

**ESTIMAÇÃO DE RIQUEZA DE ESPÉCIES DE  
MACROINVERTEBRADOS BENTÔNICOS  
UTILIZANDO O MODELO DE MICHAELIS-  
MENTEN: UMA ABORDAGEM BAYESIANA**

**EUSTÁQUIO JOSÉ MACHADO**

**2010**



**EUSTÁQUIO JOSÉ MACHADO**

**ESTIMAÇÃO DE RIQUEZA DE ESPÉCIES DE  
MACROINVERTEBRADOS BENTÔNICOS UTILIZANDO O MODELO  
DE MICHAELIS-MENTEN: UMA ABORDAGEM BAYESIANA**

Tese apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de “Doutor”.

Orientador

Prof. Dr. Joel Augusto Muniz

LAVRAS  
MINAS GERAIS – BRASIL  
2010



**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca da UFLA**

Machado, Eustáquio José.

Estimação de riqueza de espécies de macroinvertebrados bentônicos utilizando o modelo de Michaelis-Menten: uma abordagem bayesiana / Eustáquio José Machado. – Lavras : UFLA, 2010.

94 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2010.

Orientador: Joel Augusto Muniz.

Bibliografia.

1. Inferência Bayesiana. 2. Modelo não-linear. 3. Estimação em Espécies. 4. Método de Monte Carlo via cadeias de Markov. I. Universidade Federal de Lavras. II. Título.

CDD – 519.542



**EUSTÁQUIO JOSÉ MACHADO**

**ESTIMAÇÃO DE RIQUEZA DE ESPÉCIES DE  
MACROINVERTEBRADOS BENTÔNICOS UTILIZANDO O MODELO  
DE MICHAELIS-MENTEN: UMA ABORDAGEM BAYESIANA**

Tese apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de “Doutor”.

APROVADA em 18 de Fevereiro de 2010

Profa. Dra. Taciana Villela Savian	ESALQ/USP
Prof. Dr. Luiz Alberto Beijo	UNIFAL
Prof. Dr. Júlio Neil Cassa Louzada	UFLA
Profa. Dra. Thelma Sáfadi	UFLA

Prof. Dr. Joel Augusto Muniz  
UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS – BRASIL





Aos meus pais João Luiz e Maria das Dores (Mariquinha) (*in memoriam*), pelo amor, bondade e confiança.

A minha irmã Iracema (*in memoriam*) por ter me oferecido a oportunidade do conhecimento, incentivo, apoio e amor.

#### OFEREÇO

A Francisca minha esposa, peça indispensável na realização deste estudo, pela sua contribuição, apoio, confiança, paciência e amor a mim dedicado.

Ao meu filho Max que minha determinação sirva de exemplo a ser seguido e continuidade.

A minha irmãzinha Dalva pelo amor, convivência, carinho e dedicação nas horas mais difíceis.

DEDICO



## AGRADECIMENTOS

Ao senhor Deus, por ter me concedido a graça da vida!

À Universidade Federal do Acre e aos professores do Centro de Ciências Exatas e Tecnológicas da UFAC, pela oportunidade e confiança depositada.

À Universidade Federal de Lavras, pela oportunidade da realização deste curso.

A coordenação de aperfeiçoamento de pessoal de nível superior (CAPES), pela concessão da bolsa de estudo.

Ao meu orientador professor Joel Augusto Muniz, pelo estímulo, confiança, paciência, apoio e orientação no desenvolvimento deste estudo e por ser um verdadeiro amigo.

A professora Thelma Sáfadi, pela co-orientação, confiança, ensinamentos, amizade e carinho.

Aos professores e funcionários do DEX-UFLA, pela amizade e carinho.

Ao meu amigo mestre Diego pela gentileza em ceder os dados de sua pesquisa e valiosa contribuição nas discussões ecológicas para realização deste estudo.

Ao meu grande amigo professor Manoel Domingos (UFAC), pelo incentivo, confiança e ética, que mesmo de longe, tenho certeza que torce pelo meu sucesso, obrigado!

Ao meu amigo Edcarlos grande parceiro e por sua valiosa contribuição para a realização deste estudo.

Aos colegas do Mestrado e Doutorado, em especial a Taciana pela amizade e contribuição neste estudo.

Aos meus amigos Paulinho, Janice e Renata pelo apoio e amizade.

Aos amigos Altemir, Naje e Pupio pelo apoio e torcida pelo meu sucesso.



