parâmetro estipula o tamanho da área de movimentação do animal na região A ao longo do período de estudo, isto é, menores valores para σ implicarão na diminuição da área de movimento e vice-versa. Como exemplo, utilizando a aproximação normal, teremos que um animal da população terá uma região de movimento de 6σ dentro da região A com aproximadamente 99% de probabilidade.

5.2.2 Processo de captura

O histórico de captura $\mathbf{Y}_i = (Y_{i,1}, \dots, Y_{i,T})$ é definido por

$$Y_{i,t} = \begin{cases} X_{i,t} \sum_{r \in \mathcal{R}} r I(W_{i,t} \in A_r), & \text{se } t \in \{U_i, U_i+1, \dots, V_i\}, \\ 0, & \text{se } t \notin \{U_i, U_i+1, \dots, V_i\}, \end{cases}$$

para todo $t \in \mathcal{T}$. Notemos que, por construção, temos que os vetores \mathbf{X}_i e $(S_i, W_{i,u}, \dots, W_{i,v})$ são condicionalmente independentes, dados $(U_i = u, V_i = v)$. Logo, para uma dada configuração $(U_i, V_i, S_i) = (u, v, s)$ e a partir das suposições de independência condicional, obtemos a distribuição conjunta do histórico \mathbf{Y}_i definindo as probabilidades de cada componente $Y_{i,t}$, como segue.

Primeiramente, consideramos $t \in \{u, u+1, \dots, v\}$. Assim, temos

$$\begin{aligned} P(Y_{i,t} = r | U_i = u, V_i = v, S_i = s, \theta, \sigma) &= \\ &= P(X_{i,t} = 1, W_{i,t} \in A_r | U_i = u, V_i = v, S_i = s, \theta, \sigma) \\ &= P(X_{i,t} = 1 | U_i = u, V_i = v, \theta) \times P(W_{i,t} \in A_r | U_i = u, V_i = v, S_i = s, \sigma) \\ &= \theta \times \int_{A_r} f_W(w|s, \sigma^2) dw \\ &= \theta \times \left[\frac{\Phi(a_r, s, \sigma) - \Phi(a_{r-1}, s, \sigma)}{\Phi(a_R, s, \sigma) - \Phi(a_0, s, \sigma)} \right], \end{aligned}$$

para todo $r \in \mathcal{R}$, com

$$\Phi(a, s, \sigma) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{z-s}{\sigma} \right)^2 \right\} dz,$$

e

$$\begin{aligned} P(Y_{i,t} = 0 | U_i = u, V_i = v, S_i = s, \theta, \sigma) &= P(X_{i,t} = 0 | U_i = u, V_i = v, S_i = s, \theta, \sigma) \\ &= (1 - \theta). \end{aligned}$$

Naturalmente, quando $t \notin \{u, u+1, \dots, v\}$, temos

$$P(Y_{i,t} = y | U_i = u, V_i = v, S_i = s, \theta, \sigma) = \begin{cases} 1, & \text{se } y = 0, \\ 0, & \text{se } y \in \mathcal{R}. \end{cases}$$

Seja $\mathcal{Y} = \{\mathbf{y} = (y_1, \dots, y_T) : y_t \in \{0\} \cup \mathcal{R}\}$ o conjunto de todos os possíveis históricos de captura. Adicionalmente, consideramos $Q(\mathbf{y}) = \{t : y_t \neq 0\}$ a função que determina quais elementos de \mathbf{y} são diferentes de zero, isto é, todas as épocas em que houve captura para um determinado histórico $\mathbf{y} \in \mathcal{Y}$. Assim, pela suposição de independência entre capturas, temos a probabilidade condicional de observar um histórico $\mathbf{y} \in \mathcal{Y}$ dada por

$$\begin{aligned} P(\mathbf{Y}_i = \mathbf{y} | U_i = u, V_i = v, S_i = s, \theta, \sigma) &= \\ &= \prod_{t \in \mathcal{Y}} P(Y_{i,t} = y_t | U_i = u, V_i = v, S_i = s, \theta, \sigma) \\ &= (1 - \theta)^{g_0(\mathbf{y}, u, v)} \theta^{v-u+1-g_0(\mathbf{y}, u, v)} \prod_{t \in Q(\mathbf{y})} \left\{ \left[\frac{\Phi(a_{y_t}, s, \sigma) - \Phi(a_{y_t-1}, s, \sigma)}{\Phi(a_R, s, \sigma) - \Phi(a_0, s, \sigma)} \right] I(u \leq t \leq v) \right\}, \end{aligned} \quad (5.4)$$

com $g_0(\mathbf{y}, u, v) = \sum_{t=u}^v I(y_t = 0)$. Em (5.4) definimos o produtório igual à 1 se $Q(\mathbf{y}) = \emptyset$, isto é, quando $\mathbf{y} = \mathbf{0}$.

Portanto, a partir das relações (2.3), (5.2) e (5.4), determinamos a função de probabilidades marginal de \mathbf{Y}_i por

$$\begin{aligned} P(\mathbf{Y}_i = \mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\pi}, \theta, \sigma) &= \\ &= \sum_{u=1}^{f(\mathbf{y})} \sum_{v=l(\mathbf{y}, u)}^T \int_{a_0}^{a_R} P(\mathbf{Y}_i = \mathbf{y} | U_i = u, V_i = v, S_i = s, \theta, \sigma) P(U_i = u, V_i = v | \boldsymbol{\beta}, \boldsymbol{\phi}) f_S(s | \boldsymbol{\pi}) ds \\ &= \sum_{u=1}^{f(\mathbf{y})} \sum_{v=l(\mathbf{y}, u)}^T \left\{ P(U_i = u, V_i = v | \boldsymbol{\beta}, \boldsymbol{\phi}) \times \right. \\ &\quad \left. \times \left[\sum_{r \in \mathcal{R}} \pi_r \int_{a_{r-1}}^{a_r} P(\mathbf{Y}_i = \mathbf{y} | U_i = u, V_i = v, S_i = s, \theta, \sigma) \left(\frac{1}{a_r - a_{r-1}} \right) ds \right] \right\}, \end{aligned} \quad (5.5)$$

para todo $\mathbf{y} \in \mathcal{Y}$, onde as funções f e l são definidas em (2.6) e (2.7), respectivamente. A relação (5.5) é a distribuição de probabilidades do vetor “observável” gerador dos históricos de captura dos animais. Pela suposição de independência, todos animais terão histórico de captura gerados independente e igualmente distribuídos por esta distribuição de probabilidades.

5.2.3 Função de verossimilhança

Dado $N \in \{1, 2, \dots\}$, sejam $\mathbf{y}_1, \dots, \mathbf{y}_N$ os históricos de captura de todos os animais da população, $\mathbf{y}_i \in \mathcal{Y}, i \in \mathcal{N}$, e considere as frequências

- $n_{[\mathbf{y}]} = \sum_{i \in \mathcal{N}} I(\mathbf{y}_i = \mathbf{y})$ o número de animais capturados com histórico de captura $\mathbf{y}, \mathbf{y} \in \mathcal{Y}$, e
- $n = \sum_{\mathbf{y} \in \mathcal{Y}_0} n_{[\mathbf{y}]}$ o número total de animais distintos capturados,

com $\mathcal{Y}_0 = \mathcal{Y} \setminus \{\mathbf{0}\}$. No final da etapa de amostragem a amostra observada é dada pelo conjunto $\mathbf{n} = \{n_{[\mathbf{y}]}; \mathbf{y} \in \mathcal{Y}_0\}$, uma vez que a frequência $n_{[\mathbf{0}]} = N - n$ referente ao número de animais não capturados não é observada. Pela suposição de independência condicional entre as capturas dos animais e dados $(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi})$, temos que \mathbf{n} é uma realização de uma distribuição multinomial com N ensaios e probabilidade de cada evento $\mathbf{y} \in \mathcal{Y}$ dado por (5.5). Logo, a função de verossimilhança para esses parâmetros é definida baseada na amostra \mathbf{n} e dada por

$$\begin{aligned} \mathcal{L}(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}; \mathbf{n}) &= p(\mathbf{n}|N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}) \\ &= \frac{N!}{(N-n)! \prod_{\mathbf{y} \in \mathcal{Y}_0} n_{[\mathbf{y}]}!} P(\mathbf{Y}_1 = \mathbf{0} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi})^{N-n} \times \\ &\quad \times \prod_{\mathbf{y} \in \mathcal{Y}_0} \left\{ P(\mathbf{Y}_1 = \mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}) \right\}^{n_{[\mathbf{y}]}} \end{aligned} \quad (5.6)$$

para $N \in \{n, n+1, \dots\}$, $\boldsymbol{\beta} \in \Theta_{\boldsymbol{\beta}}$, $\boldsymbol{\phi} \in \Theta_{\boldsymbol{\phi}}$, $\theta \in (0, 1)$, $\sigma \in (0, +\infty)$ e $\boldsymbol{\pi} \in \Theta_{\boldsymbol{\pi}}$.

5.3 Abordagem Bayesiana

Novamente, adotaremos a abordagem bayesiana para as inferências. Supondo independência condicional a priori entre os parâmetros, definimos as seguintes distribuições a priori

marginais

$$\begin{aligned}
N &\sim \text{BN}(q, p), \\
\phi_t &\sim \text{Beta}(a_{1,t}; a_{2,t}), \quad t \in \{1, \dots, T-1\}, \\
\theta &\sim \text{Beta}(b_1, b_2), \\
\boldsymbol{\beta} &\sim \text{Dirichlet}(c_1, \dots, c_T), \\
\sigma &\sim \text{Gama}(d_1, d_2), \\
\boldsymbol{\pi} &\sim \text{Dirichlet}(e_1, \dots, e_R),
\end{aligned}$$

com os valores dos hiperparâmetros $q, p, a_{i,t}, b_i, c_t, d_i, e_r$ positivos e conhecidos, onde $\text{BN}(q, p)$ denota uma distribuição Binomial Negativa com parâmetros de tamanho q e probabilidade de sucesso p .

Pelo Teorema de Bayes, a distribuição a posteriori conjunta para os parâmetros populacionais se expressa em termos proporcionais como o produto da função de verossimilhança e a estrutura a priori especificada acima, ou seja,

$$f(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi} | \mathbf{n}) \propto \mathcal{L}(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}; \mathbf{n}) f(N) f(\boldsymbol{\beta}) f(\boldsymbol{\phi}) f(\theta) f(\sigma) f(\boldsymbol{\pi}),$$

em que, por simplicidade de notação, denotamos genericamente por f a respectiva função (densidade) de probabilidades do(s) argumento(s) especificado(s).

Observamos que, como no Capítulo 3, temos uma função de verossimilhança “incompleta”, isto é, sua especificação não contém a informação completa dos indivíduos, como em quais as épocas os animais estavam vivos em alguns históricos de captura observados e em qual região está localizado cada centro de vida. Assim, utilizaremos novamente técnicas de aumento de dados, descritas a seguir.

5.3.1 Função de verossimilhança aumentada

Dado $N \in \{1, 2, \dots\}$, sejam (u_i, v_i) o par que representa a primeira e última ocasiões em que cada animal $i \in \mathcal{N}$ esteve presente na população e s_i a localização do seu centro de vida, $s_i \in A$. Definimos por $\tilde{n}_{r, [\mathbf{y}]}^{(u, v)}$ o número de animais com centro de vida localizado na região $r \in \mathcal{R}$ capturados com histórico $\mathbf{y} \in \mathcal{Y}$ e presentes na população entre (inclusive) as ocasiões u e v , $1 \leq u \leq f(\mathbf{y})$ e $l(\mathbf{y}, u) \leq v \leq T$, isto é,

$$\tilde{n}_{r, [\mathbf{y}]}^{(u, v)} = \sum_{i \in \mathcal{N}} I(\mathbf{y}_i = \mathbf{y}, u_i = u, v_i = v, s_i \in A_r).$$

Observamos que estas frequências devem satisfazer a seguinte restrição:

$$\sum_{r \in \mathcal{R}} \sum_{u=1}^{f(\mathbf{y})} \sum_{v=l(\mathbf{y}, u)}^T \tilde{n}_{r, [\mathbf{y}]}^{(u, v)} = n_{[\mathbf{y}]}, \quad \text{para todo } \mathbf{y} \in \mathcal{Y}. \quad (5.7)$$

Portanto, dado N , consideramos o conjunto de frequências

$$\tilde{\mathbf{n}} = \left\{ \tilde{n}_{r,[\mathbf{y}]}^{(u,v)}; \mathbf{y} \in \mathcal{Y}, 1 \leq u \leq f(\mathbf{y}), l(\mathbf{y}, u) \leq v \leq T, r \in \mathcal{R} \right\}$$

como o conjunto de dados aumentados para o conjunto $\mathbf{n} \cup \{n_{[0]}\}$.

Assim, a função de verossimilhança aumentada para os parâmetros $(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi})$, associada à informação completa $(\mathbf{n}, \tilde{\mathbf{n}})$, é definida por

$$\begin{aligned} \tilde{\mathcal{L}}(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}; \mathbf{n}, \tilde{\mathbf{n}}) &= p(\mathbf{n}, \tilde{\mathbf{n}} | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}) \\ &= p(\mathbf{n} | \tilde{\mathbf{n}}, N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}) p(\tilde{\mathbf{n}} | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}), \end{aligned} \quad (5.8)$$

onde

$$p(\mathbf{n} | \tilde{\mathbf{n}}, N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}) = \begin{cases} 1, & \text{se } n_{[\mathbf{y}]} = \sum_{r \in \mathcal{R}} \sum_{u=1}^{f(\mathbf{y})} \sum_{v=l(\mathbf{y},u)}^T \tilde{n}_{r,[\mathbf{y}]}^{(u,v)} \text{ para todo } \mathbf{y} \in \mathcal{Y}, \text{ ou} \\ 0, & \text{caso contrário,} \end{cases}$$

e $p(\tilde{\mathbf{n}} | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi})$ é a função de probabilidades de uma distribuição multinomial com N_r ensaios e probabilidade relacionada a frequência $\tilde{n}_{r,[\mathbf{y}]}^{(u,v)}$ dada por

$$\begin{aligned} P(\mathbf{Y}_1 = \mathbf{y}, U = u, V = v, S \in A_r | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}) &= \\ &= \int_{A_r} P(\mathbf{Y}_1 = \mathbf{y}, U = u, V = v | S = s, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma) f_S(s | \boldsymbol{\pi}) ds \\ &= \left\{ \int_{A_r} P(\mathbf{Y}_1 = \mathbf{y} | U = u, V = v, S = s, \theta, \sigma) \frac{\pi_r}{a_r - a_{r-1}} ds \right\} P(U = u, V = v | \boldsymbol{\beta}, \boldsymbol{\phi}), \end{aligned} \quad (5.9)$$

com as expressões para as funções de probabilidades de (U, V) e \mathbf{Y}_1 dadas por (2.3) e (5.4), respectivamente. Observamos que a função de verossimilhança aumentada é não nula se e somente se as restrições em (5.7) forem satisfeitas e, neste caso, será dada por

$$\begin{aligned} \tilde{\mathcal{L}}(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}; \mathbf{n}, \tilde{\mathbf{n}}) &= p(\tilde{\mathbf{n}} | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}) \\ &= \frac{N!}{\prod_{r \in \mathcal{R}} \prod_{\mathbf{y} \in \mathcal{Y}} \prod_{u=1}^{f(\mathbf{y})} \prod_{v=l(\mathbf{y},u)}^T \tilde{n}_{r,[\mathbf{y}]}^{(u,v)}!} \prod_{r \in \mathcal{R}} \prod_{\mathbf{y} \in \mathcal{Y}} \prod_{u=1}^{f(\mathbf{y})} \prod_{v=l(\mathbf{y},u)}^T \left\{ P(\mathbf{Y}_1 = \mathbf{y}, U = u, V = v, S \in A_r | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}) \right\}^{\tilde{n}_{r,[\mathbf{y}]}^{(u,v)}} \\ &\propto \frac{N!}{\prod_{r \in \mathcal{R}} \prod_{u=1}^T \prod_{v=u}^T \tilde{n}_{r,[0]}^{(u,v)}!} \theta^{\bar{n}} (1 - \theta)^{N - \bar{n}} \left\{ \prod_{t \in \mathcal{T}} \beta_t^{R_t} \right\} \left\{ \prod_{t=1}^{T-1} \phi_t^{W_t} (1 - \phi_t)^{N_t - W_t} \right\} \\ &\times \left\{ \prod_{r \in \mathcal{R}} [\pi_r]^{C_r} \right\} \left\{ \prod_{s \in \mathcal{R}} \prod_{r \in \mathcal{R}} \left[\int_{a_{s-1}}^{a_s} \frac{\Phi(a_r, z, \sigma) - \Phi(a_{r-1}, z, \sigma)}{\Phi(a_R, z, \sigma) - \Phi(a_0, z, \sigma)} dz \right]^{Q_{s,r}} \right\} \end{aligned} \quad (5.10)$$

com

- $W_t = \sum_{r \in \mathcal{R}} \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{u=1}^{\min\{t, f(\mathbf{y})\}} \sum_{v=\max\{t+1, l(\mathbf{y}, u)\}}^T \tilde{n}_{r,[\mathbf{y}]}^{(u,v)}$ o número de animais que sobreviveram entre as ocasiões t e $t + 1$, $t \in \{1, 2, \dots, T - 1\}$,

- $N_t = \sum_{r \in \mathcal{R}} \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{u=1}^{\min\{t, f(\mathbf{y})\}} \sum_{v=\max\{t, l(\mathbf{y}, u)\}}^T \tilde{n}_{r, [\mathbf{y}]}^{(u, v)}$ o número de animais vivos na ocasião t , e $\bar{N} = \sum_{t \in \mathcal{T}} N_t$,
- $R_t = \sum_{r \in \mathcal{R}} \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{v=l(\mathbf{y}, t)}^T \tilde{n}_{r, [\mathbf{y}]}^{(t, v)}$ o número de animais recrutados exatamente na ocasião t ,
- $\bar{n} = \sum_{\mathbf{y} \in \mathcal{Y}} \left\{ \sum_{t \in \mathcal{T}} I(y_t \neq 0) \right\} n_{[\mathbf{y}]}$ o número total de capturas (e recapturas) de animais,
- $C_r = \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{u=1}^{f(\mathbf{y})} \sum_{v=l(\mathbf{y}, u)}^T \tilde{n}_{r, [\mathbf{y}]}^{(u, v)}$ o número de animais com centro de vida na região r ,
- $Q_{s, r} = \sum_{t \in \mathcal{T}} \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{u=1}^{\min\{t, f(\mathbf{y})\}} \sum_{v=\max\{t, l(\mathbf{y}, u)\}}^T \left\{ \tilde{n}_{s, [\mathbf{y}]}^{(u, v)} I(y_t = r) \right\}$ o número de animais com centro de vida na região s capturados na região r .

Vale ressaltar que, dado uma configuração fixada para o conjunto de dados observados \mathbf{n} , a função de verossimilhança (5.6) pode ser obtida somando com relação a todos os possíveis valores das variáveis latentes $\tilde{\mathbf{n}}$ na função (5.8) tal que (5.7) seja válida.

Portanto, a distribuição a posteriori conjunta das quantidades desconhecidas (parâmetros e variáveis latentes) é proporcional ao produto da função de verossimilhança aumentada e das distribuições a priori especificadas, ou seja,

$$f(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}, \tilde{\mathbf{n}} | \mathbf{n}) \propto \tilde{\mathcal{L}}(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}; \mathbf{n}, \tilde{\mathbf{n}}) f(N) f(\boldsymbol{\beta}) f(\boldsymbol{\phi}) f(\theta) f(\sigma) f(\boldsymbol{\pi}).$$

5.3.2 Algoritmo MCMC

Para a implementação do algoritmo de Monte Carlo via Cadeia de Markov (MCMC) neste capítulo, consideramos dois blocos de atualização:

- Bloco I: atualiza os valores de $(N, \tilde{\mathbf{n}})$ a partir da distribuição condicional, dados os valores fixos de $(\boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \boldsymbol{\pi}, \sigma)$ e
- Bloco II: atualiza os valores de $(\boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \boldsymbol{\pi}, \sigma)$ a partir da distribuição condicional, dados os valores fixos de $(N, \tilde{\mathbf{n}})$.

Bloco I: atualização de $(N, \tilde{\mathbf{n}})$

Pela regra do produto de probabilidades, podemos escrever a função de probabilidades condicional de $(N, \tilde{\mathbf{n}})$ como

$$f(N, \tilde{\mathbf{n}} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \mathbf{n}) = f(N | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \mathbf{n}) \times f(\tilde{\mathbf{n}} | N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \mathbf{n}). \quad (5.11)$$

Logo, é possível dividir a geração conjunta de valores para $(N, \tilde{\mathbf{n}})$ em duas partes. Primeiramente, atualizamos N através de

$$\begin{aligned} f(N|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\pi}, \sigma, \theta, \mathbf{n}) &\propto \mathcal{L}(N, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\pi}, \sigma, \theta; \mathbf{n})f(N) \\ &\propto \frac{N!}{(N-n)!} \text{P}(\mathbf{Y}_1 = \mathbf{0}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\pi}, \sigma, \theta)^N \frac{\Gamma(q+N)}{\Gamma(N+1)} (1-p)^N \\ &= \frac{\Gamma(q+N)}{\Gamma(N-n+1)} \{ \text{P}(\mathbf{Y}_1 = \mathbf{0}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\pi}, \sigma, \theta)(1-p) \}^N, \end{aligned} \quad (5.12)$$

para $N \in \{n, n+1, \dots\}$. Assim, identificamos a seguinte distribuição condicional completa

$$(N-n)|\bullet \sim \text{BN}\left\{n+q, 1 - \text{P}(\mathbf{Y}_1 = \mathbf{0}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\pi}, \sigma, \theta)(1-p)\right\}. \quad (5.13)$$

Portanto, a atualização do valor para o parâmetro N é dada por $z+n$, onde z é o valor gerado de uma variável aleatória Z com distribuição Binomial negativa de parâmetros $n+q$ e $1 - \text{P}(\mathbf{Y} = \mathbf{0}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\pi}, \sigma, \theta)(1-p)$.

Por outro lado, dado um valor de N atualizado por (5.13) e com algumas manipulações algébricas, a geração das variáveis latentes $\tilde{\mathbf{n}}$ é dada pela seguinte distribuição condicional:

$$\begin{aligned} f(\tilde{\mathbf{n}}|N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi}, \mathbf{n}) &= \frac{\text{p}(\mathbf{n}, \tilde{\mathbf{n}}|N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi})}{\text{p}(\mathbf{n}|N, \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi})} \\ &= \prod_{\mathbf{y} \in \mathcal{Y}} \left\{ \frac{n_{[\mathbf{y}]!}}{\prod_{r \in \mathcal{R}} \prod_{u=1}^{f(\mathbf{y})} \prod_{v=l(\mathbf{y}, u)}^T \tilde{n}_{r, [\mathbf{y}]}^{(u, v)}} \times \right. \\ &\quad \left. \times \prod_{r \in \mathcal{R}} \prod_{u=1}^{f(\mathbf{y})} \prod_{v=l(\mathbf{y}, u)}^T \left[\frac{\text{P}(\mathbf{Y}_1 = \mathbf{y}, U = u, V = v, S \in A_r | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi})}{\text{P}(\mathbf{Y}_1 = \mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \sigma, \boldsymbol{\pi})} \right]^{\tilde{n}_{r, [\mathbf{y}]}^{(u, v)}} \right\}, \end{aligned} \quad (5.14)$$

sujeito à restrição (5.7) sobre as variáveis latentes. De fato, para um dado histórico $\mathbf{y} \in \mathcal{Y}$ com $n_{[\mathbf{y}]} > 0$, a distribuição condicional que gera as frequências latentes $\tilde{n}_{r, [\mathbf{y}]}^{(u, v)}$, para $1 \leq u \leq f(\mathbf{y})$, $l(\mathbf{y}, u) \leq v \leq T$, $r \in \mathcal{R}$, é uma distribuição Multinomial com $n_{[\mathbf{y}]}$ ensaios e probabilidade de cada evento correspondente proporcional à (5.9). Notamos que para uma frequência $n_{[\mathbf{y}]} = 0$, as frequências $\tilde{n}_{r, [\mathbf{y}]}^{(u, v)}$ serão todas iguais à 0 considerando a restrição em (5.7). Além disso, as frequências latentes são geradas de maneira independente para cada histórico $\mathbf{y} \in \mathcal{Y}$ decorrente do produto das funções de probabilidades multinomiais observado em (5.14).

Bloco II: atualização de $(\boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \boldsymbol{\pi}, \sigma)$

Neste bloco, consideramos valores fixos para $(N, \tilde{\mathbf{n}})$. Notemos que a função de verossimilhança aumentada (5.10) é fatorada no produto de distribuições independentes para os parâmetros $(\boldsymbol{\beta}, \boldsymbol{\phi}, \theta, \boldsymbol{\pi}, \sigma)$. Além disto, a menos do parâmetro σ , as distribuições correspondentes são conjugadas com respeito às suas distribuições a priori especificadas. Logo, obtemos as

respectivas distribuições condicionais completas independentes:

$$\phi_t | \bullet \sim \text{Beta}(W_t + a_{1,t}, N_t - W_t + a_{2,t}), \quad t \in \{1, \dots, T-1\}, \quad (5.15)$$

$$\beta | \bullet \sim \text{Dirichlet}(E_1 + c_1, \dots, E_T + c_T), \quad (5.16)$$

$$\theta | \bullet \sim \text{Beta}(\bar{n} + b_1, \bar{N} - \bar{n} + b_2), \quad (5.17)$$

$$\pi | \bullet \sim \text{Dirichlet}(C_1 + e_1, \dots, C_T + e_T). \quad (5.18)$$

Para o parâmetro σ , consideremos a distribuição condicional

$$f(\sigma | \tilde{\mathbf{n}}) = c \left\{ \prod_{s \in \mathcal{R}} \prod_{r \in \mathcal{R}} \left[\int_{a_{s-1}}^{a_s} \frac{\Phi(a_r, z, \sigma) - \Phi(a_{r-1}, z, \sigma)}{\Phi(a_R, z, \sigma) - \Phi(a_0, z, \sigma)} dz \right]^{Q_{s,r}} \right\} \sigma^{d_1-1} \exp \left\{ -\frac{\sigma}{d_2} \right\}, \quad (5.19)$$

onde c é uma constante que não depende de σ . Para atualização deste parâmetro, sejam σ^* o valor atual da cadeia e $\tilde{\sigma} = \exp(x)$ um valor (proposto) com x gerado de uma distribuição normal com média $\log(\sigma^*)$ e variância conhecida κ . Assim, o valor de σ é atualizado de σ^* para $\tilde{\sigma}$ com a seguinte probabilidade de aceitação Metropolis-Hasting:

$$\alpha(\sigma^*, \tilde{\sigma}) = \min \left\{ 1, \frac{f(\tilde{\sigma} | \tilde{\mathbf{n}})}{f(\sigma^* | \tilde{\mathbf{n}})} \right\}. \quad (5.20)$$

Finalmente, o algoritmo MCMC pode ser implementado a partir de um processo iterativo baseado nas distribuições condicionais dadas por (5.13), (5.14), (5.15), (5.16), (5.17), (5.18), e (5.19). As rotinas foram implementadas no *software* R (R Development Core Team, 2014) e utilizamos a função `integrate` do pacote `stats` para avaliar a integral em (5.19).

5.4 Estudo de simulação

Nesta seção realizaremos um estudo computacional para verificar a qualidade das estimativas bayesianas para os parâmetros populacionais, a partir do modelo proposto e do processo inferencial apresentado. O estudo será feito através de dados artificiais de captura-recaptura gerados sob diferentes cenários e avaliaremos numericamente as propriedades frequentistas dos estimadores, de acordo com as medidas de desempenho definidas na seção 3.3.1.

5.4.1 Geração dos dados

Para o estudo numérico geramos dados de captura-recaptura a partir de 18 cenários diferentes, variando os valores reais do tamanho populacional (N), da probabilidade de captura (θ) e do tamanho do deslocamento dos animais na região de estudo (σ), como mostrado na Tabela 21. Tais valores foram adotados a fim de avaliar se estas quantidades podem influenciar a precisão das estimativas dos parâmetros. Especificamente, verificaremos o efeito do baixo número de animais da população ($N = 300$) nas medidas de desempenho adotadas e compará-las com tamanhos populacionais relativamente maiores ($N = 600$ e $N = 1200$). Da mesma

forma, verificaremos o impacto do aumento da probabilidade de captura nessas medidas, e se a amplitude total de movimentação dos animais exerce alguma influência nas estimativas pontuais e intervalares dos parâmetros. Os valores para o parâmetro σ foram escolhidos para que tenhamos movimentação de cada animal dentro de uma região de aproximadamente $6\sigma = 120, 180$ e 240 de amplitude, com 99% de probabilidade. Para todos os cenários, consideramos $T = 6$ ocasiões de amostragem.

Tabela 21 – Cenários utilizados no segundo estudo de simulação.

Cenário	N	θ	σ	Cenário	N	θ	σ	Cenário	N	θ	σ
1	300	0,20	20	7	600	0,20	20	13	1200	0,20	20
2	300	0,20	30	8	600	0,20	30	14	1200	0,20	30
3	300	0,20	40	9	600	0,20	40	15	1200	0,20	40
4	300	0,30	20	10	600	0,30	20	16	1200	0,30	20
5	300	0,30	30	11	600	0,30	30	17	1200	0,30	30
6	300	0,30	40	12	600	0,30	40	18	1200	0,30	40

Para obter as medidas de desempenho, $M = 200$ amostras foram geradas de cada cenário. Adicionalmente, a fim de propiciar um maior número de cenários para estudo, as probabilidades de nascimento e sobrevivência foram simuladas aleatoriamente em cada uma das $M = 200$ amostras. Especificamente, em cada amostra $m = 1, \dots, 200$, geramos $\phi = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)$ por

$$\begin{aligned} \phi_1 &\sim \text{Beta}(146, 3; 62, 7), & \phi_2 &\sim \text{Beta}(139, 875; 46, 625), & \phi_3 &\sim \text{Beta}(95, 6; 143, 4), \\ \phi_4 &\sim \text{Beta}(80, 1; 8, 9) & \text{e} & & \phi_5 &\sim \text{Beta}(127, 2; 31, 8), \end{aligned}$$

resultando os respectivos valores esperados para estes parâmetros $\mathbb{E}(\phi_1) = 0,7$, $\mathbb{E}(\phi_2) = 0,75$, $\mathbb{E}(\phi_3) = 0,4$, $\mathbb{E}(\phi_4) = 0,90$ e $\mathbb{E}(\phi_5) = 0,8$, com variância comum igual a 0.001. Esses valores esperados foram escolhidos com a finalidade de verificar o comportamento dos estimadores em períodos de baixa e alta taxas de sobrevivência. Da mesma forma, geramos em cada amostra um diferente vetor de probabilidades de recrutamento definido por

$$\beta \sim \text{Dirichlet}(40, 15, 10, 20, 10, 5),$$

condizente com valores esperados de $\mathbb{E}(\beta_1) = 0,4$, $\mathbb{E}(\beta_2) = 0,15$, $\mathbb{E}(\beta_3) = 0,1$, $\mathbb{E}(\beta_4) = 0,2$, $\mathbb{E}(\beta_5) = 0,1$ e $\mathbb{E}(\beta_6) = 0,05$. A distribuição do centro de vida foi obtida a partir da função densidade de probabilidade apresentada na Figura 14 a seguir. Esta função foi construída a partir da mistura de duas distribuições normais, com médias 150 e 350, variância comum igual a 50 e pesos dados por 0,6 e 0,4 para cada componente, respectivamente. Este cenário foi adotado para avaliar o desempenho dos estimadores em cenários com regiões de grande concentração de animais, isto é, populações que vivem em aglomerações na região pesquisada. Para simular o processo de amostragem fixamos $R = 18$ regiões de amostragem igualmente espaçadas de

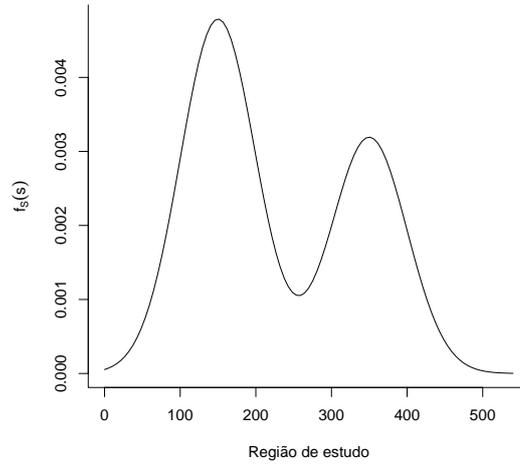


Figura 14 – Distribuição utilizada no estudo para geração dos locais de centros de vida.

amplitude 30, isto é, $A_r = (30r - 30, 30r]$, $r \in \mathcal{R}$, resultando na região $A = (0, 540]$. Neste caso, a partir dos valores de σ adotados, temos que os animais terão aproximadamente uma área (máxima) de movimento de 4, 6 e 8 regiões. A definição destas regiões foi baseada no problema prático resolvido posteriormente em nossa aplicação com dados reais. Desta forma, os pesos foram obtidos por

$$\pi_r \propto \int_{30r-30}^{30r} \left[0,6 \exp \left\{ -\frac{1}{2} \left(\frac{w-150}{50} \right)^2 \right\} + 0,4 \exp \left\{ -\frac{1}{2} \left(\frac{w-350}{50} \right)^2 \right\} \right] dw,$$

para cada $r \in \mathcal{R}$.

Uma amostra observada pode ser obtida a partir dos seguintes passos de geração de dados:

1. Para cada $i \in \mathcal{N}$, gere:

- a) $(u_i, v_i) \sim (2.3)$
- b) $s_i \sim (5.1)$
- c) $\mathbf{y}_i | (u_i, v_i, s_i) \sim (5.4)$

2. Para cada $\mathbf{y} \in \mathcal{Y}_0$, faça:

$$n_{[\mathbf{y}]} = \sum_{i=1}^N I(\mathbf{y}_i = \mathbf{y});$$

3. Faça: $\mathbf{n} = \{n_{[\mathbf{y}]}; \mathbf{y} \in \mathcal{Y}_0\}$

5.4.2 Resultados numéricos

Com respeito à análise bayesiana a seguir, os valores dos hiperparâmetros foram escolhidos para que as distribuições a priori sejam pouco informativas. Especificamos

$$\begin{aligned}\phi_t &\sim \text{Beta}(1, 1), \quad t \in \{1, 2, \dots, 5\}, \\ \theta &\sim \text{Beta}(1, 1), \\ \boldsymbol{\beta} &\sim \text{Dirichlet}(1, 1, 1, 1, 1, 1), \\ \boldsymbol{\pi} &\sim \text{Dirichlet}(1, 1, \dots, 1, 1), \\ \sigma &\sim \text{Gamma}(0, 01; 0, 001), \\ N &\sim \text{BN}(q = 1, p = 1000000^{-1}).\end{aligned}$$

Por exemplo, para o parâmetro σ , os valores dos hiperparâmetros a priori faz com que tenhamos uma variância igual a 10.000, refletindo em uma distribuição vaga sobre os reais positivos. Analogamente, fixamos $q = 1$ e um baixo valor para $p = 1000000^{-1}$ resultando em $p(N) \approx 1$, isto é, uma distribuição aproximadamente uniforme sobre $\{0, 1, 2, \dots\}$.

Determinamos as inferências pontuais e intervalares dos parâmetros a partir destas amostras geradas. Para o processo iterativo do algoritmo MCMC, geramos 100.000 iterações para cada uma das $M = 200$ amostras, descartamos as 50.000 iniciais como aquecimento da cadeia e utilizamos saltos de tamanho 10 do restante resultando uma amostra final de $B = 5.000$ valores para cada parâmetro para obter seus resumos a posteriori. Para a construção dos intervalos HPD, adotamos um nível de credibilidade de 95%. Para geração de candidatos para o parâmetro σ , fixamos uma variância de 0,25 e a probabilidade (5.20) apresentou taxas de aceitação, em geral, entre 30 e 50%.

5.4.3 Exemplo

Para fins ilustrativos, descrevemos a seguir algumas inferências a posteriori obtidas com base em uma amostra gerada segundo o Cenário 11. Primeiramente, apresentamos na Tabela 22 os valores médios, medianos, desvios-padrões (d.p.) e intervalos HPD com 95% de credibilidade a posteriori para os números de animais vivos nas ocasiões de amostragem, N_1, N_2, \dots, N_6 . Estas medidas foram calculadas utilizando os valores gerados para os dados aumentados, $\tilde{\mathbf{n}}$, a partir da relação

$$N_t = \sum_{r \in \mathcal{R}} \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{u=1}^{\min\{t, f(\mathbf{y})\}} \sum_{v=\max\{t, l(\mathbf{y}, u)\}}^T \tilde{n}_{r, [\mathbf{y}]}^{(u, v)},$$

como previamente definida na página 83. Observamos que as estimativas pontuais para estas abundâncias estão próximas dos números verdadeiros de animais vivos nas ocasiões, exibidos na coluna N_t^* , e que os intervalos de credibilidade contém os verdadeiros valores.

A Figura 15 exhibe as aproximações das funções de probabilidades a posteriori destas quantidades utilizando os valores gerados da amostra MCMC onde podemos observar distri-

Tabela 22 – Resumos a posteriori para os tamanhos populacionais em cada ocasião de amostragem para uma amostra gerada do Cenário 11.

Abundância	N_t^*	média	mediana	d.p	$IC_b(95\%)$
N_1	207	235,27	233,5	31,80	[172, 296]
N_2	265	263,01	262,0	29,36	[206, 321]
N_3	293	295,88	294,0	31,15	[236, 360]
N_4	239	248,85	247,0	28,21	[198, 306]
N_5	248	297,94	296,0	32,89	[232, 359]
N_6	240	251,37	250,0	29,05	[192, 307]

buições aproximadamente simétricas resultando em valores da média e mediana a posteriori próximos entre si.

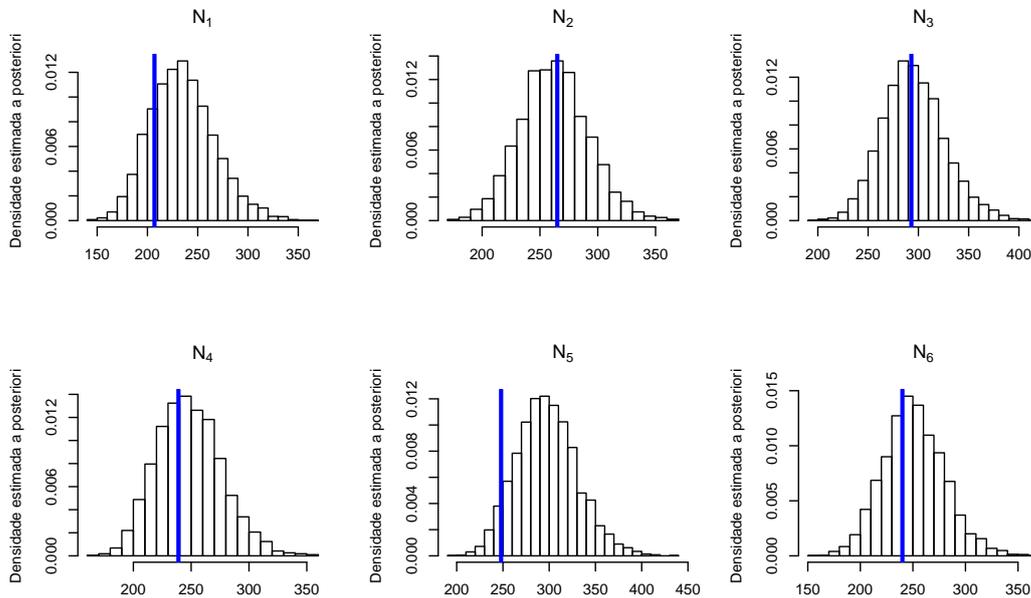


Figura 15 – Aproximações das funções de probabilidades a posteriori para os tamanhos populacionais em cada ocasião de amostragem. Linhas verticais azuis indicam o verdadeiro valor do parâmetro.

A seguir, consideraremos a estimação do número de animais vivos em cada ocasião por região de amostragem, $N_{t,r}$. Note que, no algoritmo proposto para as geração das variáveis latentes, obtemos apenas as frequências $\tilde{n}_{r,[y]}^{(u,v)}$: número de animais capturados com histórico y , centro de vida localizado na r -ésima ocasião e vivo entre (inclusive) as ocasiões u e v . Assim, calculamos os valores de $N_{r,t}$ com o seguinte esquema de geração estocástica:

1. Inicie com $N_{t,r} = 0, \forall t \in \mathcal{T}, r \in \mathcal{R}$;
2. Para $\tilde{n}_{r,[y]}^{(u,v)} > 0$, faça para cada $i \in \{1, 2, \dots, \tilde{n}_{r,[y]}^{(u,v)}\}$:
 - para $t \in \{u, u + 1, \dots, v\}$:

- (i) $w = y_t$, se $y_t \neq 0$, ou
- (ii) escolha $w \in \mathcal{R}$ com probabilidade proporcional à
- $$\int_{a_{r-1}}^{a_r} \frac{\Phi(a_w, s, \sigma) - \Phi(a_{w-1}, s, \sigma)}{\Phi(a_R, s, \sigma) - \Phi(a_0, s, \sigma)} ds, \text{ se } y_t = 0,$$

- Faça $N_{t,w} = N_{t,w} + 1$.