## SUMÁRIO

	<b>Página</b>
LISTA DE TABELAS .....	i
LISTA DE FIGURAS .....	ii
RESUMO .....	iii
ABSTRACT .....	iv
1 INTRODUÇÃO .....	01
1.1 Justificativa .....	03
2 REFERENCIAL TEÓRICO .....	05
2.1 Cenários ecológicos .....	05
2.1.1 Estimação de riqueza em espécies .....	05
2.1.2 Métodos utilizados em inventários de riquezas em espécies .....	07
2.1.3 Macro-invertebrados aquáticos: conhecendo a riqueza para estabelecimento de programas de monitoramento ambiental .....	10
2.1.4 Parque Nacional da Serra do Divisor (PNSD) .....	15
2.2 Regressão linear e não-linear .....	17
2.3 Inferência Bayesiana .....	21
2.3.1 Distribuições <i>a priori</i> .....	22
2.3.2 Função de Verossimilhança .....	23
2.3.3 Teorema de Bayes .....	24
2.3.4 Distribuições <i>a posteriori</i> .....	27
2.3.5 Intervalos de Credibilidade e de Máxima Densidade <i>a posteriori</i> (HPD).....	29
2.4 Simulação Estocástica .....	30
2.4.1 Método de Monte Carlo Simples .....	31
2.4.2 Métodos de Monte Carlo via cadeias de Markov (MCMC) .....	32



2.4.2.1 Amostrador de Gibbs .....	32
2.4.2.2 Algoritmo de Metropolis-Hasting .....	35
2.4.3 Erro de Monte Carlo .....	37
2.4.4 Avaliação da convergência .....	38
2.4.4.1 Monitoração formal da convergência .....	39
2.4.5 Autocorrelação .....	40
2.5 Modelo não-linear de Michaelis-Menten .....	41
3 MATERIAL E MÉTODO .....	43
3.1 Material .....	43
3.1.1 Dados analisados no estudo .....	43
3.2 Método .....	51
3.2.1 Obtenção da distribuição <i>a posteriori</i> conjunta para os parâmetros do modelo de Michaelis-Menten relacionado a <i>dados1</i> .....	51
3.2.2 Obtenção da distribuição <i>a posteriori</i> conjunta para os parâmetros do modelo de Michaelis-Menten relacionado a <i>dados2</i> .....	56
3.2.3 Obtenção da distribuição <i>a posteriori</i> conjunta para os parâmetros do modelo de Michaelis-Menten relacionado a <i>dados3</i> .....	59
3.3 Implementação das análises .....	62
4 RESULTADOS E DISCUSSÃO .....	63
4.1 Ajuste do modelo de Michaelis-Menten aos <i>dados1</i> , <i>dados2</i> e <i>dados3</i> .....	63
5 CONCLUSÕES .....	73
REFERÊNCIAS BIBLIOGRÁFICAS.....	74
APÊNDICE A .....	82
APÊNDICE B .....	88
APÊNDICE C .....	91





## LISTA DE TABELAS

	<b>Página</b>
TABELA 1 Representação da estrutura dos dados coletados em dez igarapés (IG) de águas claras .....	46
TABELA 2 Representação da estrutura da amostra 1 coletada em seis igarapés (IG) de águas claras .....	48
TABELA 3 Representação da estrutura da amostra 2 coletada em seis igarapés (IG) de águas claras .....	48
TABELA 4 Representação da estrutura da amostra 3 coletada em seis igarapés (IG) de águas claras .....	49
TABELA 5 Táxons acumulados por unidades amostrais (IG) relacionados a <i>dados1</i> .....	50
TABELA 6 Táxons acumulados por unidades amostrais (IG) relacionados a <i>dados2</i> .....	51
TABELA 7 Táxons acumulados por unidades amostrais (IG) relacionados a <i>dados3</i> .....	51
TABELA 8 Critérios de Convergência relacionados a <i>dados1</i> .....	64
TABELA 9 Critérios de Convergência relacionados a <i>dados2</i> .....	65
TABELA 10 Critérios de Convergência relacionados a <i>dados3</i> .....	65
TABELA 11 Moda, mediana, média, desvio padrão (DP), erro de Monte Carlo (EMC) <i>a posteriori</i> dos parâmetros das distribuições com seus respectivos intervalos HPD 95% relacionados a <i>dados1</i> .....	67
TABELA 12 Moda, mediana, média, desvio padrão (DP), erro de Monte Carlo (EMC) <i>a posteriori</i> dos parâmetros das distribuições com seus respectivos intervalos HPD 95% relacionados a <i>dados2</i> .....	67
TABELA 13 Moda, mediana, média, desvio padrão (DP), erro de Monte Carlo (EMC) <i>a posteriori</i> dos parâmetros das distribuições com seus respectivos intervalos HPD 95% relacionados a <i>dados3</i> .....	67