Exibimos na Figura 16 os números médios estimados para $N_{t,r}$ onde a primeira imagem representa os números verdadeiros destas quantidades e a segunda as estimativas utilizando o modelo proposto. Notamos que a metodologia produziu estimativas a posteriori próximas da verdadeira distribuição espaço-temporal dos animais da população simulada identificando principalmente quais regiões têm maiores e menores concentrações de animais.

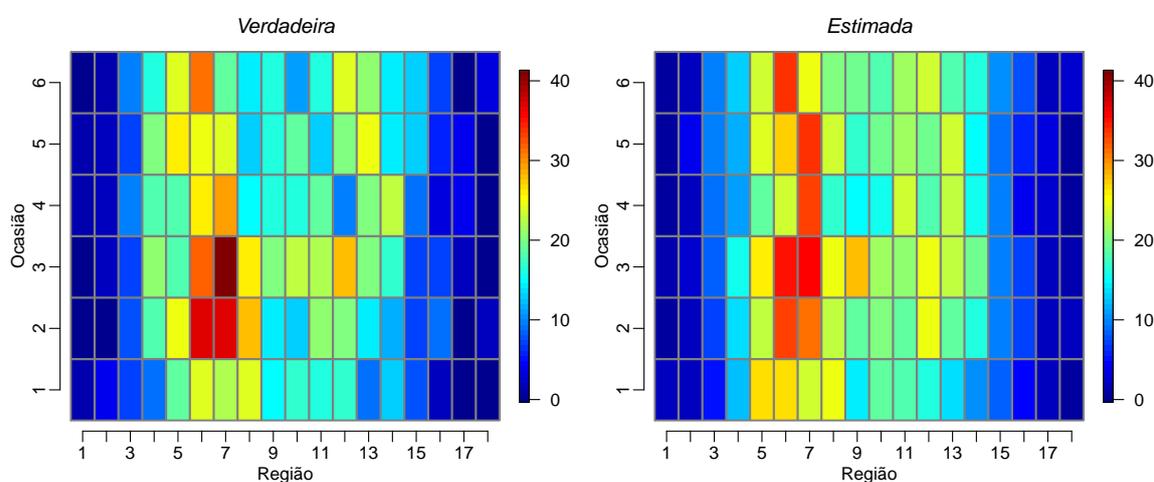


Figura 16 – Valores médios estimados a posteriori para os tamanhos populacionais em cada ocasião e região de amostragem para uma amostra gerada sob o cenário 11.

Para comparação, calculamos estas abundâncias utilizando uma amostra gerada segundo o Cenário 17. Os valores médios estimados são exibidos na Figura 17 onde observamos novamente estimativas próximas do verdadeiro comportamento territorial da população ao longo do tempo. Este gráfico apresentou visualmente melhor precisão das estimativas quando comparado ao cenário anterior. De fato, com o aumento do número de animais da população temos, em média, maior frequência de animais capturados e, conseqüentemente, mais informações amostrais para realização das inferências. Assim, notamos que o modelo apresenta comportamento esperado neste aspecto.

5.4.4 Qualidade das estimativas

Iniciamos a descrição dos resultados exibindo na Tabela 23 as medidas de desempenho dos estimadores pontuais $\mathbb{E}(N)$ e $\mathbb{M}(N)$ a posteriori para o tamanho da superpopulação, N , e do estimador intervalar por meio da amplitude média do intervalo e sua probabilidade de cobertura.

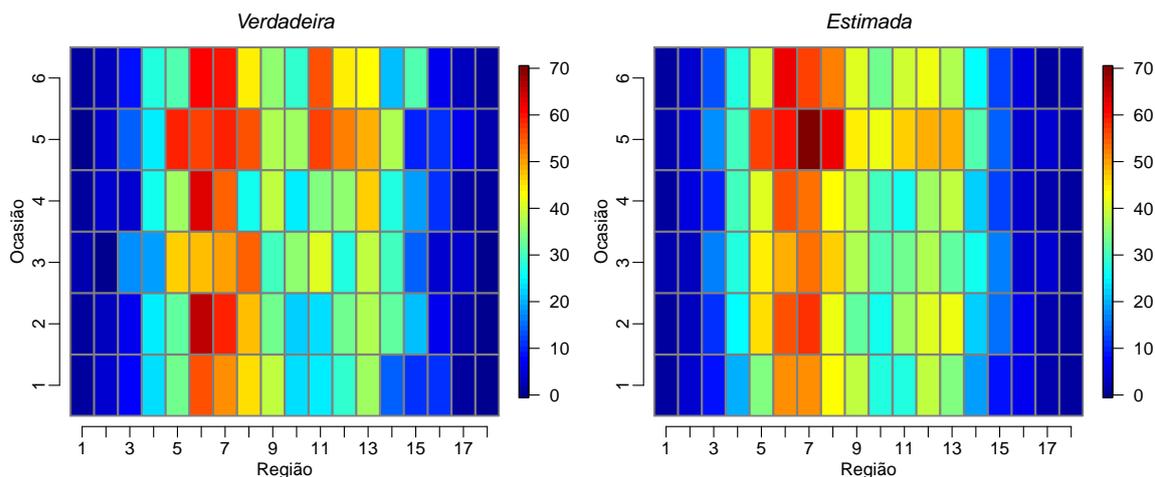


Figura 17 – Valores médios estimados a posteriori para os tamanhos populacionais em cada ocasião e região de amostragem para uma amostra gerada o cenário 17.

Tabela 23 – Medidas de desempenho dos estimadores bayesianos do parâmetro N .

Cenário	Viés		ViésR(%)		EAM		\sqrt{EQM}		AM	PC
	$\mathbb{E}(N)$	$\mathbb{M}(N)$	$\mathbb{E}(N)$	$\mathbb{M}(N)$	$\mathbb{E}(N)$	$\mathbb{M}(N)$	$\mathbb{E}(N)$	$\mathbb{M}(N)$		
1	22,86	14,86	7,62	4,95	45,88	42,09	60,78	55,56	225,20	96,0
2	22,83	14,80	7,61	4,93	45,78	41,93	60,44	55,20	224,94	96,0
3	22,99	14,96	7,66	4,99	46,05	42,12	61,04	55,76	224,82	94,5
4	11,39	8,12	3,80	2,71	26,56	25,53	34,33	32,79	131,71	96,0
5	11,38	8,02	3,79	2,67	26,55	25,57	34,29	32,78	132,13	95,0
6	11,26	7,95	3,75	2,65	26,51	25,40	34,18	32,58	131,75	95,5
7	20,72	13,53	3,45	2,25	61,41	59,35	77,95	74,86	295,77	97,0
8	20,62	13,48	3,44	2,25	61,33	59,37	77,57	74,51	296,06	96,5
9	20,82	13,65	3,47	2,27	61,40	59,48	78,11	75,05	295,42	96,5
10	8,34	5,32	1,39	0,89	33,18	32,64	41,35	40,44	177,19	96,5
11	8,20	5,09	1,37	0,85	32,97	32,48	41,03	40,19	177,11	97,0
12	8,41	5,34	1,40	0,89	33,11	32,58	41,21	40,39	177,20	97,5
13	18,57	11,64	1,55	0,97	92,28	90,41	113,91	111,87	407,71	94,5
14	19,15	12,09	1,60	1,01	92,44	90,43	114,13	111,88	407,72	94,0
15	19,11	12,10	1,59	1,01	92,50	90,57	113,92	111,88	407,50	94,0
16	5,00	1,92	0,42	0,16	48,98	48,70	61,11	60,62	247,81	95,5
17	4,92	1,86	0,41	0,15	48,98	48,62	61,02	60,54	246,97	95,5
18	5,04	1,92	0,42	0,16	48,95	48,70	61,16	60,68	247,59	96,0

Primeiramente, notamos um viés positivo para ambos estimadores pontuais, embora este viés seja menor para o valor mediano. De fato, a distribuição de probabilidades a posteriori para o tamanho populacional, em geral, tem um decaimento mais suave à direita resultando uma assimetria nesta direção. Para ilustrar este comportamento, apresentamos na Figura 18 a distribuição a posteriori aproximada para o parâmetro N a partir de uma amostra gerada no primeiro cenário. Contudo, o viés relativo não ultrapassou 8% em nenhum cenário para a média a posteriori e 5% para a mediana a posteriori, com valores se aproximando de zero na medida

em que o tamanho da população cresce ou quando aumenta a probabilidade de captura. De fato, nestes dois casos, o número de animais capturados na amostra aumenta e, conseqüentemente, mais dados amostrais são utilizados para obter as inferências. A probabilidade de captura mais elevada ($\theta = 0,3$) implicou também em uma diminuição dos erros quadrático e absoluto médios, quando comparado com $\theta = 0,2$, bem como na diminuição, em média, de pelo menos 35% da amplitude do intervalo de credibilidade HPD. Além disto, a probabilidade de cobertura permaneceu próxima do valor nominal adotado na construção do intervalo HPD de 95% de credibilidade. A variação do parâmetro σ aparentemente não exerce influência sobre os valores das medidas de desempenho dos estimadores.

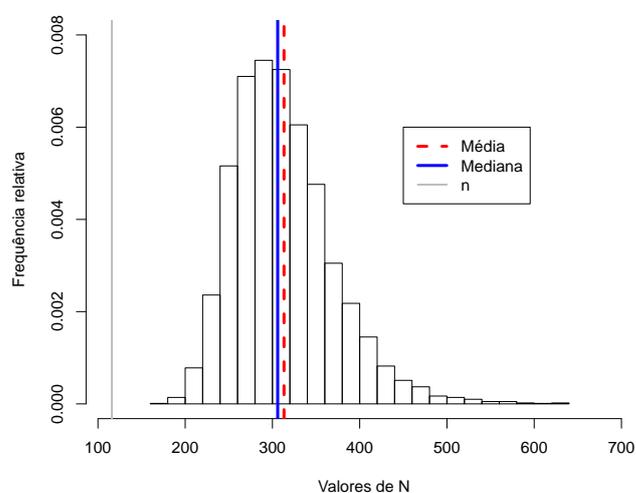


Figura 18 – Distribuição a posteriori marginal aproximada do parâmetro N em uma amostra gerada sob o cenário 1.

Para os demais parâmetros do modelo consideraremos a média a posteriori como estimador pontual na apresentação das medidas descritivas e ressaltamos que o uso da mediana como estimador pontual apresentou resultados similares aos que serão discutidos a seguir.

Na Tabela 24 são apresentadas as medidas de desempenho dos estimadores pontual e intervalar para a probabilidade de captura. Os resultados exibem viés positivo, mas com tendência para zero quando aumentamos o tamanho da população. O mesmo comportamento foi observado para os erros médios. Para os intervalos de credibilidade, notamos que a verdadeira probabilidade de captura não teve impacto significativo na amplitude média dos intervalos de credibilidade, mas houve um ganho razoável (menor amplitude) com o aumento do tamanho da população. Em todos os cenários, a probabilidade de cobertura do estimador intervalar foi superior ao nível nominal adotado de 95%.

Na Tabela 25 são apresentadas as medidas de desempenho dos estimadores pontual e intervalar para o vetor de probabilidades de recrutamento, β . Para o estimador pontual, notamos que vieses mais elevados foram observados nos primeiros cenários referentes ao tamanho

Tabela 24 – Medidas de desempenho dos estimadores bayesianos para o parâmetro θ .

Cenário	Viés	ViésR(%)	EAM	EQM	AM	PC
1	0,0147	7,3268	0,0352	0,0020	0,1973	99,0
2	0,0147	7,3368	0,0351	0,0020	0,1975	98,5
3	0,0145	7,2434	0,0353	0,0020	0,1969	99,0
4	0,0085	2,8345	0,0369	0,0021	0,1880	98,0
5	0,0085	2,8204	0,0369	0,0021	0,1882	98,0
6	0,0086	2,8756	0,0370	0,0021	0,1883	98,5
7	0,0092	4,5826	0,0252	0,0011	0,1424	98,0
8	0,0094	4,6996	0,0251	0,0011	0,1434	98,0
9	0,0091	4,5377	0,0250	0,0011	0,1429	97,5
10	0,0034	1,1246	0,0236	0,0009	0,1365	99,5
11	0,0034	1,1437	0,0234	0,0008	0,1362	99,0
12	0,0033	1,1025	0,0235	0,0009	0,1363	99,0
13	0,0049	2,4569	0,0204	0,0007	0,1043	96,0
14	0,0046	2,2888	0,0203	0,0007	0,1039	95,5
15	0,0046	2,2845	0,0204	0,0007	0,1041	96,0
16	0,0013	0,4299	0,0182	0,0006	0,1011	97,0
17	0,0011	0,3806	0,0183	0,0006	0,1012	97,5
18	0,0013	0,4305	0,0181	0,0006	0,1015	97,5

relativamente baixo do número de animais da população. Todavia, as medidas de desempenho dos estimadores pontuais são próximas de zero com o aumento do número de animais da população ou da probabilidade de captura. Para o estimador intervalar, suas medidas de desempenho se mostraram coerentes com o aumento da probabilidade de captura e tamanho populacional no sentido de que houve a diminuição da amplitude do intervalo, mantendo aproximadamente constante uma probabilidade de cobertura próxima ao verdadeiro nível de credibilidade adotado em sua construção, embora a maioria dos cenários a probabilidade de cobertura para o parâmetro β_4 tenha valor abaixo do esperado.

Na Tabela 26 apresentamos as medidas de desempenho dos estimadores pontual e intervalar para o vetor de probabilidades de sobrevivência, ϕ . Primeiramente, observamos que os vieses mais elevados são apresentados nos cenários com menores tamanhos populacionais. Em resumo, temos valores subestimados quando o valor esperado do parâmetro na geração dos dados é superior à 0,5 e sobreestimados no caso inferior à 0,5. Este último caso dá-se especificamente para o parâmetro ϕ_3 com $\mathbb{E}(\phi_3) = 0,4$. Entretanto, a partir do cenário 10, baixos valores para os vieses foram apresentados para todos os parâmetros e as probabilidades de cobertura ficaram próximas ao valor nominal esperado de 95%.

Tabela 25 – Medidas de desempenho dos estimadores bayesianos do vetor de parâmetro β .

Cenário	β_1			β_2			β_3					
	Viés	EAM	PC	Viés	EAM	PC	Viés	EAM	PC			
1	-0,0392	0,0553	0,2813	95,5	0,0001	0,0456	0,2644	96,0	0,0211	0,0410	0,2281	98,0
2	-0,0395	0,0553	0,2811	95,5	-0,0001	0,0452	0,2639	95,5	0,0213	0,0414	0,2282	97,0
3	-0,0393	0,0556	0,2814	95,5	0,0003	0,0453	0,2648	96,0	0,0209	0,0413	0,2280	97,0
4	-0,0214	0,0473	0,2420	94,5	-0,0039	0,0421	0,2324	94,0	0,0141	0,0354	0,1984	95,0
5	-0,0214	0,0472	0,2425	94,5	-0,0040	0,0422	0,2323	94,5	0,0140	0,0357	0,1977	95,5
6	-0,0213	0,0476	0,2424	95,0	-0,0043	0,0423	0,2318	93,5	0,0141	0,0356	0,1978	95,0
7	-0,0230	0,0458	0,2303	96,5	-0,0045	0,0387	0,2296	96,0	0,0147	0,0377	0,1969	96,5
8	-0,0235	0,0456	0,2306	96,0	-0,0041	0,0391	0,2296	96,5	0,0148	0,0377	0,1973	96,5
9	-0,0230	0,0456	0,2307	96,0	-0,0040	0,0387	0,2302	96,5	0,0142	0,0371	0,1970	96,5
10	-0,0103	0,0349	0,1911	93,0	-0,0047	0,0344	0,1953	96,0	0,0068	0,0339	0,1633	92,5
11	-0,0104	0,0349	0,1914	92,5	-0,0047	0,0344	0,1954	95,5	0,0070	0,0340	0,1635	92,0
12	-0,0102	0,0348	0,1912	94,0	-0,0046	0,0346	0,1950	96,0	0,0067	0,0340	0,1629	93,0
13	-0,0113	0,0346	0,1826	95,0	-0,0073	0,0370	0,1884	94,0	0,0065	0,0330	0,1648	95,5
14	-0,0109	0,0344	0,1828	95,0	-0,0075	0,0367	0,1895	95,0	0,0062	0,0331	0,1647	94,5
15	-0,0106	0,0345	0,1827	95,0	-0,0081	0,0370	0,1888	93,5	0,0063	0,0327	0,1651	95,0
16	-0,0033	0,0285	0,1462	95,5	-0,0049	0,0300	0,1524	94,5	-0,0004	0,0274	0,1296	93,5
17	-0,0031	0,0285	0,1461	95,5	-0,0048	0,0300	0,1523	96,0	-0,0005	0,0274	0,1305	93,5
18	-0,0034	0,0284	0,1464	95,5	-0,0046	0,0301	0,1520	96,5	-0,0006	0,0272	0,1296	93,5
β_4												
1	-0,0500	0,0596	0,2433	85,5	0,0267	0,0430	0,2250	97,0	0,0412	0,0454	0,1849	97,5
2	-0,0498	0,0595	0,2427	86,0	0,0266	0,0426	0,2252	97,5	0,0415	0,0456	0,1847	96,5
3	-0,0499	0,0594	0,2436	86,5	0,0271	0,0429	0,2260	96,5	0,0410	0,0453	0,1842	97,5
4	-0,0290	0,0447	0,2118	91,0	0,0121	0,0347	0,1887	96,5	0,0281	0,0345	0,1480	96,0
5	-0,0290	0,0448	0,2118	90,0	0,0122	0,0347	0,1889	97,0	0,0282	0,0345	0,1480	96,0
6	-0,0289	0,0448	0,2113	91,0	0,0121	0,0346	0,1889	97,5	0,0283	0,0347	0,1481	95,5
7	-0,0359	0,0500	0,2082	85,5	0,0197	0,0401	0,1916	95,5	0,0290	0,0345	0,1534	98,0
8	-0,0358	0,0498	0,2082	86,5	0,0195	0,0398	0,1912	95,5	0,0291	0,0345	0,1538	97,5
9	-0,0356	0,0497	0,2080	86,0	0,0194	0,0395	0,1913	95,0	0,0291	0,0346	0,1541	97,5
10	-0,0136	0,0338	0,1657	96,0	0,0058	0,0282	0,1569	96,5	0,0159	0,0249	0,1188	96,0
11	-0,0137	0,0338	0,1654	96,0	0,0061	0,0281	0,1566	96,5	0,0158	0,0250	0,1185	96,5
12	-0,0136	0,0339	0,1656	96,0	0,0059	0,0280	0,1565	97,0	0,0158	0,0248	0,1188	95,5
13	-0,0211	0,0361	0,1638	91,0	0,0166	0,0335	0,1596	92,5	0,0166	0,0256	0,1237	97,0
14	-0,0210	0,0363	0,1636	90,5	0,0169	0,0340	0,1593	93,0	0,0164	0,0255	0,1239	98,0
15	-0,0206	0,0359	0,1645	92,5	0,0165	0,0336	0,1604	93,5	0,0165	0,0255	0,1237	97,5
16	-0,0035	0,0260	0,1224	95,5	0,0054	0,0241	0,1246	93,5	0,0067	0,0194	0,0951	94,0
17	-0,0033	0,0262	0,1226	95,5	0,0053	0,0239	0,1251	94,0	0,0065	0,0195	0,0952	94,0
18	-0,0035	0,0261	0,1225	95,5	0,0055	0,0241	0,1250	94,5	0,0066	0,0195	0,0954	95,0

Tabela 26 – Medidas de desempenho dos estimadores bayesianos do vetor de parâmetro ϕ .