## LISTA DE FIGURAS

	<b>Página</b>
FIGURA 1 Curva do coletor .....	09
FIGURA 2 Eventos mutuamente exclusivos .....	25
FIGURA 3 Mapa do PNSD, área norte, contendo as duas bacias de coleta. Bacia do rio Moa para coletas de águas claras e bacia do rio Azul para coleta de águas brancas. ....	44
FIGURA 4 Imagem dos riachos (igarapés) de águas claras (1,2 e 3) e águas brancas (4) .....	45
FIGURA 5 Gráfico de Traços dos parâmetros da cadeia <i>dados1</i> .....	64
FIGURA 6 Gráfico de Traços dos parâmetros da cadeia <i>dados2</i> .....	64
FIGURA 7 Gráfico de Traços dos parâmetros da cadeia <i>dados3</i> .....	65
FIGURA 8 Gráfico de densidade da distribuição <i>a posteriori</i> dos parâmetros <i>dados1</i> .....	68
FIGURA 9 Gráfico de densidade da distribuição <i>a posteriori</i> dos parâmetros <i>dados2</i> .....	68
FIGURA 10 Gráfico de densidade da distribuição <i>a posteriori</i> dos parâmetros <i>dados3</i> .....	68
FIGURA 11 Gráfico de densidade da distribuição <i>a posteriori</i> dos parâmetros. As curvas em vermelho, azul e preto são referentes aos <i>dados1</i> , <i>dados2</i> e <i>dados3</i> respectivamente.....	69
FIGURA 12 Gráfico de ajuste ao modelo relacionado à <i>dados1</i> utilizando como parâmetro para o modelo a moda <i>a posteriori</i> .....	70
FIGURA 13 Gráfico de ajuste ao modelo relacionado à <i>dados2</i> utilizando como parâmetro para o modelo a moda <i>a posteriori</i> .....	71
FIGURA 14 Gráfico de ajuste ao modelo relacionado à <i>dados3</i> utilizando como parâmetro para o modelo a moda <i>a posteriori</i> .....	72

## RESUMO

MACHADO, Eustáquio José. **Estimação de riqueza de espécies de macroinvertebrados bentônicos utilizando o modelo Michaelis-Menten: uma abordagem Bayesiana.** 2010. 94 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.\*

O levantamento do número de espécies é o primeiro passo em qualquer estudo sobre diversidade biológica e essencial para avaliação do *status* de conservação por ser a extinção um processo irreversível, e a falta de informação implicar em estratégias de conservação equivocadas, aumentando, conseqüentemente a perda de espécies que, em grande parte, ainda não foram estudadas. Neste estudo, estimou-se o número de espécies biológicas denominadas macroinvertebrados bentônicos, por meio do modelo de regressão não-linear conhecido no contexto ecológico como equação de Michaelis-Menten, a partir de amostras coletadas em igarapés (riachos) de águas brancas e claras em estudo de Lima (2008). A metodologia utilizada fez uso da inferência bayesiana para a estimação dos parâmetros do modelo por meio de métodos de Monte Carlo via cadeias de Markov usando o amostrador de *Gibbs* e o algoritmo de Metropolis-Hastings. Utilizou-se a função de autocorrelação para produzir independência para a amostra obtida. A análise de convergência das cadeias foi monitorada pelos critérios: Raftery & Lewis ((1992), Gelman & Rubin (1992), e Geweke (1992), sendo que a implementação das análises foi realizada por meio de rotinas e do pacote BOA executável no software R (2010). Os resultados encontrados indicaram o aumento considerável da precisão das estimativas dos parâmetros utilizando como critérios: o intervalo HPD (*Highest Posterior Density*), erro de Monte Carlo e desvio padrão. O modelo de Michaelis-Menten ajustou-se muito bem aos dados observacionais nas três massas de dados utilizadas. Em virtude dos resultados apresentados, conclui-se que são bastante satisfatórios e promissores com a utilização da abordagem Bayesiana em estudos de estimação em espécies na ecologia.

**Palavras chaves:** Inferência Bayesiana, modelo não-linear, estimação em espécies, método de Monte Carlo via cadeias de Markov, macroinvertebrados bentônicos.

---

Comitê Orientador: Joel Augusto Muniz – UFLA (orientador), Thelma Sáfyadi – UFLA

## ABSTRACT

MACHADO, Eustáquio José. **Estimating species richness of benthic macroinvertebrates using Michaelis-Menten model: a Bayesian approach.** 2010. 94 p. Thesis (Doctorate in Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, MG, \*

A survey of number of species is the first step in any study about biological diversity and essential for evaluating the conservation status being the extinction an irreversible process, and the lack of information implies in misdeeds in conservation strategies, increasing consequently, the species loss which in most part, have not been studied yet. In this study, the number of biological species called benthic macroinvertebrates was estimated through the non-linear regression model known as Michaelis-Menten equation in the ecological context, from samples collected in clear and clean-water stream in Lima's study (2008). The methodology used was the Bayesian inference for the model parameters estimation through Monte Carlo methods via Markov chains using the Gibbs' sampler and Metropolis-Hastings algorithm. The autocorrelation function was used for producing independence to the sample obtained. The convergence analysis of the chains was monitored through criteria: Raftery & Lewis (1992), Gelman & Rubin (1992) and Geweke (1992), so that the implementation of the analysis was carried out through routines and BOA pack executable in the R (2010) software. The results found indicated a considerable increasing of the accuracy in parameter estimation using HPD (highest posterior density) interval, Monte Carlo error and standard deviation as criteria. Michaelis-Menten model fitted well to observational data in three data mass used. Because of the results presented, one can conclude they are quite satisfactory and promising by using the Bayesian approach in ecological studies on estimation of species.