Cenário	ϕ_1				ϕ_2				ϕ_3				PC
	Viés	EAM	AM	PC	Viés	EAM	AM	PC	Viés	EAM	AM	PC	
1	-0,0281	0,1050	0,5726	96,5	-0,0843	0,1125	0,5547	97,0	0,1175	0,1486	0,5787	86,0	
2	-0,0280	0,1063	0,5709	96,5	-0,0848	0,1131	0,5554	96,5	0,1172	0,1475	0,5787	86,5	
3	-0,0276	0,1055	0,5725	96,5	-0,0846	0,1119	0,5560	97,5	0,1178	0,1483	0,5804	87,0	
4	0,0003	0,0914	0,4745	95,5	-0,0418	0,0910	0,4557	95,0	0,0743	0,1057	0,4275	93,0	
5	0,0004	0,0914	0,4743	95,5	-0,0423	0,0909	0,4556	95,0	0,0748	0,1060	0,4290	93,5	
6	0,0009	0,0918	0,4731	95,0	-0,0420	0,0911	0,4556	94,0	0,0740	0,1053	0,4285	93,0	
7	0,0035	0,0927	0,4911	95,5	-0,0498	0,0941	0,4672	93,5	0,0791	0,1075	0,4506	93,5	
8	0,0030	0,0931	0,4923	96,0	-0,0509	0,0944	0,4671	93,0	0,0793	0,1076	0,4523	94,5	
9	0,0035	0,0926	0,4922	97,0	-0,0493	0,0941	0,4672	94,5	0,0789	0,1074	0,4503	93,5	
10	0,0172	0,0833	0,3958	94,0	-0,0149	0,0707	0,3784	95,0	0,0399	0,0653	0,2977	96,5	
11	0,0174	0,0839	0,3957	92,0	-0,0153	0,0707	0,3775	96,5	0,0398	0,0654	0,2976	95,5	
12	0,0167	0,0828	0,3957	93,5	-0,0147	0,0705	0,3778	95,5	0,0398	0,0653	0,2985	96,5	
13	0,0114	0,0796	0,4219	96,0	-0,0217	0,0772	0,3926	94,5	0,0409	0,0681	0,3154	96,5	
14	0,0123	0,0806	0,4212	96,5	-0,0207	0,0766	0,3914	95,0	0,0411	0,0679	0,3163	98,0	
15	0,0132	0,0798	0,4212	95,5	-0,0207	0,0775	0,3912	94,5	0,0405	0,0681	0,3150	97,5	
16	0,0096	0,0685	0,3177	92,5	0,0007	0,0557	0,3083	95,0	0,0115	0,0392	0,2052	97,0	
17	0,0098	0,0686	0,3180	93,0	0,0014	0,0562	0,3091	96,0	0,0115	0,0394	0,2054	96,0	
18	0,0095	0,0686	0,3182	92,5	0,0012	0,0555	0,3100	95,5	0,0115	0,0392	0,2057	96,0	
ϕ_4													
1	-0,1579	0,1586	0,5045	94,5	-0,1102	0,1219	0,5612	96,5	-	-	-	-	
2	-0,1575	0,1583	0,5054	94,0	-0,1109	0,1226	0,5622	97,5	-	-	-	-	
3	-0,1584	0,1592	0,5058	94,5	-0,1104	0,1219	0,5620	97,0	-	-	-	-	
4	-0,1006	0,1053	0,3908	95,5	-0,0659	0,0929	0,4586	94,5	-	-	-	-	
5	-0,1008	0,1053	0,3910	95,5	-0,0663	0,0929	0,4582	94,0	-	-	-	-	
6	-0,1001	0,1046	0,3904	97,0	-0,0670	0,0931	0,4602	95,5	-	-	-	-	
7	-0,1058	0,1093	0,4075	98,0	-0,0634	0,0863	0,4663	98,0	-	-	-	-	
8	-0,1063	0,1100	0,4095	97,5	-0,0632	0,0857	0,4682	97,0	-	-	-	-	
9	-0,1050	0,1089	0,4079	97,0	-0,0623	0,0845	0,4680	97,5	-	-	-	-	
10	-0,0614	0,0715	0,3107	97,5	-0,0238	0,0649	0,3783	99,5	-	-	-	-	
11	-0,0617	0,0719	0,3112	96,5	-0,0237	0,0640	0,3772	99,0	-	-	-	-	
12	-0,0614	0,0714	0,3113	98,0	-0,0236	0,0641	0,3779	99,0	-	-	-	-	
13	-0,0714	0,0790	0,3301	96,0	-0,0301	0,0693	0,3860	97,0	-	-	-	-	
14	-0,0716	0,0789	0,3301	96,5	-0,0291	0,0681	0,3852	97,5	-	-	-	-	
15	-0,0713	0,0785	0,3303	97,0	-0,0297	0,0684	0,3857	98,0	-	-	-	-	
16	-0,0380	0,0522	0,2508	97,5	-0,0007	0,0517	0,3106	97,5	-	-	-	-	
17	-0,0380	0,0521	0,2508	97,0	-0,0002	0,0525	0,3097	97,0	-	-	-	-	
18	-0,0384	0,0525	0,2509	97,0	-0,0007	0,0518	0,3100	97,5	-	-	-	-	

Neste estudo de simulação verificamos o comportamento das estimativas do número de animais vivos na t -ésima ocasião de amostragem N_t , $t \in \mathcal{T}$. Para tal, utilizamos os valores gerados desta quantidade durante as atualizações das variáveis latentes, com expressão definida na página 83. Observamos que o número esperado de animais vivos em cada ocasião é definido por

$$\mathbb{E}(N_1) = N \cdot \mathbb{E}(\beta_1)$$

e

$$\mathbb{E}(N_t) = \mathbb{E}(N_{t-1})\mathbb{E}(\phi_{t-1}) + N \cdot \mathbb{E}(\beta_t),$$

para $t \in \{2, \dots, T\}$. Por exemplo, no primeiro cenário, temos os tamanhos populacionais esperados nas seis ocasiões dados respectivamente por 120, 129, 126,75, 110,7, 129,63 e 118,704 enquanto que para o cenário 18 estes valores são 480, 516, 507, 442,8, 518,52 e 474,816.

Na Figura 19 apresentamos os vieses relativos e os erros absolutos médios dos estimadores pontuais para os tamanhos populacionais em cada ocasião de amostragem. No entanto, exibimos estes critérios apenas nos cenários onde $\sigma = 20$ uma vez que para os demais as estimativas mostraram-se análogas a este caso. Observamos que, apesar dos vieses relativos não diminuírem significativamente com o aumento da probabilidade de captura, houve uma diminuição dos erros médios. Além disto, nota-se que os vieses relativos não ultrapassaram, em valores absolutos, mais de 4% do verdadeiro valor do tamanho populacional, evidenciando um comportamento aceitável para estes estimadores.

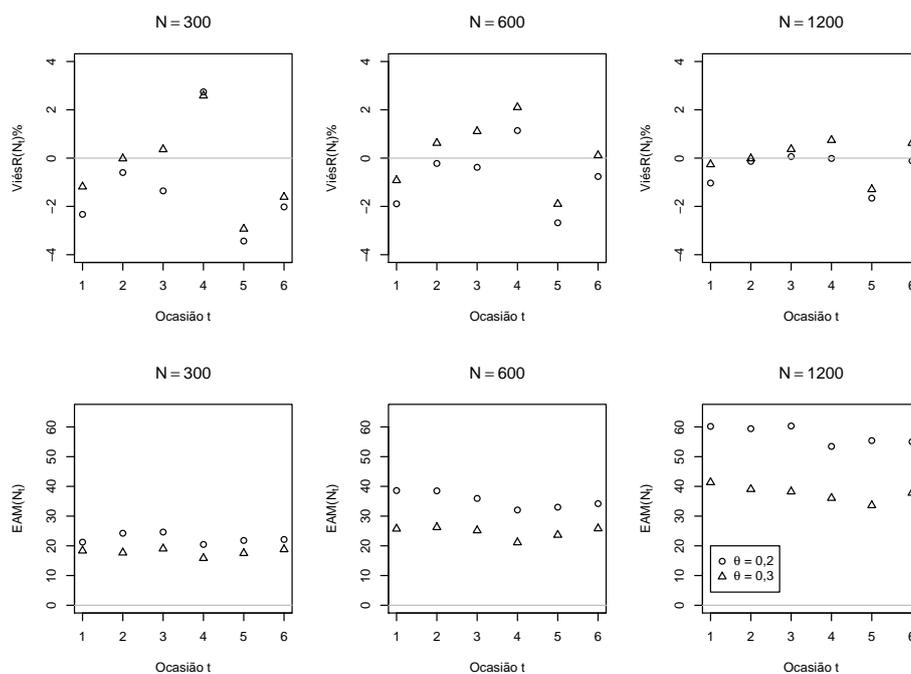


Figura 19 – Viés Relativo (%) e EAM do número estimado de animais vivos em cada ocasião de captura.

Na Tabela 27 são apresentadas as medidas de desempenho do estimador intervalar HPD para os tamanhos populacionais em cada ocasião. Em geral, notamos uma diminuição de pelo menos 20% na amplitude média do intervalo de credibilidade com o aumento da probabilidade de captura. Além disto, a probabilidade de cobertura está próxima do nível de credibilidade adotado para construção dos intervalos HPD. Analogamente ao que foi observado nos resultados para o tamanho da superpopulação, N , o valor verdadeiro de σ não apresentou influência significativa nas qualidades das estimativas produzidas. Em suma, os vieses para os parâmetros de recrutamento (β) e sobrevivência (ϕ) não influenciaram significativamente as estimativas dos tamanhos populacionais em cada ocasião.

Tabela 27 – Medidas de desempenho do estimador intervalar bayesiano para os parâmetros N_t .

Cenário	N_1		N_2		N_3		N_4		N_5		N_6	
	AM	PC										
1	130,5	97,0	132,4	95,5	131,0	96,0	123,3	97,5	127,3	96,0	128,5	95,0
2	130,3	96,0	131,8	95,0	131,0	96,0	122,8	97,5	126,9	96,0	128,2	94,5
3	130,4	97,0	132,2	93,5	130,4	96,0	123,0	97,5	127,1	96,0	128,1	94,0
4	96,5	95,5	94,3	97,0	94,2	94,5	86,6	98,5	88,1	95,5	92,8	94,5
5	96,7	97,0	94,5	96,5	94,0	94,0	86,6	98,0	88,1	96,0	92,6	94,0
6	96,7	95,5	94,4	96,0	94,3	96,0	86,7	98,5	88,0	96,0	92,6	94,5
7	194,6	94,5	195,9	96,5	195,0	96,0	178,0	95,0	186,7	97,0	191,3	96,0
8	195,2	94,5	196,4	96,5	195,1	97,0	178,9	95,5	186,7	96,5	192,1	96,0
9	195,5	94,5	196,4	96,5	195,7	96,5	178,7	95,0	187,1	98,0	192,6	96,0
10	143,4	95,5	139,7	97,5	139,3	97,5	125,9	97,5	128,5	96,5	138,8	96,5
11	143,3	96,0	139,3	97,5	139,0	98,0	126,0	97,0	128,4	97,0	138,1	97,5
12	143,3	96,5	139,3	97,0	139,2	97,5	126,1	97,5	128,5	97,5	138,5	96,5
13	293,1	95,5	294,6	93,0	290,7	95,0	261,9	96,5	277,6	94,5	287,8	96,0
14	293,7	94,5	294,9	92,5	291,5	94,5	262,9	96,5	277,6	93,0	287,6	96,0
15	293,5	95,5	294,4	93,0	291,5	93,5	262,5	96,5	277,9	92,5	287,7	96,0
16	213,8	95,0	206,6	96,0	204,3	95,5	184,1	95,5	190,3	94,5	207,6	97,0
17	213,6	96,5	206,9	95,0	204,6	95,5	183,8	97,0	190,9	95,0	207,9	96,5
18	213,9	95,0	206,8	95,0	204,8	95,5	184,5	95,5	190,9	95,0	207,7	96,5

Por fim, apresentamos os resultados numéricos dos parâmetros referentes ao comportamento territorial da população. Na Tabela 28 são exibidas as medidas descritivas para os estimadores pontual e intervalar do parâmetro σ . Observamos que o estimador pontual apresenta um padrão de sobreestimação para este parâmetro, apesar de considerarmos baixa a discrepância para o verdadeiro valor do parâmetro. Isto pode ser visto pelo viés relativo que não ultrapassa 5% em nenhum dos cenários considerados no estudo. De fato, para um mesmo valor de σ , o aumento da probabilidade de captura ou tamanho populacional refletiu em uma leve diminuição dos erros quadrático e absoluto médios para o estimador e acarretou uma diminuição média de aproximadamente 50% da amplitude do intervalo, mostrando uma efetiva contribuição destes fatores na precisão das estimativas. A probabilidade de cobertura esteve próxima ao nível considerado na construção do intervalo HPD.

Por outro lado, a fim de verificar o impacto do número de regiões – especificamente do tamanho de cada região – na estimativa de σ , consideramos um estudo computacional com dois casos de partição da região de amostragem: o primeiro, com 12 regiões igualmente espaçadas

Tabela 28 – Medidas de desempenho dos estimadores bayesianos do parâmetro σ .

Cenário	Viés	ViésR(%)	EAM	EQM	AM	PC
1	0,98	4,91	3,09	14,97	14,58	95,00
2	0,88	2,94	3,86	22,36	18,21	96,00
3	1,56	3,90	4,92	39,10	23,68	95,50
4	0,86	4,29	2,10	7,07	10,09	95,50
5	0,84	2,81	2,78	11,95	12,65	94,00
6	1,18	2,96	3,40	17,66	16,27	95,00
7	0,88	4,39	2,14	7,25	10,17	94,50
8	0,92	3,08	2,64	11,15	12,87	94,50
9	1,38	3,46	3,81	21,43	16,59	96,00
10	0,75	3,74	1,51	3,55	7,03	95,50
11	0,81	2,71	1,97	6,19	9,00	93,50
12	1,15	2,89	2,60	10,05	11,59	94,00
13	0,90	4,48	1,56	3,94	7,21	91,50
14	0,78	2,59	1,97	6,20	9,25	94,50
15	1,72	4,30	2,99	13,52	12,22	90,00
16	0,56	2,80	1,18	2,20	5,01	91,00
17	0,57	1,91	1,39	2,99	6,39	95,00
18	1,44	3,60	2,13	6,83	8,47	93,00

de amplitude 45 e, o segundo, com 36 regiões igualmente espaçadas de amplitude 15. Isto é, consideramos o aumento ou a diminuição de 50% no amplitude da região inicial. Para fins de comparação, utilizamos os cenários 7, 8 e 9. Os resultados das medidas de desempenho para σ são dados na Tabela 29. De fato, com o aumento do número de regiões de amostragem, de 18 para 36, houve um ganho significativo na precisão das estimativas pontuais e intervalares. Para o estimador pontual, observamos um viés quase nulo nas estimativas e os menores erros médios observados quando $R = 36$. Além disto, a probabilidade de cobertura permaneceu próxima ao verdadeiro valor nominal de 95% e com uma diminuição da amplitude média dos intervalos HPD produzidos exibindo assim, de forma geral, uma melhor precisão para ambos estimadores. Isto já era esperado uma vez que há mais precisão do local específico da captura do animal ao longo da região de estudo. Contudo, podemos observar que não houve uma diferença tão expressiva na comparação de $R = 12$ contra $R = 18$. Com efeito, o aumento da região de busca em 50% acarretou apenas em uma diminuição de 6 regiões de amostragem. Em resumo, concluímos que o modelo proposto apresenta estimativas satisfatórias para as partições da região de amostragem consideradas.

Na Figura 20 exibimos os vieses de π nos 18 cenários para cada região. Observamos que os vieses absolutos foram mais aparentes nas primeiros cenários considerados. Em suma, foram positivos para regiões de baixa densidade de centros de vida e negativos para regiões com maiores concentrações de animais. No entanto, os vieses se aproximam de zero nos últimos cenários em todas as regiões (baixas e altas densidades). Observamos que, quanto maior o valor

Tabela 29 – Comparação das medidas de desempenho dos estimadores bayesianos para o parâmetro σ em diferentes partições da região de amostragem.

R	Viés	ViésR(%)	EAM	EQM	AM	PC
12	0,93	4,65	2,13	6,76	9,10	94,50
12	1,37	4,56	2,34	8,91	10,83	92,50
12	2,45	6,13	3,56	19,88	13,46	87,00
18	0,88	4,39	2,14	7,25	10,17	94,50
18	0,92	3,08	2,64	11,15	12,87	94,50
18	1,38	3,46	3,81	21,43	16,59	96,00
36	0,27	1,36	1,16	2,02	5,52	94,50
36	0,23	0,78	1,63	4,21	7,76	93,00
36	0,37	0,92	2,13	6,99	10,34	96,00

de σ , maiores foram os vieses para π . De fato, isto é compreensível uma vez que, quando os animais se movimentam com maior distância entre as regiões há uma provável complicação em estimar a exata posição do seu centro de vida. O erro quadrático médio do estimador pontual de π , exibido na Figura 21, apresentou maiores valores para as regiões com maiores valores para o parâmetro, como esperado. Contudo, o aumento do número de animais da população ou probabilidade de captura acarretou uma diminuição, em média, de 0,01 nos valores dos EAM's sendo mais aparente nas regiões com baixa frequência de centros de vida, especificamente nas bordas da região de estudo.

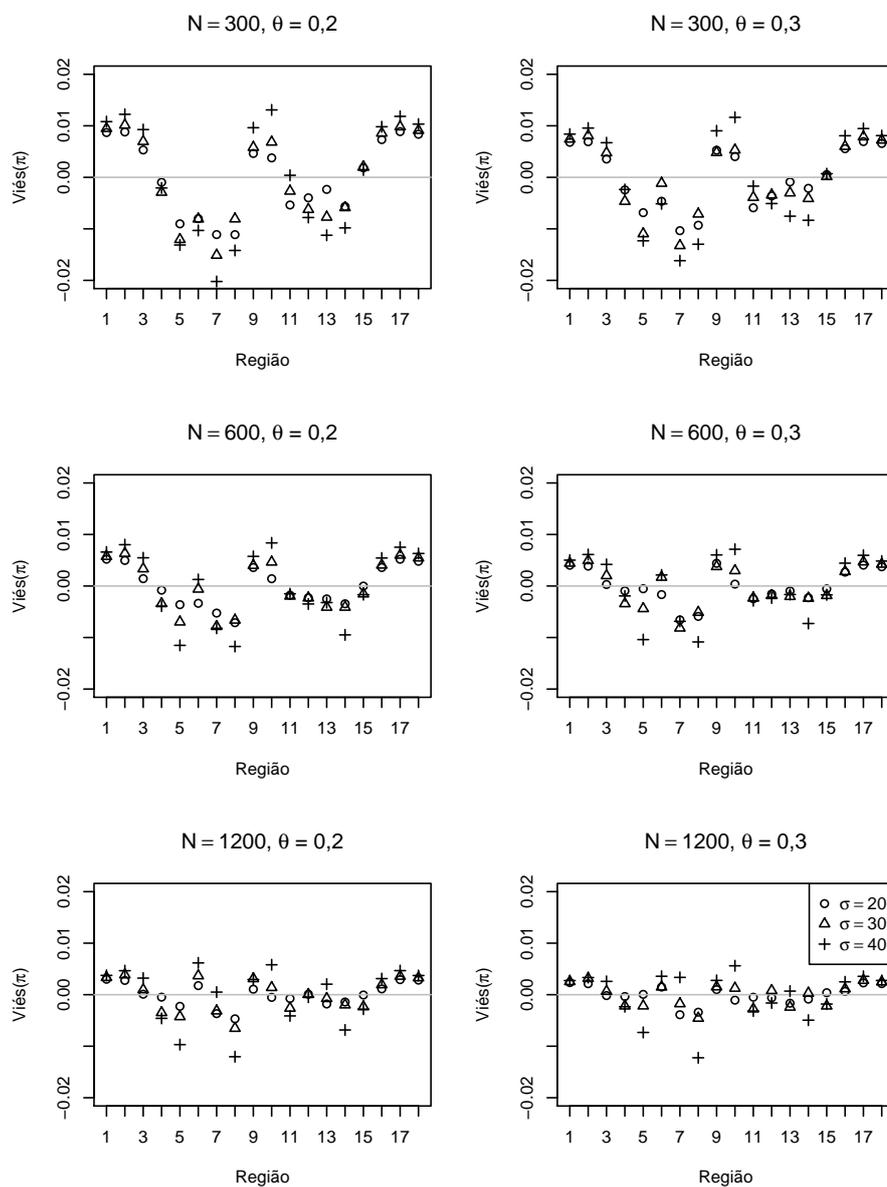


Figura 20 – Viés Relativo (%) dos estimadores bayesianos pontuais para o vetor de parâmetros $\boldsymbol{\pi}$.