Keywords: Bayesian inference, non-linear model, estimation of species, Monte Carlo method via Markov chain, benthic macroinvertebrates.

---

Guidance Comitee: Prof. DSc. Joel Augusto Muniz – UFLA (Adviser) and Profa.Dsc. Thelma Sáfiadi.

## 1 INTRODUÇÃO

Da Amazônia ao cerrado, da mata atlântica à caatinga, do pantanal aos campos sulinos, o Brasil exibe uma das mais ricas diversidades do mundo em ambientes naturais e espécies de plantas e animais do planeta. No entanto, a informação de quantos (riqueza) e quais organismos (composição) pode ser encontrada em determinados ambientes gera desdobramento que podem fornecer subsídios para a escolha de áreas prioritárias para a conservação e permitem comparações entre comunidades de áreas distintas com vistas a vislumbrar sobre eventos anteriores de incidência sobre as espécies bem como a adaptações evolutivas.

Inevitavelmente, ocorre na natureza processos de extinção em longo prazo resultante de interações intra e inter-específicas aliado a eventos naturais com implicações diretas e indiretas em comunidades de espécies dos diferentes ambientes existentes na terra. No entanto, este processo de extinção natural é por demais alterados em virtude da interferência humana a qual promove a quebra deste ciclo evolutivo antes do período proposto pelos eventos naturais, de maneira que muitas espécies desaparecem dos ecossistemas sem sequer serem descritas pela ciência.

A falta de informação sobre o levantamento do número de espécies e conseqüentemente a diversidade biológica implica em estratégias equivocadas, e com isso há a perda de espécies nas condições já abordadas anteriormente. Esta problemática amplia-se em magnitude quando percebemos que o Brasil possui vários *hotspots* (zona quente) de diversidade biológica sendo considerados verdadeiros tesouros naturais para vários grupos científicos no mundo todo.

Não é difícil aceitarmos que o aumento de registros de espécies novas em qualquer campanha amostral é resultante da ampliação do esforço de campo. Não obstante, ecólogos com atividades voltadas a pesquisas de campo têm

reconhecido as limitações em detectar as espécies existentes em um ambiente utilizando um número limitado de amostras.

A biota de determinado ecossistema é constituída de grupos que apresentam comportamentos diferentes dentro de suas populações. Estas características intensificam as dificuldades na retenção de medidas apropriadas para registros de espécies. Como exemplo, citamos a mobilidade de alguns vertebrados. Enquanto as capivaras (*Hydrochaeris hydrochaeris*) apresentam-se em grupos tanto isolados quanto em pequenos grupos, outros animais como os falcões (família Falconidae) apresentam vôos elevados e técnicas de contagem diferentes. Estas peculiaridades oneram as atividades de inventário faunístico e muitas vezes resultam em trabalhos poucos expressivos.

Considerando essa problemática para invertebrados, as dificuldades de estimação ganham proporções maiores, pois cerca de 90% da fauna conhecida em todo mundo é classificada dentro deste grupo. Neste grupo inserem-se os macroinvertebrados aquáticos (animais de água doce passível de percepção a olho nu), em especial a comunidade bentônica (MIB), caracterizada pelo seu importante papel nas relações ecológicas em fluxo de energia e comprovada utilidade para monitoramento dos diferentes fontes de impactos em sistemas aquáticos.

O presente estudo teve como objetivo estimar o número de espécies em uma população de macroinvertebrados bentônicos a partir de amostras coletadas em igarapés (riachos) de águas brancas e claras no Parque Nacional da Serra do Divisor (PNSD), uma Unidade de Conservação (UC), localizada no estado do Acre.

Para estimar o número de espécies utilizou-se de um dos principais métodos que tem como base as curva de acumulação em espécies fazendo uso de um modelo de regressão não-linear denominado equação de Michaelis-Menten sob um enfoque Bayesiano.

## 1.1 Justificativa

Os princípios norteadores que resultaram na execução deste trabalho têm sua origem do entendimento da necessidade de possibilitar ferramentas-resposta para dois problemas: a atualidade da escassez de água de qualidade e a necessidade de subsidiar elaboração de programas de conservação para mananciais naturais com macroinvertebrados bentônicos no oeste amazônico frente aos avanços urbanísticos esperados para região.

O reconhecimento da água doce de qualidade como um recurso limitado não é difícil de ser compreendido no mundo hodierno. Pouco menos de 0,005% do total de água presente no planeta está disponível para o consumo dos seres vivos sem comprometer o bem-estar e a saúde. Ainda assim, diante das necessidades vitais saciadas apenas com água é surpreendente o avanço dos usos múltiplos da água para atender os setores industriais, agricultura, comércio e atividades recreativas. Estima-se que os grandes utentes deste recurso podem ser categorizados conforme as seguintes porcentagens de consumo anual de água: uso em agricultura 69 %; uso em indústria 23 %; e uso doméstico (pessoal, familiar e municipal) 8 %.

Os números para o Brasil são surpreendentes. Nosso país concentra em torno de 12% da água doce do mundo disponível em rios e abriga o maior rio em extensão e volume do planeta, o Amazonas. Não obstante, pouco mais de 90% de todo território brasileiro tem incidência de chuvas superior a várias regiões do mundo durante o ano. Isto torna as condições climáticas e geológicas propícias à formação de uma extensa e densa rede de rios, com exceção do semi-árido, onde os rios são pobres e temporários. Essa água, no entanto, é distribuída de forma irregular, apesar da abundância em termos gerais. A Amazônia, onde estão as mais baixas concentrações populacionais, possui 78% da água superficial. Enquanto isso, no Sudeste, essa relação se inverte: a maior concentração populacional do País tem disponíveis 6% do total da água.

Os macroinvertebrados aquáticos apresentam uma grande diversidade de espécies e são encontrados em quase todos os tipos de habitats de água doce, sob diferentes condições ambientais. Conseqüentemente, tornam-se extremamente úteis para monitorar a qualidade da água, destacando-se os macroinvertebrados bentônicos por possuírem grande potencial como bioindicadores ambientais, apresentando desde espécies muito sensíveis até fortemente tolerantes a poluição, além de serem de fácil amostragem. O conhecimento deles é ainda muito escasso no Estado do Acre, mas ocorre a tendência ao crescimento deste estudo evidenciado por trabalhos pontuais de pós-graduação e iniciação científica apresentados em eventos em nível nacional e internacional (*Seminário UFAC-FLÓRIDA ocorrido nos anos de 2006 a 2009*), aliado à riqueza natural existente na região.

O problema em estimar o número de espécies de uma população a partir de amostras vem se tornando um tema importante em estudos ecológicos e considerando que a literatura é basicamente frequentista e sob a qual, várias tentativas de solução para estimar espécies foram realizadas, mas não se tem indicação de que os resultados obtidos tenham sido satisfatórios.

Métodos Bayesianos constitui um conjunto de técnicas alternativas à análise clássica, uma vez que pode facilitar a interpretação dos resultados e permitem uma incorporação de informações sobre os parâmetros do modelo, antes da observação dos dados. Portanto, neste estudo utilizar-se-á um modelo de regressão não-linear denominada equação de Michaelis-Menten para estimar o número de espécies de uma população usando uma abordagem sob a perspectiva Bayesiana.



## **2 REFERENCIAL TEÓRICO**

### **2.1 Cenários ecológicos**

#### **2.1.1 Estimação de riqueza em espécies**

A ecologia prima pelo estudo dos componentes de determinado ambiente (escalas diversas) visando não apenas descrevê-los, como também analisar os entes que o interagem promovendo o equilíbrio natural. No entanto, diagnósticos por meio de censos de todos os componentes de determinado grupo em qualquer ambiente é por demais onerosos devido a vários fatores (tempo, espaço, custo benefício, praticidade).

Segundo Wilson (1997), citado por Santos (2004), o conhecimento sobre a diversidade biológica do planeta continua escasso, apesar de todos os avanços da ciência. Para afirmar quantas espécies de um determinado grupo taxonômico existe no mundo, ou ainda em um pequeno fragmento de floresta, é extremamente complexo, e especialmente preocupante quando se considera o ritmo atual de destruição de ecossistemas naturais, aliado a altas taxas de extinção de espécies no planeta.

Diante deste cenário diferentes grupos da ciência vêm convergindo esforços no intuito de promover meios práticos e confiáveis para atingir o número total de espécies em qualquer ambiente. A estimação de riqueza de espécies nos diferentes grupos da biota de determinado ambiente trouxe muitos avanços para a determinação do número de espécies presentes em uma comunidade. Esta perspectiva impulsionou Melo (2004) a concluir que o inventário faunístico, tanto para a fauna quanto para flora, é uma ferramenta de extrema relevância para qualquer ecossistema, tornando-se o eixo principal para elaboração de políticas e gestões públicas que visam à conservação de ambientes. Conhecer as espécies existentes em um ambiente bem como a magnitude de suas populações geram bases norteadores para elaboração de

planos de manejo e promoção de técnicas que sustentem a diversidade local. No entanto, a falta de conhecimento da diversidade e riqueza de alguns grupos de organismos aquáticos geram sub-estimação devido ao conhecimento disponível ser ainda aquém da realidade de número de espécies (Martens et al., 2005).

A opção factível nesse caso é inventariar alguns grupos taxonômicos e algumas características físicas do ambiente, na tentativa de construir uma visão geral sobre toda a biota. O que se espera desses organismos selecionados para o inventário é que sua riqueza em um determinado local reflita a riqueza total, ou pelo menos de uma parte de seus componentes (Pearson, 1994) citado por Santos (2004). Isso não significa que os táxons não selecionados não sejam importantes, mas apenas que, diante da urgência de se obter dados úteis para projetos de conservação, é necessário concentrar esforços em alguns grupos. Esta temática tem sido discutida sobre a perspectiva de estabelecimento de espécies *surrogates*, as quais possuem benefícios amplos para outras espécies geralmente não contempladas em estudos, porém menos exigentes para permanência em determinados ambientes (Caro & O'doherty, 1998).