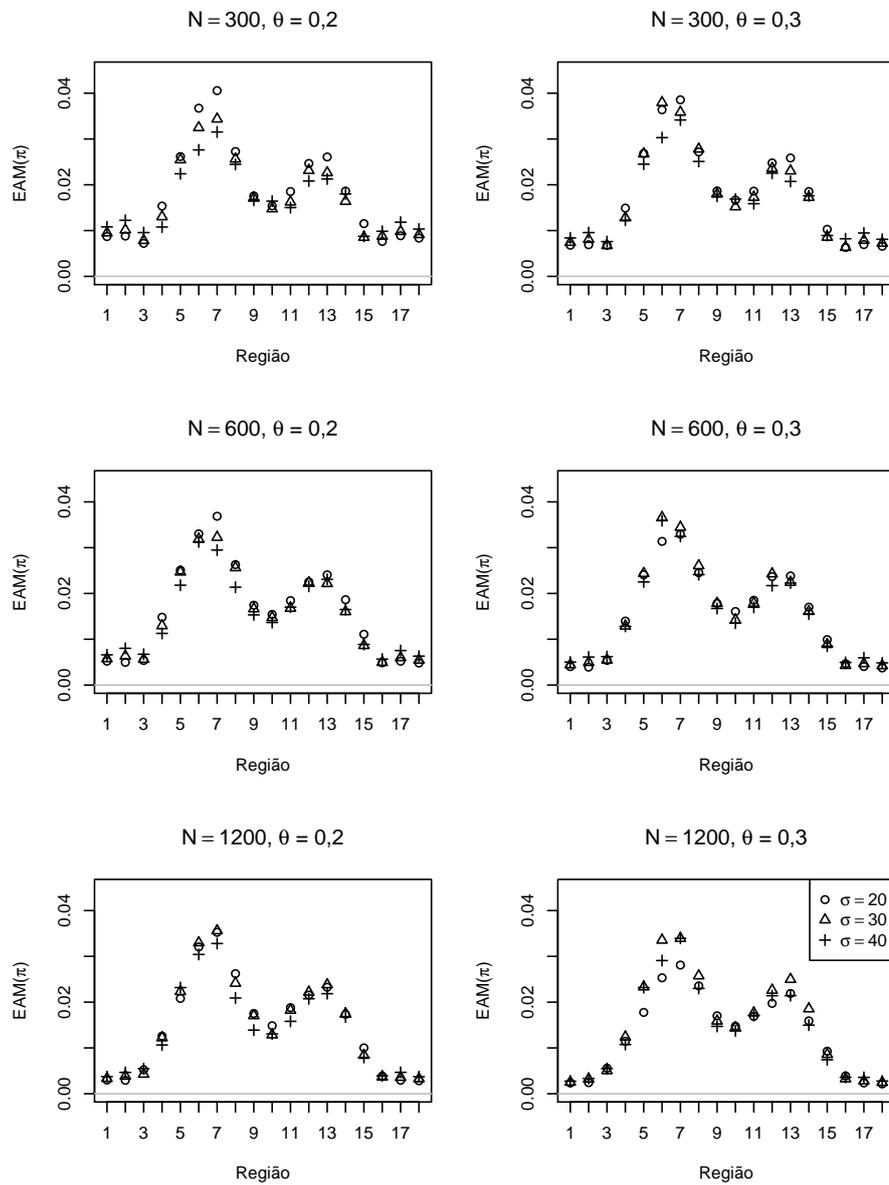


Figura 21 – EAM dos estimadores bayesianos pontuais para o vetor de parâmetros π .

APLICAÇÃO: ESPÉCIES *EUSARCUS CAVERNICOLA* E *EUSARCUS ADUNCUS*

Neste capítulo aplicaremos a metodologia proposta no capítulo anterior para estimação dos parâmetros populacionais utilizando os conjuntos de dados reais de captura das espécies *Eusarcus cavernicola* e *Eusarcus aduncus* apresentados por Monte (2017). Como comentado na Seção 1.1 do Capítulo 1, o processo de captura destas espécies foi realizado em seis ocasiões de amostragem, respectivamente, em Junho, Agosto e Outubro de 2014 e Junho, Setembro e Outubro de 2015, no interior da caverna Gruta Olhos d'Água, Minas Gerais. A caverna foi subdividida em 50 regiões de 30 metros cada em sentido horizontal a partir da sua entrada, resultando em uma região total de 1.500 metros para a realização da amostragem. A partir do processo de marcação adotado, foi possível registrar o histórico completo de captura dos animais, isto é, em quais ocasiões e respectivas regiões cada animal foi capturado. A seguir, apresentamos uma descrição desses dois conjuntos de dados .

6.1 Análise descritiva dos dados

Primeiramente, apresentamos na Tabela 30 um resumo do total de capturas e recapturas para as duas espécies em estudo, onde n_t representa o número total de animais capturados (marcados e não marcados) na t -ésima ocasião de amostragem, f_t o número de animais não marcados na t -ésima ocasião de amostragem e m_t' o número de animais marcados na t -ésima ocasião de amostragem recapturados na ocasião t' , $t < t'$. Na última linha de cada tabela, temos os totais (soma) de cada coluna.

Observamos que, no final do processo de CR, foram contabilizadas 144 e 160 capturas para as espécies *Eusarcus cavernicola* e *Eusarcus aduncus*, respectivamente. Destes montantes, 26 (18,06%) foram por recapturas da *Eusarcus cavernicola* e 32 (20%) foram por recapturas de

Tabela 30 – Frequências de capturas para as espécies *Eusarcus cavernicola* e *Eusarcus aduncus*.

Ocasião (t)	<i>Eusarcus cavernicola</i>							<i>Eusarcus aduncus</i>						
	n_t	f_t	$m_t^{(2)}$	$m_t^{(3)}$	$m_t^{(4)}$	$m_t^{(5)}$	$m_t^{(6)}$	n_t	f_t	$m_t^{(2)}$	$m_t^{(3)}$	$m_t^{(4)}$	$m_t^{(5)}$	$m_t^{(6)}$
1	21	21	0	1	4	0	0	10	10	0	0	1	1	0
2	11	11	-	2	0	0	0	24	24	-	5	2	2	1
3	29	26	-	-	1	0	0	33	28	-	-	1	0	0
4	28	23	-	-	-	8	4	31	27	-	-	-	10	3
5	27	19	-	-	-	-	6	39	26	-	-	-	-	6
6	28	18	-	-	-	-	-	23	13	-	-	-	-	-
Total	144	118	0	3	5	8	10	160	128	0	5	4	13	10

aracnídeos da outra espécie. Logo, de acordo com estes números, observou-se 118 aracnídeos distintos da espécie *Eusarcus cavernicola* e 128 animais distintos da espécie *Eusarcus aduncus* durante todo o estudo. Observamos também que o padrão das frequências capturas e recapturas foi semelhante para as duas espécies com exceção das duas primeiras ocasiões quando houve, de certa forma, uma alternância dos números totais de animais capturados. Os aracnídeos da espécie *Eusarcus aduncus*, marcados na primeira ou na segunda ocasião de amostragem, foram recapturados em períodos mais longos no tempo se comparados aos aracnídeos da outra espécie marcados nestas ocasiões, o que pode induzir taxas diferentes de sobrevivência entre as espécies.

A seguir, comentaremos sobre o comportamento territorial das espécies. Inicialmente, apresentamos na Tabela 31 os totais de recapturas de acordo com as distâncias entre as regiões e das ocasiões de amostragem. As linhas desta tabela estão associadas à distância entre as regiões onde ocorreu a recaptura, sendo que uma diferença igual a 0 (zero) representa recaptura na mesma região, igual a 1 (um) representa regiões vizinhas, igual a 2 representa regiões separadas por uma região entre elas, e assim por diante. A mesma interpretação vale para diferenças entre ocasiões de amostragem e são exibidas nas colunas.

Tabela 31 – Totais de recapturas entre diferentes regiões.

Diferença entre regiões	<i>E. cavernicola</i>					<i>E. aduncus</i>				
	entre ocasiões					entre ocasiões				
	1	2	3	4	5	1	2	3	4	5
0	10	4	0	0	0	15	2	2	1	0
1	7	1	2	0	0	6	1	0	0	0
2	2	0	2	0	0	1	1	0	1	0
3	0	0	0	0	0	1	1	1	0	0
4	0	0	0	0	0	0	0	0	0	0
5	1	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0

Por exemplo, 14 aracnídeos da espécie *Eusarcus cavernicola* foram recapturados na mesma região de uma captura prévia enquanto observou-se 20 para a outra espécie. Por outro lado, observamos 7 aracnídeos da espécie *Eusarcus cavernicola* recapturados em regiões vizinhas entre consecutivas ocasiões de amostragem, dois animais com diferença de duas regiões e um

animal com diferença de 5 regiões de amostragem. Em geral, notamos que os animais exibem comportamentos de baixa movimentação no sentido de serem capturados e recapturados em regiões próximas uma das outras, embora diferença de até 6 regiões foi observada para a espécie *Eusarcus aduncus*. Resumidamente, 15 e 14 animais foram recapturados em regiões distintas ao longo do tempo para as espécies *Eusarcus cavernicola* e *Eusarcus aduncus* representando 10,42% e 8,75% do total de capturas realizadas, respectivamente.

Ilustramos na Figura 22 os números totais de capturas (e recapturas) das duas espécies ao longo da área de estudo. Primeiramente, destacamos a similaridade na distribuição espacial dos aracnídeos para as duas populações como, por exemplo, a região 12 em que se concentra o maior número de aracnídeos para as ambas espécies e as três primeiras regiões sem aracnídeos capturados. Adicionalmente, nenhum aracnídeo destas espécies foi encontrado a partir da região 19. Especificamente, Monte (2017) destaca que há uma constrição de aproximadamente 0,4 m de largura e 0,9 m de altura no final da região 18 que pode ser um delimitante para a vida animal destas duas espécies até a esta região. Por esta razão, fixamos na aplicação um habitat nas primeiras 18 regiões, que corresponde aos 540 metros iniciais da caverna.

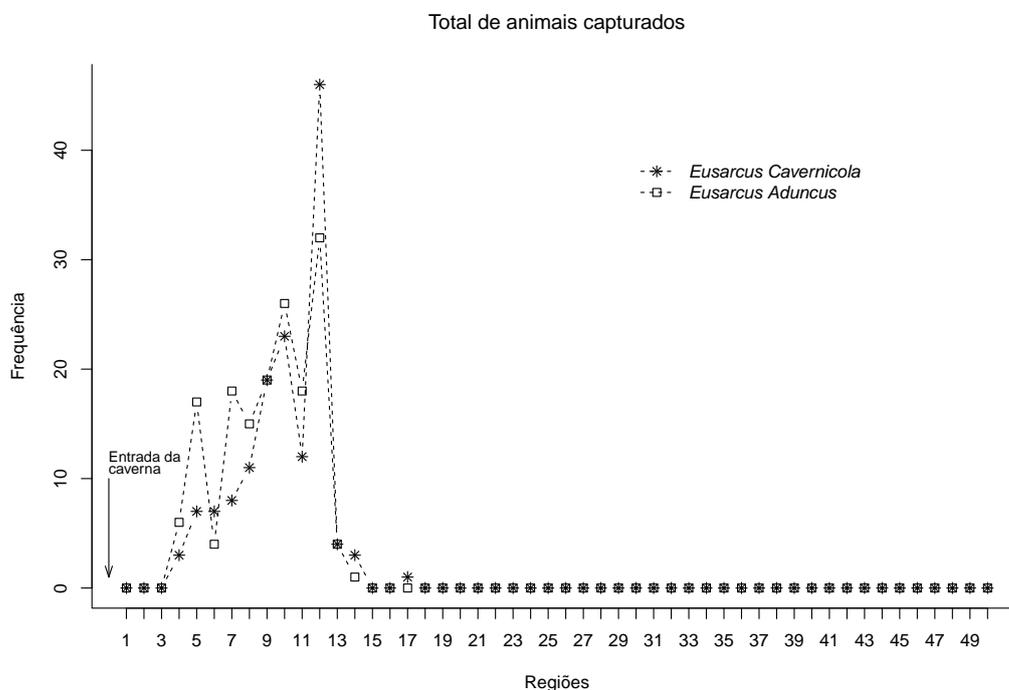


Figura 22 – Número de animais capturados em cada região de amostragem.

6.2 Análise Inferencial

Primeiramente, como foi observado certa similaridade entre os dois conjuntos de dados na análise descritiva, realizaremos juntamente com a estimação dos tamanhos populacionais um estudo comparativo de igualdade entre parâmetros das duas espécies. Particularmente, compararemos as probabilidades de recrutamento (β), morte (ϕ), captura (θ), movimentação (σ) e distribuição espacial (π) entre as duas populações. Para isto, consideramos novamente o critério de comparação *Deviance Information Criterion* (DIC) (SPIEGLHALTER *et al.*, 2002). Considerando $p = 1$ o índice relacionado à espécie *Eusarcus cavernicola* e $p = 2$ para a espécie *Eusarcus aduncus*, sejam $\mathbf{n}^{(1)}$ e $\mathbf{n}^{(2)}$ os respectivos conjuntos de dados de captura-recaptura destas duas espécies. Supomos que as amostras observadas são realizações independentes de captura no sentido que podemos associar uma função de verossimilhança conjunta por

$$\begin{aligned} \mathcal{L}_c(N^{(1)}, \beta^{(1)}, \phi^{(1)}, \theta^{(1)}, \sigma^{(1)}, \pi^{(1)}, N^{(2)}, \beta^{(2)}, \phi^{(2)}, \theta^{(2)}, \sigma^{(2)}, \pi^{(2)}; \mathbf{n}^{(1)}, \mathbf{n}^{(2)}) &= \\ &= \mathcal{L}(N^{(1)}, \beta^{(1)}, \phi^{(1)}, \theta^{(1)}, \sigma^{(1)}, \pi^{(1)}; \mathbf{n}^{(1)}) \mathcal{L}(N^{(2)}, \beta^{(2)}, \phi^{(2)}, \theta^{(2)}, \sigma^{(2)}, \pi^{(2)}; \mathbf{n}^{(2)}) \end{aligned}$$

onde \mathcal{L} é a função de verossimilhança definida em (5.6).

Adotamos para a abordagem bayesiana as seguintes distribuições a priori:

$$\begin{aligned} \phi_t^{(p)} &\stackrel{\text{ind.}}{\sim} \text{Beta}(1, 1), \quad t \in \{1, 2, \dots, 5\}, \\ \theta^{(p)} &\sim \text{Beta}(1, 1), \\ \beta^{(p)} &\sim \text{Dirichlet}(1, 1, 1, 1, 1), \\ \pi^{(p)} &\sim \text{Dirichlet}(1, 1, \dots, 1, 1), \\ \sigma^{(p)} &\sim \text{Gamma}(0, 01; 0, 001), \\ N^{(p)} &\sim \text{BN}(1, 1000000^{-1}), \end{aligned}$$

para $p = 1, 2$.

Os valores dos hiperparâmetros foram escolhidos para que as distribuições a priori sejam pouco informativas. Utilizamos o algoritmo MCMC para obter uma amostra de 150.000 valores gerados para cada parâmetro. Para a realização dos resumos a posteriori das quantidades de interesse, descartamos as 50.000 primeiras amostras geradas e, do restante, tomamos valores a cada 10 para minimizar o efeito de autocorrelação da cadeia, resultando em uma amostra final de 10.000 valores. Analisamos a convergência das cadeias através do critério de Gelman e Rubin (1992) (GR) utilizando o pacote CODA (BEST; COWLES; VINES, 1995) do software R (R Development Core Team, 2014). O critério CR e o monitoramento gráfico das cadeias MCMC pelas figuras anexadas no Anexo A apontam indicativos de convergência do algoritmo.

Iniciamos a etapa inferencial apresentando na Tabela 32 os valores do critério DIC para a comparação dos 32 diferentes modelos determinados de acordo com a combinação de igualdade (ou não) entre os processos. De acordo com o critério, o modelo que melhor se ajusta

aos dados é aquele cuja diferença reside apenas nas probabilidades de recrutamento entre as populações. Isto já era esperado de acordo com a diferença apresentada no padrão do total de animais não marcados capturados em cada ocasião (veja Tabela 30), essencialmente nas duas primeiras, resultando maior diferença entre as estimativas nestes períodos (veja Tabela 35). Observamos também que o modelo que leva em conta as diferenças entre o parâmetro de movimento foi o segundo com melhor ajuste. Em particular, as estimativas pontuais para o parâmetro σ considerando este modelo alternativo foram 23,18 e 26,43 com intervalos de credibilidades (17,09; 30,30) e (19,95; 33,26) para as espécies *Eusarcus cavernicola* e *Eusarcus aduncus*, respectivamente. De fato, esta inferência é coerente com os dados observados pois os animais da espécie *Eusarcus aduncus* foram recapturados em maiores distâncias entre duas regiões quando comparados com os da outra espécie, como exibido na Tabela 31. Ressaltamos que os modelos com melhores ajustes não diferenciam as probabilidades de sobrevivência, captura e distribuição espacial entre as populações. No tocante à escolha do modelo, adotamos aquele com diferença apenas nas probabilidades de recrutamento, como indicado pelo critério DIC.

Na Tabela 33 a seguir são apresentadas as estimativas pontuais e intervalares das abundâncias para as duas espécies, bem como o desvio-padrão (d.p.) dos estimadores pontuais. Para a espécie *Eusarcus cavernicola*, obtivemos uma estimativa média da superpopulação de 298,91 aracnídeos, pouco menos do que para espécie *Eusarcus aduncus*, com estimativa média de 302,54. As estimativas intervalares para estas quantidades informam que, com 95% de probabilidade, o verdadeiro número de aracnídeos da população durante o período de estudo esta aproximadamente entre o dobro e o triplo do número de animais observados na amostra (118 animais para a espécie *Eusarcus cavernicola* e 128 para *Eusarcus aduncus*). Adicionalmente, utilizando uma largura média da caverna de 2,5 m totalizando aproximadamente 1350 m² de área pesquisada (até a décima oitava região), temos densidades médias estimadas de 0,21 e 0,23 animais por metro quadrado para as espécies *Eusarcus cavernicola* e *Eusarcus aduncus*, respectivamente.

A seguir, realizamos a inferência para o número total de animais vivos em cada ocasião de amostragem. A Tabela 34 exhibe os números estimados para estas quantidades que também são ilustrados graficamente na Figura 23. Os resultados mostram uma maior variação entre os tamanhos populacionais nas primeiras ocasiões de amostragem, com uma situação de crescimento da abundância populacional neste período e leve decaimento nas ocasiões seguintes. Ao compararmos o número total de animais no final de cada ano, isto é, ocasiões 3 e 6 de amostragem, temos uma diminuição de aproximadamente 13% e 23% no número total de animais vivos das espécies *Eusarcus cavernicola* e *Eusarcus aduncus*, respectivamente.

Na Tabela 35 são apresentadas as estimativas para os parâmetros de recrutamento, sobrevivência, captura e movimentação. Exceto os parâmetros de recrutamento, os demais são compartilhados entre as duas espécies de acordo com o modelo escolhido pelo critério DIC.

Tabela 32 – Valores do Critério DIC para comparação dos modelos.

Comparação entre os parâmetros de					
nascimento	sobrevivência	captura	dist. espacial	movimentação	$\Delta_i(\text{DIC})$
diferentes	iguais	iguais	iguais	iguais	0
diferentes	iguais	iguais	diferentes	iguais	1,64
iguais	iguais	iguais	iguais	iguais	2,84
iguais	iguais	iguais	diferentes	iguais	4,48
diferentes	diferentes	iguais	iguais	iguais	5,15
diferentes	diferentes	iguais	diferentes	iguais	6,64
iguais	diferentes	iguais	iguais	iguais	7,66
iguais	diferentes	iguais	diferentes	iguais	9,73
diferentes	iguais	diferentes	iguais	iguais	13,13
diferentes	iguais	diferentes	diferentes	iguais	13,66
iguais	iguais	diferentes	iguais	iguais	14,86
iguais	iguais	diferentes	diferentes	iguais	16,32
diferentes	iguais	iguais	iguais	diferentes	16,36
diferentes	iguais	iguais	diferentes	diferentes	17,57
diferentes	diferentes	diferentes	iguais	iguais	18,49
diferentes	diferentes	diferentes	diferentes	iguais	19,78
iguais	iguais	iguais	iguais	diferentes	19,83
iguais	iguais	iguais	diferentes	diferentes	20,06
iguais	diferentes	diferentes	iguais	iguais	20,74
diferentes	diferentes	iguais	diferentes	diferentes	21,73
iguais	diferentes	diferentes	diferentes	iguais	22,03
diferentes	diferentes	iguais	iguais	diferentes	22,41
iguais	diferentes	iguais	iguais	diferentes	24,34
iguais	diferentes	iguais	diferentes	diferentes	24,71
diferentes	iguais	diferentes	iguais	diferentes	29,08
diferentes	iguais	diferentes	diferentes	diferentes	29,96
iguais	iguais	diferentes	iguais	diferentes	31,7
iguais	iguais	diferentes	diferentes	diferentes	32,15
diferentes	diferentes	diferentes	iguais	diferentes	35,28
diferentes	diferentes	diferentes	diferentes	diferentes	35,65
iguais	diferentes	diferentes	iguais	diferentes	37,2
iguais	diferentes	diferentes	diferentes	diferentes	37,76

Tabela 33 – Resumos a posteriori para as abundâncias populacionais.