Com o avanço das respostas fornecidas pelos estimadores de riqueza e a ampliação de suas aplicações no contexto ecológico, diferentes estimadores foram elaborados com base em diversos parâmetros visando melhorar a qualidade de tais estimadores e com isto alcançar maior acurácia e precisão dos parâmetros. Todavia alguns problemas são ocorrentes em diferentes trabalhos promovendo diferentes discussões sobre o tema. Por exemplo, a mensuração da riqueza em espécies de um determinado grupo taxonômico em um determinado local com metodologias diversas, conforme exposto por Walther et al. (1995); Keddy et al. (1996), citado por Santos (2004), outro exemplo resulta na pobreza de informações sobre o esforço amostral aplicado a cada campanha de inventário de riqueza de diversos grupos de biota. Quando não se sabe quanto esforço de coleta foi aplicado a um determinado local, é difícil determinar se este apresenta

maior diversidade em espécies do que outros (Esbérard & Bergallo, 2008).

Diante disto, algumas sugestões foram aos poucos desenvolvidas visando o melhoramento da qualidade e da eficácia da informação gerada pela estimativa de riqueza, entre elas destaca-se a criação de protocolos de procedimentos em campo para realização de inventário biológico. Santos (2004) destaca em seu trabalho as conclusões de Longino et al. (1997), onde as conclusões obtidas com aplicação de protocolos foi mais eficaz para comparação entre ambientes diferentes do que as realizações de inventários com metodologias aleatórias.

### **2.1.2 Métodos utilizados em inventários de riquezas em espécies**

Segundo estudos de Santos (2004), os métodos de estimativa mais utilizados podem ser divididos em três categorias, com suas respectivas premissas e natureza de dados necessários para estimar o número total de espécies da comunidade. Os métodos do primeiro grupo se baseiam na distribuição de abundâncias das espécies (métodos paramétricos), os do segundo grupo no número de espécies raras nas amostras (métodos não-paramétricos), e os do terceiro nas curvas de acumulação de espécies ou curvas do coletor (extrapolação da curva de espécies).

Métodos Paramétricos utilizam dados de abundância de cada espécie para o ajuste a uma determinada distribuição. Conforme May (1975), citado por Santos (2004), a distribuição mais utilizada é a LogNormal, dado que a distribuição é simétrica, basta estimar a área da distribuição ainda não amostrada. As estimativas baseadas em ajustes de modelos de distribuição de abundância têm como principal vantagem, o fato de não exigirem coletas divididas em unidades amostrais. Por outro lado, para usar este método é extremamente importante testar estatisticamente o ajuste dos dados ao modelo em questão; no caso da curva LogNormal, é essencial que se obtenha uma moda

definida na distribuição de abundâncias. Além disso, é importante mencionar que o uso desse modelo exige conjuntos de dados com grande riqueza observada (no mínimo 100 espécies), uma vez que não é possível determinar com certeza o ajuste do modelo com amostras pequenas.

De acordo com Colwell & Coddington (1994); Schimit et al.(1999), apesar de certa elegância matemática e a existência de corpo teórico sobre a distribuição o método é pouco usado, porque depende de como se resolve a separação de classe de abundância, implicando em resultados muito diferentes, talvez ainda mais importante é que não foi incluído no programa de computador EstimateS (Colwell, 2009), sem dúvida o mais utilizado por pesquisadores atualmente.

Métodos Não-Paramétricos são bem mais simples e necessitam apenas da informação da riqueza observada na amostra e o número de espécies raras, definidas dependendo do método como aquelas com presença em 1, 2, 3, ..., unidades amostrais ou com 1 ou 2 indivíduos. Muitos destes métodos foram criados na década de 1980 e são relativamente bem utilizados nos dias atuais.

A extrapolação da curva da acumulação de espécies é composta por métodos bem intuitivos e precisam apenas da informação de presença/ausência das espécies coletadas em cada unidade amostral. Em conformidade com os estudos de Flather (1996); Soberón & Llorente (1993), esta categoria de métodos se baseia na modelagem de curvas de coletor, na tentativa de estimar a riqueza total prevendo o ponto de estabilização da curva, sendo que os modelos disponíveis para modelagem de curva podem ser divididos em não assintóticos e assintóticos, conforme Figura 1.

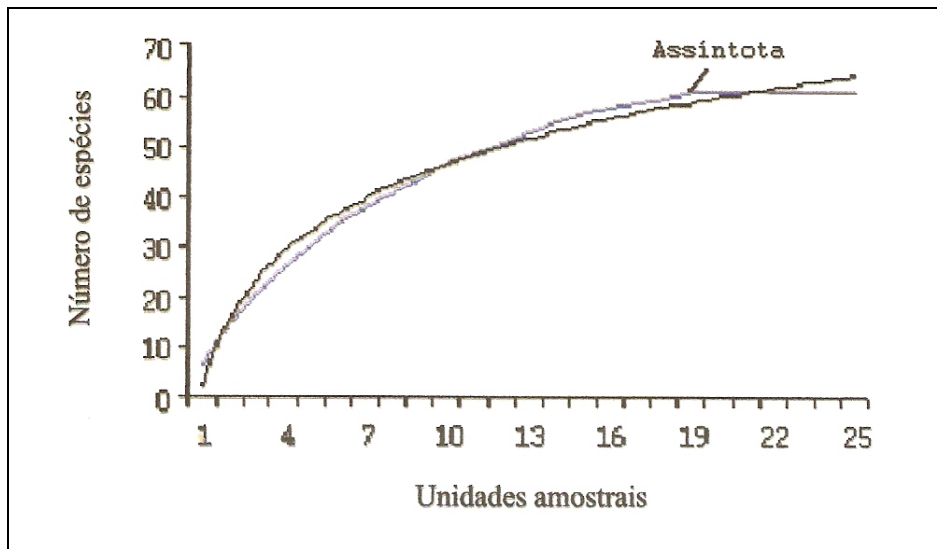


FIGURA 1 Curva do coletor

Os modelos não assintóticos não prevêm um ponto de estabilização para a curva, que cresceria infinitamente. Assim, esses modelos são úteis para prever o aumento do número de espécies diante de um aumento de esforço de coleta. No modelo assintótico, extrapola-se a curva ajustada até o ponto em que a taxa de acréscimo de espécies tenda a zero, ou seja, quando a curva estabiliza.

Teoricamente qualquer função matemática assintótica pode ser usada, embora na prática observa-se que algumas funções produzem em alguns casos, estimativas menores do que os valores de riqueza observados. Uma vantagem deste tipo de estimadores é que muitas das funções utilizadas possuem em suas fórmulas parâmetros que podem ser interpretados biologicamente, justificando a adoção de modelos não-lineares para o estudo do fenômeno estimação de espécies.

Apesar da grande diversidade de modelos de ajuste de curvas disponíveis, existe um que têm sido muito utilizados para estimativas de riqueza o modelo de Michaelis-Menten (Raaijmakers, 1987), sendo este modelo,

originalmente desenvolvido para modelar os mecanismos cinéticos enzimáticos *michaelianos* (Machado, 2006) e que permite estimar a assíntota a partir de qualquer ponto da curva, sendo um modelo facilmente ajustável a diferentes conjuntos de dados.

Mencionado por Coddington et al. (1996), e Colwell & Coddington (1994), mesmo que não seja possível estimar a riqueza total, a modelagem de curvas de acumulação de espécies pode ser muito útil para prever o esforço necessário para obter novas espécies em um inventário, permitindo inclusive discutir o esforço mínimo necessário para capturar uma determinada proporção das espécies da comunidade.

Um estudo de aplicação prática desse método é o trabalho de Flather (1996), citado por Santos (2004), em que o pesquisador observou os efeitos de modificações de habitat naturais sobre comunidades de aves, por meio de comparações de curvas de acumulação. Esse pesquisador testou vários modelos de ajuste e, através de análise de regressão entre taxas de incremento das curvas e índices de intensidade de uso antrópico para várias localidades no leste dos EUA, detectou reduções de riqueza em espécies em áreas muito modificadas.

### **2.1.3 Macro-invertebrados aquáticos: conhecendo a riqueza para estabelecimento de programas de monitoramento ambiental**

A destruição dos recursos naturais é uma preocupação atual crescente e que, apesar de inquietante, pouco tem sido feito para ser controlada. A água é um recurso natural imprescindível para a manutenção da vida no planeta. Entretanto, muitos dos nossos mananciais estão contaminados por lançamentos de esgotos domésticos e industriais não adequadamente tratados, contaminação por agrotóxicos, resíduos de atividades de mineração, entre outros (Nonato et al., 2007). Estas, e muitas outras formas de degradação ambiental têm impactado os

ecossistemas aquáticos o que tem acarretado em queda da qualidade da água e perda da diversidade aquática (Goulart & Callisto, 2003; Mazzini, 2007).

A busca por parâmetros que viabilizem a ação de gestores ambientais sobre estas áreas é ponto passivo para elaboração de estratégias que promovam a conservação/recuperação/restauração ambiental. Segundo Ribeiro & Uieda (2005), o uso de parâmetros biológicos para medir a qualidade da água se baseia nas respostas dos organismos em relação ao meio onde vivem, torna-os um promissor bioindicador da qualidade da água. Utilizar respostas biológicas como indicadores de degradação ambiental é mais vantajoso que usar medidas físicas e químicas da água, pois estas registram apenas o momento em que foram coletadas, como uma fotografia do rio (Metcalf, 1989). No entanto, os índices bióticos têm sido uma importante ferramenta em estudos de monitoramento de condições ecológicas, em geral considerando a composição taxonômica e dominância de alguns grupos tolerantes a poluição, (Callisto et al., 2002).