Espécie	Quantidade	Estimativas à posteriori			$IC_b(95\%)$
		Média	Mediana	d.p.	
<i>Eusarcus cavernicola</i>	Abundância	298,90	295	40,8378	(226, 382)
	Densidade (m ²)	0,2214	0,2185	0,0303	(0,1681, 0,2837)
<i>Eusarcus aduncus</i>	Abundância	302,54	299	39,5929	(230, 381)
	Densidade (m ²)	0,2241	0,2215	0,0293	(0,1704, 0,2822)

Primeiramente, observamos que as estimativas para o vetor β foram bem distintas entre as duas populações. Contudo, para cada uma das espécies, aproximadamente 58% do total de animais

Tabela 34 – Resumos a posteriori para as abundâncias populacionais em cada ocasião de captura.

	<i>Eusarcus cavernicola</i>				<i>Eusarcus aduncus</i>			
	Média	Mediana	d.p.	IC _b (95%)	Média	Mediana	d.p.	IC _b (95%)
N_1	84,17	82	19,88	(48, 122)	43,72	42	14,23	(20, 73)
N_2	63,25	61	17,53	(34, 98)	106,66	104	24,44	(61, 154)
N_3	127,05	123	27,36	(77, 181)	142,33	139	28,41	(91, 198)
N_4	106,15	104	22,27	(64, 149)	127,89	126	25,60	(79, 17)7
N_5	111,58	109	22,83	(69, 157)	136,40	134	24,00	(90, 182)
N_6	110,82	108	24,10	(66, 157)	108,55	105	25,46	(61, 157)

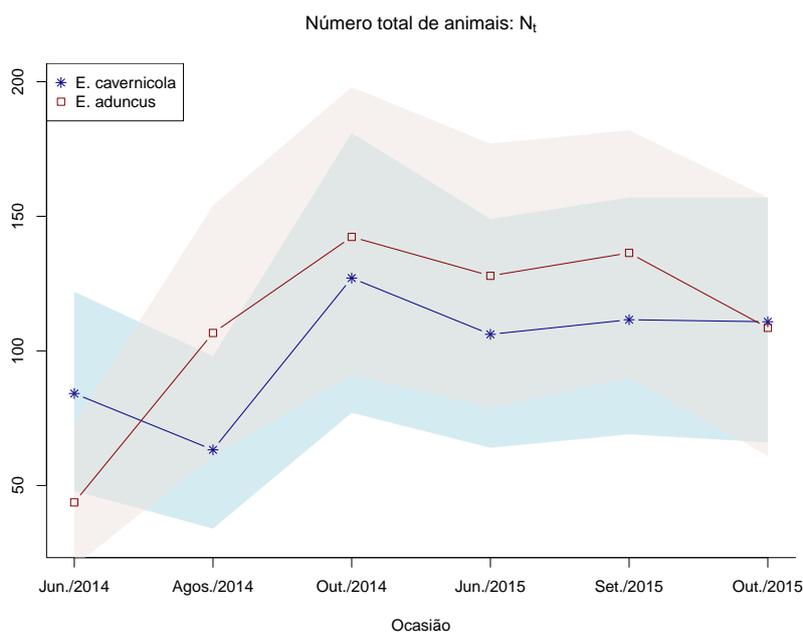


Figura 23 – Abundâncias populacionais estimadas por ocasião.

presentes na população durante o estudo foram recrutados ainda no primeiro ano de pesquisa, isto é, até a terceira ocasião de amostragem. Para as probabilidades de sobrevivência, observamos os maiores valores para ϕ_2 , ϕ_4 e ϕ_5 e menor valor para ϕ_3 , este último correspondente ao período de transição entre os dois anos. Assim, o período de tempo entre as ocasiões de amostragem teve, aparentemente, efeito nas probabilidades de sobrevivência. Observamos também que a estimativa de ϕ_1 esteve próximo à 50%, e acreditamos que devido ao número baixo de capturas na primeira ocasião de amostragem – e conseqüentemente a presença de poucos dados amostrais para estimar a sobrevivência da espécie nesta ocasião – levou a inferências a posteriori proporcionais a informação contida na distribuição a priori. A estimativa da probabilidade de captura, θ , induz que aproximadamente 25% dos animais vivos em uma determinada ocasião de amostragem são capturados pelos pesquisadores.

Em seguida, temos as inferências a posteriori para o parâmetro σ , com estimativa média 24,85 e intervalo de credibilidade (19,93, 30,06). Em razão da aproximação normal para modelagem da movimentação dos animais, temos que cada aracnídeo concentra, com 95% de

Tabela 35 – Resumos a posteriori para os parâmetros de nascimento, sobrevivência, captura e movimentação

Parâm.	Média	Mediana	d.p.	IC _b (95%)	Média	Mediana	d.p.	IC _b (95%)
β_1	0,2786	0,2755	0,0541	(0,1794, 0,3886)	0,1443	0,1397	0,0434	(0,0665, 0,2310)
β_2	0,0704	0,0638	0,0466	(0,0001, 0,1570)	0,2760	0,2738	0,0688	(0,1441, 0,4118)
β_3	0,2367	0,2361	0,0697	(0,1008, 0,3755)	0,1588	0,1568	0,0754	(0,0068, 0,2929)
β_4	0,2223	0,2208	0,0654	(0,1015, 0,3577)	0,2701	0,2695	0,0686	(0,1417, 0,4067)
β_5	0,0876	0,0803	0,0560	(0,0001, 0,1895)	0,1058	0,1019	0,0619	(0,0001, 0,2152)
β_6	0,1045	0,1004	0,0567	(0,0004, 0,2040)	0,0451	0,0370	0,0356	(0,0000, 0,1144)
ϕ_1	0,5080	0,4969	0,1511	(0,2230, 0,8095)	-	-	-	-
ϕ_2	0,8741	0,8960	0,0985	(0,6804, 1,0000)	-	-	-	-
ϕ_3	0,3161	0,3077	0,0887	(0,1548, 0,4942)	-	-	-	-
ϕ_4	0,8105	0,8220	0,1144	(0,6061, 1,0000)	-	-	-	-
ϕ_5	0,7028	0,7003	0,1327	(0,4658, 0,9728)	-	-	-	-
θ	0,2464	0,2435	0,0411	(0,1694, 0,3268)	-	-	-	-
σ	24,85	24,62	2,63	(19,93, 30,06)	-	-	-	-

probabilidade, seus movimentos em torno do seu centro de vida em uma área de amplitude aproximadamente $4\hat{\sigma} \approx 100$ m. Para ilustrar a obtenção desta inferência, suponhamos um animal com centro de vida localizado em $S^* = 250$ m. Realizamos um procedimento computacional para obter uma amostra gerada da distribuição preditiva da localização instantânea deste animal, W^* . Uma amostra w_1, \dots, w_B de tamanho B da distribuição preditiva de W^* , dado $S^* = s$, pode ser obtida por meio de simulações MCMC considerando a seguinte distribuição:

$$(W^* | S^* = 250) \sim \text{TNormal}(250, \sigma_b^2, (0, 540]) \quad (6.1)$$

onde $\text{TNormal}(250, \sigma_b^2, (0, 540])$ denota uma distribuição normal com média 250 e variância σ_b^2 truncada no intervalo $(0, 540]$ com σ_b^2 sendo o b -ésimo valor gerado da distribuição a posteriori do parâmetro σ . Utilizando o algoritmo MCMC definido no capítulo anterior para ter uma amostra da distribuição a posteriori de σ , a geração em (6.1) torna-se de fácil implementação. Apresentamos na Figura 24 estes valores gerados desta predição. Observamos que a forma da distribuição é parecida como de uma distribuição normal com média 250 e variância $\hat{\sigma}^2 = 24,85^2$, com a função densidade desta distribuição apresentada em linha azul. Diante disto, é possível concluir que este animal terá seus movimentos dentro de uma região de 198,98 m e 297,15 m com probabilidade de pelo menos 95%, ou numa região entre 154,1 m e 356,1 m com probabilidade de pelo menos 99%. Um comportamento similar desta distribuição preditiva foi observado para quaisquer centros de vida localizados entre a quinta e décima segunda regiões, as quais concentram maior parte dos centros destas espécies, como discutido a seguir.

Apresentamos na Figura 25 as estimativas pontuais e a região de credibilidade 95% para os componentes do vetor $\boldsymbol{\pi}$. Como esperado, a região 12 contém a maior probabilidade de vida animal, com estimativa média a posteriori para π_{12} igual à 0,2452, que corresponde a concentração de quase um quarto de todos os centros de vida dos animais. Observamos também que as quatro primeiras e as seis últimas regiões têm baixa probabilidade de vida animal, com aproximadamente 6,6% de probabilidade nestas regiões. Em geral, os animais concentram seus

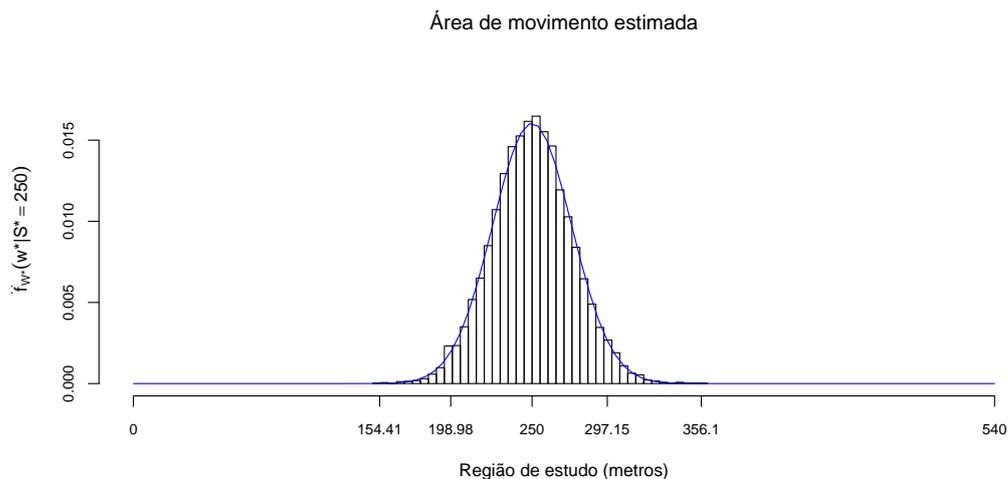


Figura 24 – Região de movimentos estimada para um animal com centro de vida localizado em 250 m.

movimentos entre (inclusive) as regiões 5 e 12 com probabilidade 93,4%, correspondente a uma área de aproximadamente 240 m no sentido horizontal.

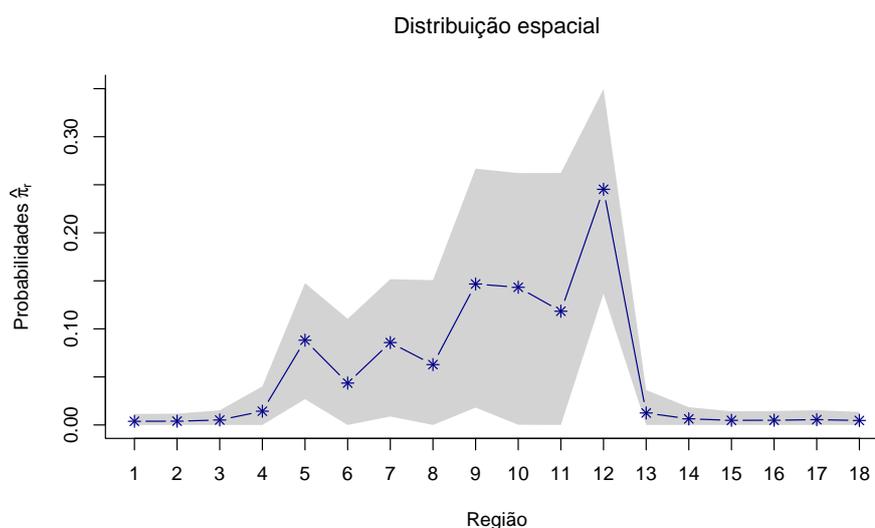


Figura 25 – Distribuição espacial estimada dos centros de vida.

Por fim, estimamos o número de animais em cada ocasião e região de amostragem, $N_{t,r}$. Para isto, utilizaremos a etapa de geração destas quantidades como definido na página 90. Apresentamos na Figura 26 os valores médios estimados destas quantidades. Observamos pela imagem à esquerda da figura, relacionado à espécie *Eusarcus cavernicola*, que a terceira ocasião teve maior dispersão dos animais (da região 5 até à 12 com razoável abundância). A região 12, em geral, apresentou alto valor para a abundância das espécies, embora em algumas ocasiões suas regiões vizinhas apresentassem maior concentração de animais. Por exemplo, para a espécie *Eusarcus cavernicola*, na ocasião 4, a região 11 foi aquela com maior abundância populacional,

enquanto que isto ocorreu para a espécie *Eusarcus aduncus* na ocasião 3. Aparentemente, os aracnídeos destas duas espécies não ocupam as mesmas regiões em uma mesma ocasião. Por exemplo, na terceira ocasião a região 12 apresentou a maior abundância para a espécie *Eusarcus cavernicola* enquanto que para *Eusarcus aduncus* foi na região 11, fato que se alternou na ocasião seguinte. Novamente, na sexta ocasião a maioria dos animais da primeira espécie se concentrou na região 12 enquanto os da outra espécie estavam mais concentrados na décima região. De forma menos intensa, o fato também ocorreu na segunda ocasião com relação à região 10. Estes fatos podem nos levar a pensar em uma espécie de disputa de território entre as duas populações.

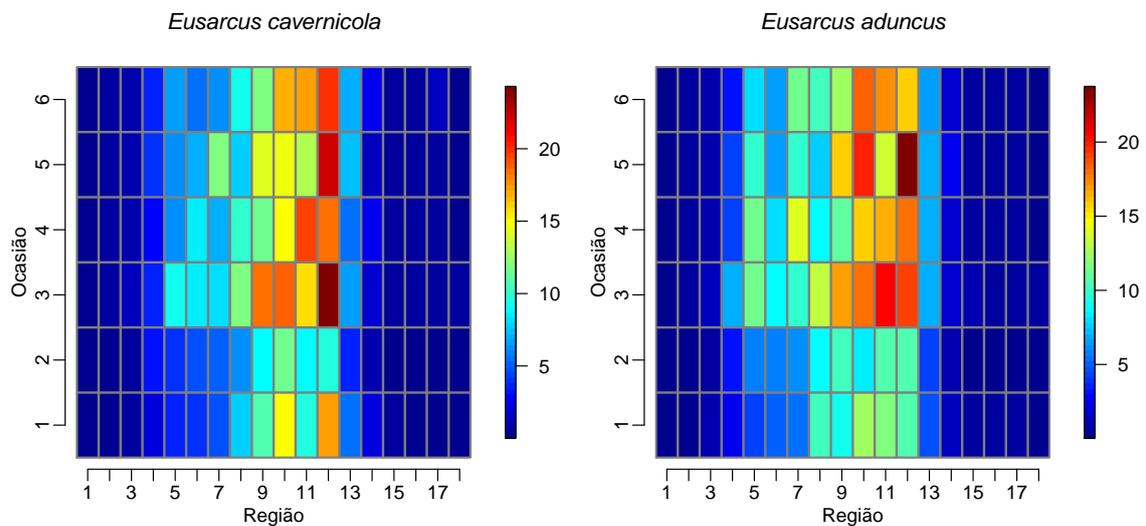


Figura 26 – Distribuição espaço-temporal estimada dos animais na caverna ao longo do tempo.

CONCLUSÕES

O presente trabalho ocupou-se de modelos espaciais de captura-recaptura para populações abertas. Nossos modelos estatísticos permitiram estimar as abundâncias populacionais em cada ocasião de captura, a partir de taxas de nascimento e sobrevivência das espécies, e em cada região de amostragem, a partir da incorporação dos registros espaciais da etapa amostral.

No primeiro modelo, conjuntamente com as estimativas dos parâmetros populacionais, a estrutura hierárquica proposta possibilitou realizar processos preditivos de abundâncias populacionais em locais não visitados durante a etapa amostral. Nesta modelagem, uma abordagem a partir de dados inflacionados de zeros evidenciou melhores resultados preditivos quando comparado com a abordagem usual nos casos que a amostragem é realizada em diversas regiões sem a presença da espécie. No segundo modelo, incorporamos os possíveis deslocamentos geográficos dos animais entre diferentes regiões de amostragem. Neste cenário, supomos que os animais concentram seus movimentos em torno de um determinado centro de vida e, neste contexto, o modelo proposto permitiu estimar a distribuição espacial dos animais sobre a região de amostragem e a área de movimento em torno dos seus centros de vida. Em ambos os modelos, procedimentos bayesianos foram adotados para o processo inferencial e o uso de técnicas de dados aumentados proporcionou algoritmos MCMC de simples implementação computacional. Estudos simulados constataram que os estimadores bayesianos dos parâmetros populacionais apresentam, em geral, boas propriedades frequentistas. Destacamos, ainda, que as propostas de modelagem apresentadas nesta tese contribuíram, via análise estatística de dados, para uma descrição mais detalhada de três espécies de aracnídeos apresentadas por [Monte \(2017\)](#).

Por fim, destacamos alguns tópicos abertos para serem abordados em pesquisas futuras. No contexto espacial, outras estruturas hierárquicas a priori para os tamanhos populacionais locais podem ser adotadas para o modelo apresentado no Capítulo 3 como, por exemplo, levando em consideração o grau de dependência entre as regiões a partir das distâncias geográficas entre elas ou, até mesmo, uma abordagem bayesiana não paramétrica ([DORAZIO et al., 2008](#)), como

o processo Dirichlet (FERGUSON, 1973; FERGUSON, 1974) que permite agrupar regiões que possuem o mesmo tamanho populacional. No modelo proposto no Capítulo 5, outras formas podem ser utilizadas para a função densidade para o centro de vida animal. Em particular, este modelo pode ser estendido para considerar a região de amostragem como um subconjunto do plano cartesiano (EFFORD, 2004; GLENNIE *et al.*, 2017). Ao considerarmos pesquisas futuras no contexto de população aberta, os dois modelos podem ser estendidos para considerar diferentes probabilidades de recrutamento, morte e captura entre as regiões de amostragem e a inclusão de covariáveis na modelagem destas quantidades (SCHOFIELD; BARKER, 2011; WU; HOLAN, 2017). Destacamos, por fim, que estudos a partir da teoria de estimação por máxima verossimilhança podem ser realizados futuramente. Como motivação, as técnicas de aumento de dados apresentadas nesta tese podem ser adaptadas para o contexto clássico utilizando o algoritmo EM (DEMPSTER; LAIRD; RUBIN, 1977).

REFERÊNCIAS

AMSTRUP, S. C.; MCDONALD, T. L.; MANLY, B. F. **Handbook of capture-recapture analysis**. [S.l.]: Princeton University Press, 2010. Citado na página 18.

ARNASON, A.; MILLS, K. Bias and loss of precision due to tag loss in jolly–seber estimates for mark–recapture experiments. **Canadian Journal of Fisheries and Aquatic Sciences**, NRC Research Press, v. 38, n. 9, p. 1077–1095, 1981. Citado na página 19.

ARNASON, A. N. The estimation of population size, migration rates and survival in a stratified population. **Researches on population ecology**, Springer, v. 15, n. 2, p. 1–8, 1973. Citado na página 74.

AZIZ, M.; TOLLINGTON, S.; BARLOW, A.; GREENWOOD, C.; GOODRICH, J.; SHAM-SUDDOHA, M. Estimating density and population size of bengal tigers in the bangladesh sundarbans using non-invasively collected genetic data. **Population status, threats, and evolutionary conservation genetics of Bengal tigers in the Sundarbans of Bangladesh**, p. 91, 2017. Citado nas páginas 19 e 75.

BERNARDINI, C.; RUSSO, C. D.; RAMPINI, M.; CERSARONI, D.; SBORDONI, V. A recent colonization of dolichopoda cave crickets in the poscola cave (orthoptera, raphidophoridae). **International Journal of Speleology**, v. 25, n. 1, p. 2, 1996. Citado na página 19.

BERTULLI, C. G.; GUÉRY, L.; MCGINTY, N.; SUZUKI, A.; BRANNAN, N.; MARQUES, T.; RASMUSSEN, M. H.; GIMENEZ, O. Capture-recapture abundance and survival estimates of three cetacean species in icelandic coastal waters using trained scientist-volunteers. **Journal of Sea Research**, Elsevier, v. 131, p. 22–31, 2018. Citado na página 19.

BEST, N.; COWLES, M. K.; VINES, K. Coda* convergence diagnosis and output analysis software for gibbs sampling output version 0.30. **MRC Biostatistics Unit, Cambridge**, v. 52, 1995. Citado nas páginas 66 e 106.

BORCHERS, D.; DISTILLER, G.; FOSTER, R.; HARMSSEN, B.; MILAZZO, L. Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap survey. **Methods in Ecology and Evolution**, Wiley Online Library, v. 5, n. 7, p. 656–665, 2014. Citado na página 75.

BORCHERS, D.; FEWSTER, R. *et al.* Spatial capture–recapture models. **Statistical Science**, Institute of Mathematical Statistics, v. 31, n. 2, p. 219–232, 2016. Citado nas páginas 76 e 78.

BORCHERS, D. L.; EFFORD, M. Spatially explicit maximum likelihood methods for capture–recapture studies. **Biometrics**, Wiley Online Library, v. 64, n. 2, p. 377–385, 2008. Citado nas páginas 19 e 75.

BROWNIE, C.; HINES, J.; NICHOLS, J.; POLLOCK, K.; HESTBECK, J. Capture-recapture studies for multiple strata including non-markovian transitions. **Biometrics**, JSTOR, p. 1173–1187, 1993. Citado na página 74.

- BUCKLAND, S. T. A modified analysis of the jolly-seber capture-recapture model. **Biometrics**, JSTOR, p. 419–435, 1980. Citado nas páginas 18 e 19.
- BUCKLAND, S. T.; ANDERSON, D. R.; BURNHAM, K. P.; LAAKE, J. L.; BORCHERS, D. L.; THOMAS, L. Introduction to distance sampling estimating abundance of biological populations. Oxford University Press., 2001. Citado na página 76.
- BULMER, M. On fitting the poisson lognormal distribution to species-abundance data. **Biometrics**, JSTOR, p. 101–110, 1974. Citado na página 36.
- BURNHAM, K. On a unified theory for release-resampling of animal populations. In: **Proceedings of 1990 Taipei Symposium in Statistics. Taipei, Taiwan: Institute of Statistical Science, Academia Sinica.** [S.l.: s.n.], 1991. p. 11–36. Citado nas páginas 18 e 19.
- BURNHAM, K. P. A unified theory for release-resampling studies of animal populations. 1987. Citado nas páginas 18 e 19.
- CARCHINI, G.; RAMPINI, M.; SBORDONI, V. Life cycle and population ecology of the cave cricket *dolichopoda geniculata* (costa) from valmarino cave (central italy). **International Journal of Speleology**, v. 23, n. 3, p. 6, 1994. Citado na página 19.
- CASELLA, G.; GEORGE, E. I. Explaining the gibbs sampler. **The American Statistician**, Taylor & Francis, v. 46, n. 3, p. 167–174, 1992. Citado na página 41.
- CASTLEDINE, B. A bayesian analysis of multiple-recapture sampling for a closed population. **Biometrika**, Biometrika Trust, v. 68, p. 197–210, 1981. Citado na página 35.
- CHANDLER, R. B.; ROYLE, J. A. *et al.* Spatially explicit models for inference about density in unmarked or partially marked populations. **The Annals of Applied Statistics**, Institute of Mathematical Statistics, v. 7, n. 2, p. 936–954, 2013. Citado na página 75.
- CHAO, A. An overview of closed capture-recapture models. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer, v. 6, n. 2, p. 158–175, 2001. Citado na página 18.
- CHEN, M.-H.; SHAO, Q.-M. Monte carlo estimation of bayesian credible and hpd intervals. **Journal of Computational and Graphical Statistics**, Taylor & Francis Group, v. 8, n. 1, p. 69–92, 1999. Citado na página 45.
- COLVIN, K. F. Kruschke, jk (2011). doing bayesian data analysis: A tutorial with r and bugs. burlington, ma: Academic press. **Journal of Educational Measurement**, Wiley Online Library, v. 50, n. 4, p. 469–471, 2013. Citado na página 45.
- COOPER, J. E.; COOPER, M. R. Long-term mark-recapture studies of population sizes in the stygobiotic crayfishes (decapoda: Cambaridae) of shelta cave, alabama, usa. **Subterranean Biology**, v. 7, p. 35–40, 2009. Citado na página 19.
- CORMACK, R. Estimates of survival from the sighting of marked animals. **Biometrika**, JSTOR, v. 51, n. 3/4, p. 429–438, 1964. Citado na página 76.
- CORMACK, R. M. Log-linear models for capture-recapture. **Biometrics**, JSTOR, p. 395–413, 1989. Citado nas páginas 18 e 19.

COWEN, L.; SCHWARZ, C. J. The jolly–seber model with tag loss. **Biometrics**, Wiley Online Library, v. 62, n. 3, p. 699–705, 2006. Citado na página 19.

CROSBIE, S.; MANLY, B. Parsimonious modelling of capture-mark-recapture studies. **Biometrics**, JSTOR, p. 385–398, 1985. Citado na página 26.

DAIL, D.; MADSEN, L. Models for estimating abundance from repeated counts of an open metapopulation. **Biometrics**, Wiley Online Library, v. 67, n. 2, p. 577–587, 2011. Citado na página 18.

DARROCH, J. N. The multiple-recapture census: Ii. estimation when there is immigration or death. **Biometrika**, JSTOR, v. 46, p. 336–351, 1959. Citado na página 18.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 1–38, 1977. Citado na página 114.

DIEBOLT, J.; ROBERT, C. P. Estimation of finite mixture distributions through bayesian sampling. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 363–375, 1994. Citado na página 38.

DORAZIO, R. M.; MUKHERJEE, B.; ZHANG, L.; GHOSH, M.; JELKS, H. L.; JORDAN, F. Modeling unobserved sources of heterogeneity in animal abundance using a dirichlet process prior. **Biometrics**, Wiley Online Library, v. 64, n. 2, p. 635–644, 2008. Citado na página 113.

DUPUIS, J. A. Bayesian estimation of movement and survival probabilities from capture-recapture data. **Biometrika**, Oxford University Press, v. 82, n. 4, p. 761–772, 1995. Citado na página 74.

DUPUIS, J. A.; SCHWARZ, C. J. A bayesian approach to the multistate jolly–seber capture–recapture model. **Biometrics**, Wiley Online Library, v. 63, n. 4, p. 1015–1022, 2007. Citado nas páginas 26, 73 e 74.

DYK, D. A. V.; MENG, X.-L. The art of data augmentation. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 10, n. 1, p. 1–50, 2001. Citado na página 38.

EFFORD, M. Density estimation in live-trapping studies. **Oikos**, Wiley Online Library, v. 106, n. 3, p. 598–610, 2004. Citado nas páginas 74, 77 e 114.

EFFORD, M. G. Estimation of population density by spatially explicit capture–recapture analysis of data from area searches. **Ecology**, Wiley Online Library, v. 92, n. 12, p. 2202–2207, 2011. Citado na página 75.

EFFORD, M. G.; BORCHERS, D. L.; BYROM, A. E. Density estimation by spatially explicit capture–recapture: likelihood-based methods. In: **Modeling demographic processes in marked populations**. [S.l.]: Springer, 2009. p. 255–269. Citado nas páginas 75 e 77.

ERGON, T.; GARDNER, B. Separating mortality and emigration: modelling space use, dispersal and survival with robust-design spatial capture–recapture data. **Methods in Ecology and Evolution**, Wiley Online Library, v. 5, n. 12, p. 1327–1336, 2014. Citado na página 76.

FERGUSON, T. S. A bayesian analysis of some nonparametric problems. **The annals of statistics**, JSTOR, p. 209–230, 1973. Citado na página 114.

- _____. Prior distributions on spaces of probability measures. **The annals of statistics**, JSTOR, p. 615–629, 1974. Citado na página 114.
- GARDNER, B.; REPPUCCI, J.; LUCHERINI, M.; ROYLE, J. A. Spatially explicit inference for open populations: estimating demographic parameters from camera-trap studies. **Ecology**, Eco Soc America, v. 91, n. 11, p. 3376–3383, 2010. Citado nas páginas 75 e 76.
- GARDNER, B.; ROYLE, J. A.; WEGAN, M. T. Hierarchical models for estimating density from dna mark-recapture studies. **Ecology**, Eco Soc America, v. 90, n. 4, p. 1106–1115, 2009. Citado na página 75.
- GARDNER, B.; ROYLE, J. A.; WEGAN, M. T.; RAINBOLT, R. E.; CURTIS, P. D. Estimating black bear density using dna data from hair snares. **The Journal of Wildlife Management**, Wiley Online Library, v. 74, n. 2, p. 318–325, 2010. Citado na página 75.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. [S.l.]: Chapman and Hall/CRC, 1995. Citado na página 35.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical science**, JSTOR, p. 457–472, 1992. Citado nas páginas 66 e 106.
- GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. **Markov chain Monte Carlo in practice**. [S.l.]: CRC press, 1995. Citado na página 41.
- GLENNIE, R.; BORCHERS, D. L.; MURCHIE, M.; HARMSSEN, B.; FOSTER, R. Open population maximum likelihood spatial capture-recapture. **Biometrics**, 2017. Citado nas páginas 75, 76, 77 e 114.
- GNASPINO, P. Population ecology of goniosoma spelaum, a cavernicolous harvestman from south-eastern brazil (arachnida: Opiliones: Gonyleptidae). **Journal of Zoology**, Wiley Online Library, v. 239, n. 3, p. 417–435, 1996. Citado na página 19.
- HARA, M. R.; PINTO-DA-ROCHA, R. Systematic review and cladistic analysis of the genus Eusarcus Perty 1833 (Arachnida, Opiliones, Gonyleptidae). **Zootaxa**, v. 17441, p. 1–136, 2010. ISSN 0101-8175. Citado na página 20.
- JOLLY, G. M. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. **Biometrika**, JSTOR, v. 52, p. 225–247, 1965. Citado nas páginas 18 e 76.
- JOSEPH, L. N.; ELKIN, C.; MARTIN, T. G.; POSSINGHAM, H. P. Modeling abundance using n-mixture models: the importance of considering ecological mechanisms. **Ecological Applications**, Wiley Online Library, v. 19, n. 3, p. 631–642, 2009. Citado na página 37.
- JUGOVIC, J.; PRAPROTNIK, E.; BUZAN, E.; LUŽNIK, M. Estimating population size of the cave shrimp trogloniscus anophthalmus (crustacea, decapoda, caridea) using mark–release–recapture data. **Animal Biodiversity and Conservation**, v. 38, n. 1, p. 77–86, 2015. Citado na página 19.
- KENDALL, W. L. Robustness of closed capture–recapture methods to violations of the closure assumption. **Ecology**, Wiley Online Library, v. 80, n. 8, p. 2517–2525, 1999. Citado na página 18.

- KERY, M.; GARDNER, B.; STOECKLE, T.; WEBER, D.; ROYLE, J. A. Use of spatial capture-recapture modeling and dna data to estimate densities of elusive animals. **Conservation Biology**, Wiley Online Library, v. 25, n. 2, p. 356–364, 2011. Citado na página 75.
- KING, R.; BROOKS, S. Bayesian model discrimination for multiple strata capture-recapture data. **Biometrika**, Oxford University Press, v. 89, n. 4, p. 785–806, 2002. Citado na página 74.
- LAAKE, J.; REXSTAD, E. Rmark—an alternative approach to building linear models in mark. **Program MARK: a gentle introduction**, p. C1–C113, 2008. Citado na página 19.
- LAPLACE, P. S. Sur les naissances, les mariages et les morts. **Histoire de l'Academie Royale des Sciences**, Impr. Royale Paris, p. 693, 1783. Citado na página 18.
- LIEURY, N.; DEVILLARD, S.; BESNARD, A.; GIMENEZ, O.; HAMEAU, O.; PONCHON, C.; MILLON, A. Designing cost-effective capture-recapture surveys for improving the monitoring of survival in bird populations. **Biological Conservation**, Elsevier, v. 214, p. 233–241, 2017. Citado na página 19.
- LINCOLN, F. C. Calculating waterfowl abundance on the basis of banding returns. US Dept. of Agriculture, 1930. Citado na página 17.
- LINK, W. A.; BARKER, R. J. Modeling association among demographic parameters in analysis of open population capture–recapture data. **Biometrics**, Wiley Online Library, v. 61, n. 1, p. 46–54, 2005. Citado nas páginas 18, 19 e 30.
- MACDONALD, I. L.; ZUCCHINI, W. **Hidden Markov and other models for discrete-valued time series**. [S.l.]: CRC Press, 1997. v. 110. Citado na página 73.
- MARQUES, T. A.; THOMAS, L.; ROYLE, J. A. A hierarchical model for spatial capture-recapture data: comment. **Ecology**, JSTOR, v. 92, n. 2, p. 526–528, 2011. Citado na página 75.
- MARQUES, T. A.; THOMAS, L.; WARD, J.; DIMARZIO, N.; TYACK, P. L. Estimating cetacean population density using fixed passive acoustic sensors: An example with blainville's beaked whales. **The Journal of the Acoustical Society of America**, ASA, v. 125, n. 4, p. 1982–1994, 2009. Citado na página 19.
- MATECHOU, E.; PLEDGER, S.; EFFORD, M.; MORGAN, B. J.; THOMSON, D. L. Estimating age-specific survival when age is unknown: open population capture–recapture models with age structure and heterogeneity. **Methods in Ecology and Evolution**, Wiley Online Library, v. 4, n. 7, p. 654–664, 2013. Citado na página 19.
- MCCREA, R. S.; MORGAN, B. J. **Analysis of capture-recapture data**. [S.l.]: CRC Press, 2014. Citado na página 18.
- MCDONALD, T. L.; AMSTRUP, S. C.; MANLY, B. F. Tag loss can bias jolly-seber capture-recapture estimates. **Wildlife Society Bulletin**, JSTOR, p. 814–822, 2003. Citado na página 19.
- MELLO-LEITÃO, C. d. Oito novos opiliões do espírito santo. **Boletim do Museu nacional**, JSTOR, p. 14: 1–11, 1942. Citado na página 20.

- MEREDITH, M.; KRUSCHKE, J. **HDIInterval: highest (posterior) density intervals. R package version 0.1.3**. 2017. Citado na página 45.
- MOLDOVAN, O.; FEJÉR, A. Population size and dispersal patterns for a drimeotus (coleoptera, leiodidae, leptodirini) cave population. **Subterranean Biology**, Pensoft Publishers, v. 11, p. 31, 2013. Citado na página 19.
- MONTE, B. G. O. **Ecologia populacional do amblipígio Charinus eleonora**e Baptista e Giupponi 2003 (Arachnida: Amblypygi: Charinidae) e dos opiliões Iandumoema uai Pinto da Rocha 1996, Eusarcus cavernicola Hara e Pinto da Rocha 2010 e Eusarcus aduncus Mello Leitao 1942 (Arachnida: Opiliones: Gonyleptidae) da Gruta Olhos d Agua, Itacarambi, Minas Gerais. [S.l.]: Tese de Doutorado, Departamento de Ecologia e Biologia Evolutiva, Universidade Federal de São Carlos, 2017. Citado nas páginas 11, 19, 20, 21, 63, 103, 105 e 113.
- O'CONNELL, A. F.; NICHOLS, J. D.; KARANTH, K. U. **Camera traps in animal ecology: methods and analyses**. [S.l.]: Springer Science & Business Media, 2010. Citado na página 75.
- O'SHEA, T. J.; ELLISON, L. E.; STANLEY, T. R. Survival estimation in bats: historical overview, critical appraisal, and suggestions for new approaches. **Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters**. Island Press, Washington, DC, USA, p. 297–336, 2004. Citado na página 19.
- OTIS, D. L.; BURNHAM, K. P.; WHITE, G. C.; ANDERSON, D. R. Statistical inference from capture data on closed animal populations. **Wildlife monographs**, JSTOR, p. 3–135, 1978. Citado na página 18.
- PETERSEN, C. The yearly immigration of young plaice into the limfjord from the german sea. **Report of the Danish Biological Station**, v. 6, p. 1–48, 1896. Citado na página 17.
- PINTO-DA-ROCHA, R. Iandumoema uai, a new genus and species of troglobitic harvestman from brazil (arachnida, opiliones, gonyleptidae). **Revista brasileira de Zoologia**, SciELO Brasil, v. 13, n. 4, p. 843–848, 1996. Citado na página 20.
- PLEDGER, S.; POLLOCK, K. H.; NORRIS, J. L. Open capture–recapture models with heterogeneity: Ii. jolly-seber model. **Biometrics**, Wiley Online Library, v. 66, n. 3, p. 883–890, 2010. Citado na página 19.
- POLLOCK, K. H. A capture-recapture design robust to unequal probability of capture. **The Journal of Wildlife Management**, JSTOR, v. 46, n. 3, p. 752–757, 1982. Citado na página 30.
- PRADEL, R. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. **Biometrics**, JSTOR, p. 703–709, 1996. Citado nas páginas 18 e 19.
- R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2014. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Citado nas páginas 45, 66, 86 e 106.
- RAABE, J. K.; GARDNER, B.; HIGHTOWER, J. E. A spatial capture–recapture model to estimate fish survival and location from linear continuous monitoring arrays. **Canadian Journal of Fisheries and Aquatic Sciences**, NRC Research Press, v. 71, n. 1, p. 120–130, 2013. Citado na página 76.

RIVERS, N. M.; BUTLIN, R. K.; ALTRINGHAM, J. D. Autumn swarming behaviour of natterer's bats in the uk: population size, catchment area and dispersal. **Biological conservation**, Elsevier, v. 127, n. 2, p. 215–226, 2006. Citado na página 19.

RODRIGUES, J. Bayesian analysis of zero-inflated distributions. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 32, n. 2, p. 281–289, 2003. Citado na página 38.

ROYLE, J. A. N-mixture models for estimating population size from spatially replicated counts. **Biometrics**, Wiley Online Library, v. 60, n. 1, p. 108–115, 2004. Citado na página 37.

ROYLE, J. A.; CHANDLER, R. B.; SOLLMANN, R.; GARDNER, B. **Spatial capture-recapture**. [S.l.]: Academic Press, 2013. Citado nas páginas 18, 76 e 78.

ROYLE, J. A.; CHANDLER, R. B.; YACKULIC, C.; NICHOLS, J. D. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. **Methods in Ecology and Evolution**, Wiley Online Library, v. 3, n. 3, p. 545–554, 2012. Citado na página 75.

ROYLE, J. A.; DORAZIO, R. M. Hierarchical models of animal abundance and occurrence. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer, v. 11, n. 3, p. 249–263, 2006. Citado nas páginas 36 e 37.

_____. **Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities**. [S.l.]: Academic Press, 2008. Citado nas páginas 18 e 30.

ROYLE, J. A.; FULLER, A. K.; SUTHERLAND, C. Unifying population and landscape ecology with spatial capture–recapture. **Ecography**, Wiley Online Library, v. 41, n. 3, p. 444–456, 2018. Citado na página 77.

ROYLE, J. A.; GOPALASWAMY, A. M.; DORAZIO, R. M.; NICHOLS, J. D.; JATHANNA, D.; PARAMESHWARAN, R.; KARANTH, K. U. Concepts: Assessing tiger population dynamics using capture–recapture sampling. In: **Methods For Monitoring Tiger And Prey Populations**. [S.l.]: Springer, 2017. p. 163–189. Citado nas páginas 19 e 75.

ROYLE, J. A.; KÉRY, M.; GAUTIER, R.; SCHMID, H. Hierarchical spatial models of abundance and occurrence from imperfect survey data. **Ecological Monographs**, Wiley Online Library, v. 77, n. 3, p. 465–481, 2007. Citado na página 75.

ROYLE, J. A.; KERY, M.; GUELAT, J. Spatial capture-recapture models for search-encounter data. **Methods in Ecology and Evolution**, Wiley Online Library, v. 2, n. 6, p. 602–611, 2011. Citado na página 75.

ROYLE, J. A.; NICHOLS, J. D.; KARANTH, K. U.; GOPALASWAMY, A. M. A hierarchical model for estimating density in camera-trap studies. **Journal of Applied Ecology**, Wiley Online Library, v. 46, n. 1, p. 118–127, 2009. Citado na página 77.

ROYLE, J. A.; YOUNG, K. V. A hierarchical model for spatial capture-recapture data. **Ecology**, Eco Soc America, v. 89, n. 8, p. 2281–2289, 2008. Citado na página 75.

- RUSSELL, R. E.; ROYLE, J. A.; DESIMONE, R.; SCHWARTZ, M. K.; EDWARDS, V. L.; PILGRIM, K. P.; MCKELVEY, K. S. Estimating abundance of mountain lions from unstructured spatial sampling. **The Journal of Wildlife Management**, Wiley Online Library, v. 76, n. 8, p. 1551–1561, 2012. Citado na página 75.
- SCHOFIELD, M. R.; BARKER, R. J. Full open population capture–recapture models with individual covariates. **Journal of agricultural, biological, and environmental statistics**, Springer, v. 16, n. 2, p. 253–268, 2011. Citado nas páginas 19 e 114.
- SCHWARZ, C. J. The jolly-seber model: more than just abundance. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer, v. 6, n. 2, p. 195–205, 2001. Citado nas páginas 18 e 29.
- SCHWARZ, C. J.; ARNASON, A. N. A general methodology for the analysis of capture–recapture experiments in open populations. **Biometrics**, JSTOR, p. 860–873, 1996. Citado nas páginas 18, 19 e 26.
- _____. Jolly-seber models in mark. **Program MARK “A Gentle Introduction”**. Online, p. 1–53, 2006. Citado nas páginas 19 e 29.
- SCHWARZ, C. J.; SCHWEIGERT, J. F.; ARNASON, A. N. Estimating migration rates using tag-recovery data. **Biometrics**, JSTOR, p. 177–193, 1993. Citado na página 74.
- SCHWARZ, C. J.; SEBER, G. A. Estimating animal abundance: review iii. **Statistical Science**, JSTOR, p. 427–456, 1999. Citado na página 18.
- SEBER, G. A review of estimating animal abundance. **Biometrics**, JSTOR, p. 267–292, 1986. Citado na página 18.
- SEBER, G. A. **The estimation of animal abundance**. [S.l.]: Griffin London, 1982. Citado na página 18.
- _____. A review of estimating animal abundance ii. **International Statistical Review/Revue Internationale de Statistique**, JSTOR, p. 129–166, 1992. Citado na página 18.
- SEBER, G. A. F. A note on the multiple-recapture census. **Biometrika**, JSTOR, p. 249–259, 1965. Citado nas páginas 18 e 76.
- _____. **The estimation of animal abundance and related parameters**. New York (USA) Macmillan, 1982. Citado nas páginas 18 e 74.
- SOLLMANN, R.; FURTADO, M. M.; GARDNER, B.; HOFER, H.; JÁCOMO, A. T.; TÔRRES, N. M.; SILVEIRA, L. Improving density estimates for elusive carnivores: accounting for sex-specific detection and movements using spatial capture–recapture models for jaguars in central brazil. **Biological Conservation**, Elsevier, v. 144, n. 3, p. 1017–1024, 2011. Citado na página 75.
- SOLLMANN, R.; GARDNER, B.; CHANDLER, R. B.; ROYLE, J. A.; SILLETT, T. S. An open-population hierarchical distance sampling model. **Ecology**, Wiley Online Library, v. 96, n. 2, p. 325–331, 2015. Citado na página 76.
- SPIEGELHALTER, D. J.; BEST, N.; CARLIN, B. P.; LINDE, A. Van der. **Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models**. [S.l.], 1998. Citado na página 66.

SPIEGLHALTER, D.; BEST, N.; CARLIN, B.; LINDE, A. Van der. Bayesian measures of model complexity and fit (with discussion). **Journal of the Royal Statistical, Series B**, v. 64, p. 583–616, 2002. Citado na página 106.

TANNER, M. A.; WONG, W. H. The calculation of posterior distributions by data augmentation. **Journal of the American statistical Association**, Taylor & Francis, v. 82, n. 398, p. 528–540, 1987. Citado na página 38.

THOMPSON, C. M.; ROYLE, J. A.; GARNER, J. D. A framework for inference about carnivore density from unstructured spatial sampling of scat using detector dogs. **The Journal of Wildlife Management**, Wiley Online Library, v. 76, n. 4, p. 863–871, 2012. Citado na página 75.

THOMSON, D. L.; COOCH, E. G.; CONROY, M. J. **Modeling demographic processes in marked populations**. [S.l.]: Springer Science & Business Media, 2008. v. 3. Citado na página 18.

TOWNSEND, C. R.; BEGON, M.; HARPER, J. L. *et al.* **Essentials of ecology**. [S.l.]: Blackwell Science, 2003. Citado na página 17.

TRAJANO, E. Movements of cave bats in southeastern brazil, with emphasis on the population ecology of the common vampire bat, *desmodus rotundus* (chiroptera). **Biotropica**, JSTOR, p. 121–129, 1996. Citado na página 19.

VENARSKY, M. P.; HURY, A. D.; BENSTEAD, J. P. Re-examining extreme longevity of the cave crayfish *orconectes australis* using new mark–recapture data: a lesson on the limitations of iterative size-at-age models. **Freshwater Biology**, Wiley Online Library, v. 57, n. 7, p. 1471–1481, 2012. Citado na página 19.

WHITE, G. C.; BURNHAM, K. P. Program mark: survival estimation from populations of marked animals. **Bird study**, Taylor & Francis, v. 46, n. sup1, p. S120–S139, 1999. Citado na página 19.

WILLEMART, R. H.; GNASPINI, P. Spatial distribution, mobility, gregariousness, and defensive behaviour in a brazilian cave harvestman *goniosoma albiscriptum* (arachnida, opiliones, gonyleptidae). **Animal Biology**, Brill, v. 54, n. 3, p. 221–235, 2004. Citado na página 19.

WU, G.; HOLAN, S. H. Bayesian hierarchical multi-population multistate jolly–seber models with covariates: Application to the pallid sturgeon population assessment program. **Journal of the American Statistical Association**, Taylor & Francis, v. 112, n. 518, p. 471–483, 2017. Citado nas páginas 19, 74 e 114.

ZUCCHINI, W.; MACDONALD, I. L.; LANGROCK, R. **Hidden Markov models for time series: an introduction using R**. [S.l.]: Chapman and Hall/CRC, 2016. Citado na página 73.

GRÁFICOS DE CONVERGÊNCIA

A seguir, exibimos os gráficos dos valores gerados da cadeia MCMC para os parâmetros populacionais estimados nas aplicações apresentados nos Capítulos 4 e 6.

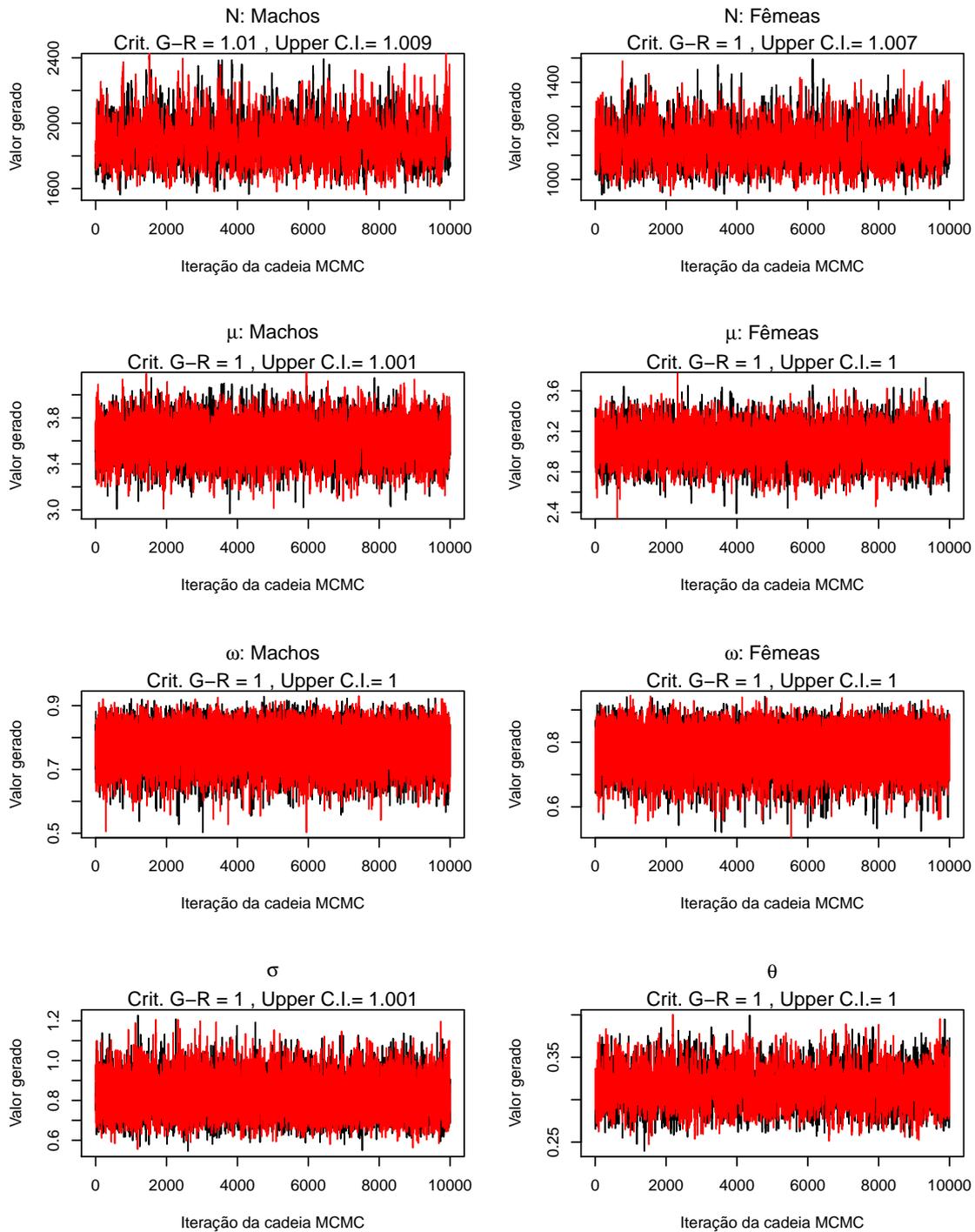


Figura 27 – Valores gerados para os parâmetros na primeira aplicação real.

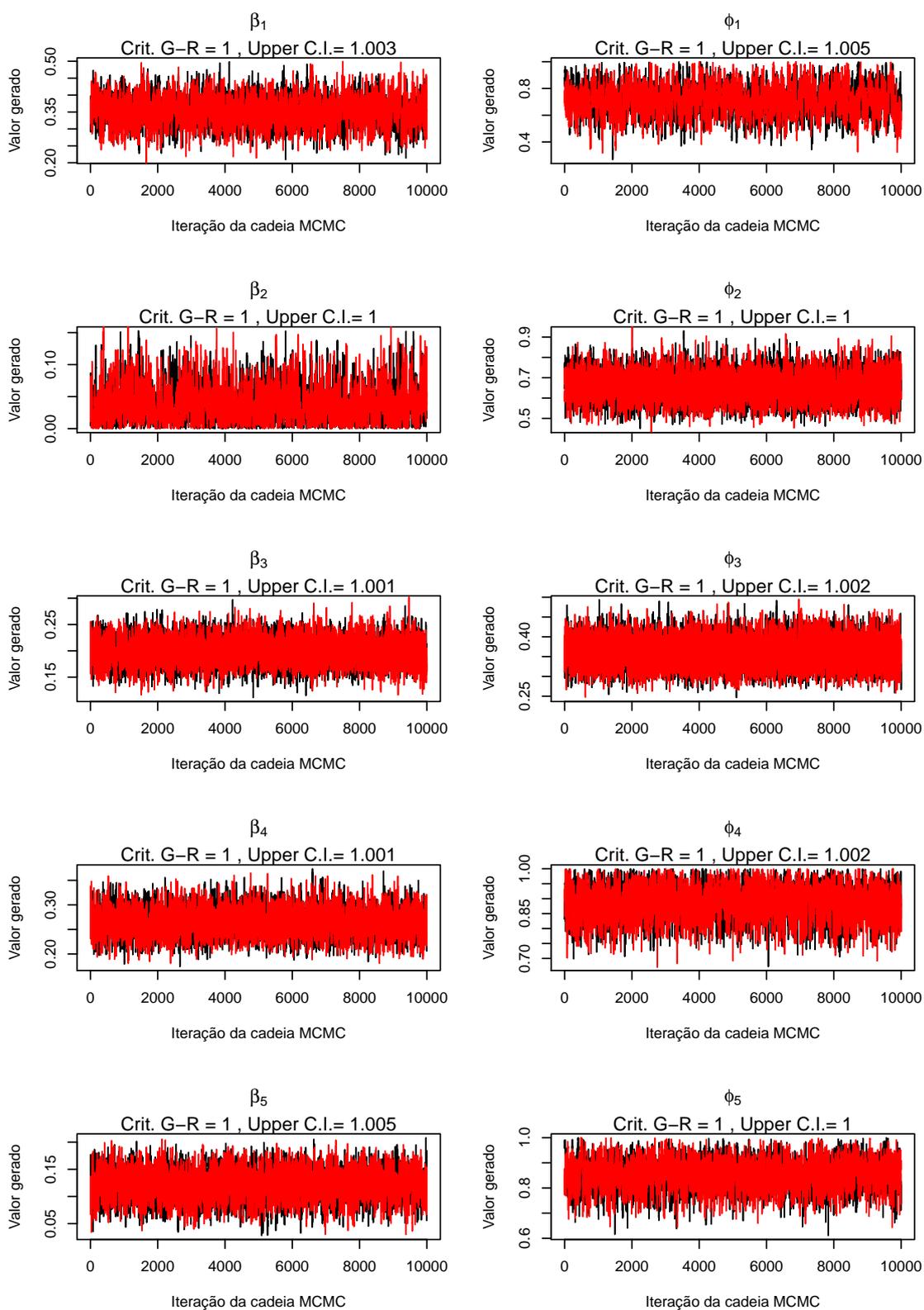


Figura 28 – Valores gerados para os parâmetros na primeira aplicação real.

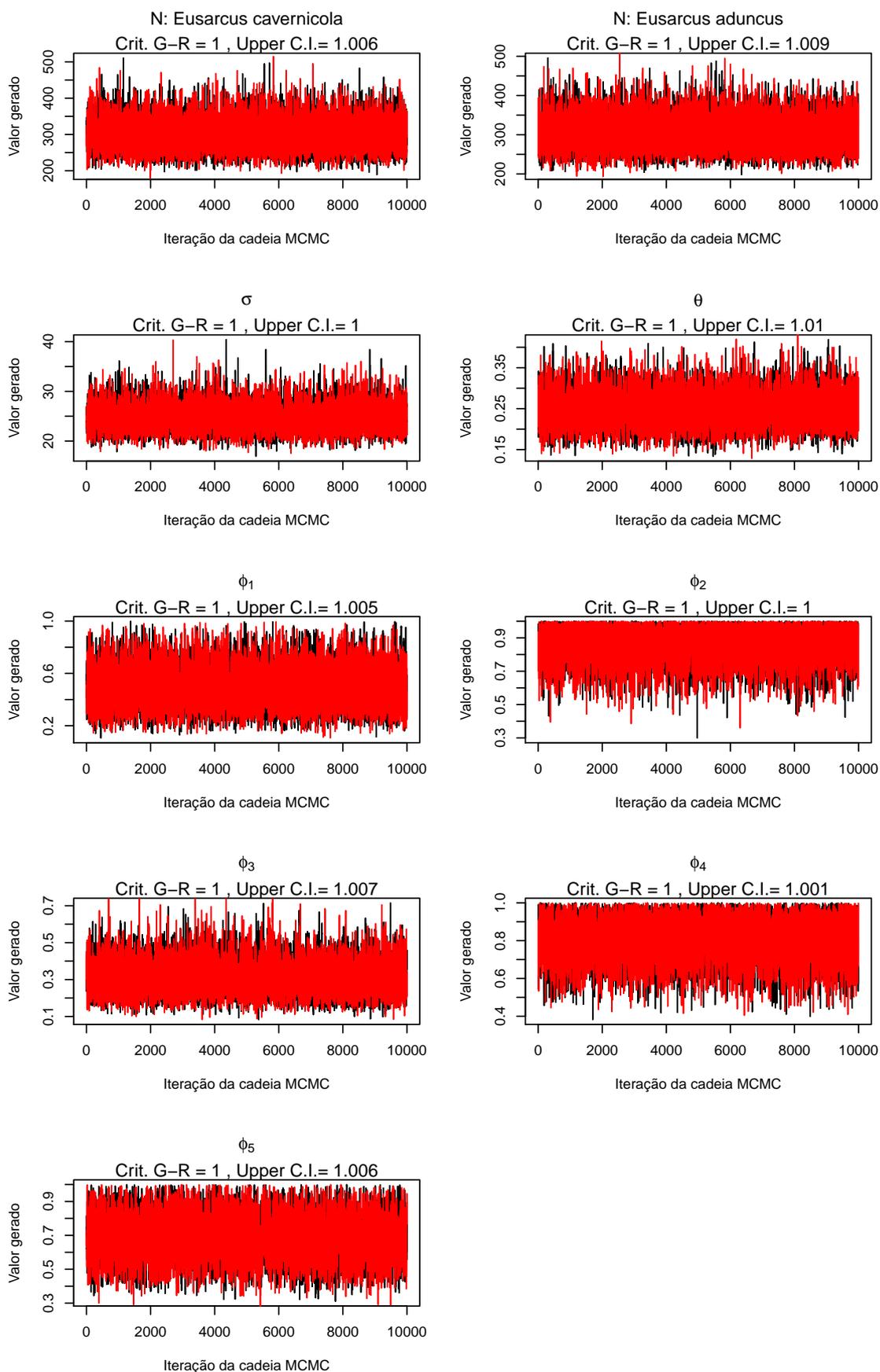


Figura 29 – Valores gerados para os parâmetros na segunda aplicação real.

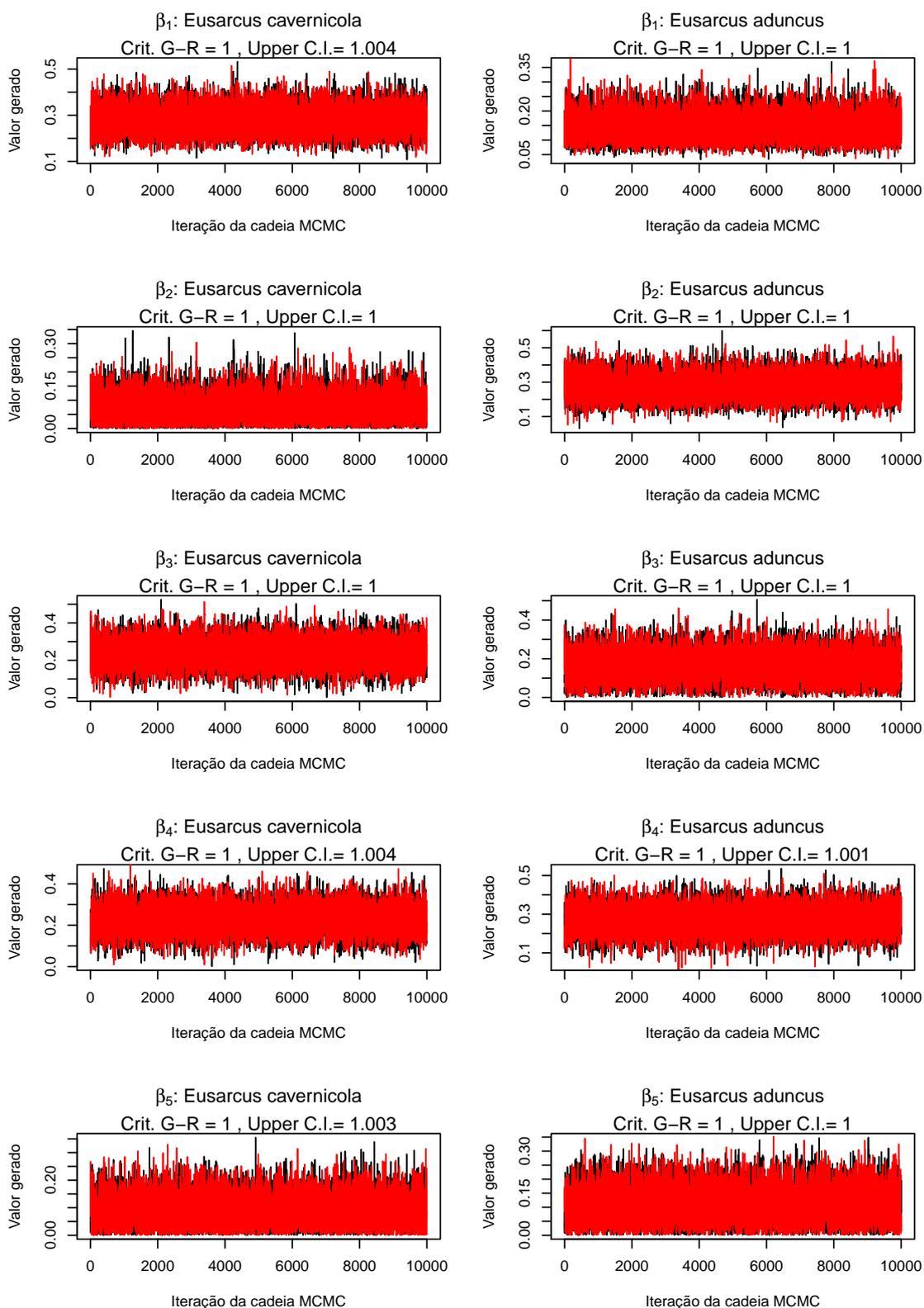


Figura 30 – Valores gerados para os parâmetros na segunda aplicação real.