Bioindicadores são organismos ou comunidades que respondem à poluição ambiental, alterando suas funções vitais ou acumulando toxinas (Andréa, 2008). Estes organismos ou comunidades são potencialmente capazes de reagir a alterações ambientais visíveis em atributos de comunidade como declínios em abundância, diversidade, riqueza, densidade e com isso fornecem informações sobre a situação ambiental. Isto só é possível devido à existência de funções vitais que se correlacionam tão estreitamente com determinados fatores ambientais, podendo ser empregados como indicadores na avaliação de uma área (Galdean et al., 2000; Andréa, 2008).

Entre os organismos potencialmente conhecidos como bioindicadores está os macroinvertebrados bentônicos, os quais se constituem num importante grupo de organismos presentes em rios, riachos e lagoas, servindo de alimento para peixes e crustáceos e participando da reciclagem de nutrientes. Nos escritos de Esteves (1998), o seu uso como bioindicadores, é recomendado, pois refletem

as mudanças em variáveis do ambiente aquático, tais como o pH, oxigênio dissolvido e matéria orgânica.

A definição operacional para macroinvertebrado bentônico para este trabalho foi baseada em Callisto et al. (2001) e Maenpaa et al. (2003), os quais o definem como a fauna de invertebrados que fica retida em uma malha 0,5mm, sendo ela constituída por diversos táxons, tais como Arthropoda, Mollusca, Annelida, Nematoda e Platyhelminthes, dentre outros. A maioria dessas espécies está associada à habitats de fundo de corpos d'água e por isso é denominada de bentônica (do grego: *benthos* = fundo).

Nos apontamentos de Callisto & Gonçalves Junior (2005), os macroinvertebrados bentônicos têm sido amplamente utilizados como bioindicadores de qualidade de água e saúde de ecossistemas por apresentarem as seguintes características: (i) ciclos de vida longo, comparando-se com os organismos do plâncton que em geral tem ciclos de vida em torno de horas, dias, uma (1) ou duas (2) semanas; os macroinvertebrados bentônicos podem viver entre semanas, meses e mesmo mais de 1 ano, caracterizando-se como "organismos sentinelas"; (ii) em geral, são organismos grandes (maiores que 125 ou 250  $\mu\text{m}$ ), sésseis ou de pouca mobilidade, ou seja, são relativamente sedentários e mais fáceis de serem amostrados do que os organismos nectônicos, como os peixes; (iii) fácil amostragem, com custos relativamente baixos; (iv) elevada diversidade taxonômica e de identificação relativamente fácil (ao nível de família e alguns gêneros); (v) organismos sensíveis a diferentes concentrações de poluentes no meio, fornecendo ampla faixa de respostas frente a diferentes níveis de contaminação ambiental.

Segundo Goulart & Callisto (2003), os macroinvertebrados bentônicos podem ser classificados em organismos sensíveis ou intolerantes, organismos tolerantes ou facultativos e organismos resistentes, de acordo com sua tolerância frente às condições do ambiente. A predominância de determinado grupo de



organismos no meio pode oferecer indícios sobre as condições de qualidade da água, permitindo avaliar os efeitos da poluição sobre o corpo d'água (Moulton, 1998). Os estresses que alteram as características físico-químicas da água estão entre os que mais influenciam na composição da comunidade bentônica, podendo determinar a ocorrência e distribuição dos organismos no ambiente aquático.

Alguns parâmetros físico-químicos, quando atinge valores específicos podem afetar a sobrevivência de determinados organismos e favorecer a proliferação de outros, tornando-se fatores limitantes para a comunidade bentônica. Para Junqueira et al. (2000), apesar do emprego dos macroinvertebrados bentônicos ainda ser restrito no Brasil desde a década passada instituições de pesquisa de proteção ambiental têm desenvolvido estudos baseados no uso dessas comunidades para avaliar a qualidade das condições ambientais dos rios.

Para aplicação operacional de bioindicadores é necessário o conhecimento da comunidade em condições naturais na ausência de fatores impactantes para então compará-los com situações de comunidades bióticas em áreas impactadas. Além disso, ocorre que muitas áreas apresentam diferenças naturalmente impostas por arranjos ambientais. Na Amazônia, o rio Amazonas e seus afluentes representam o maior sistema hidrológico da Terra. Suas áreas de captação de água são bastante diversificadas, possuindo características distintas tanto em relação ao tipo de vegetação, quanto ao tipo de solo, relevo, processos erosivos e quantidade de chuvas.

Nos estudos de Junk & Silva (1995), os rios da bacia Amazônica podem ser classificados, de forma geral, em três tipos de águas: *brancas*, *claras e pretas*. As principais causas dessas variações seriam as alterações na quantidade de sedimentos suspensos e matéria orgânica dissolvida, decorrentes, principalmente, das características naturais da área de captação dos rios.

As águas *brancas* correspondem aos rios da Amazônia que apresentam maior turbidez, rios com essa característica são originados na cordilheira dos Andes, onde os processos erosivos são muito intensos devido à formação montanhosa recente. Portanto, Junk & Silva (1995), possuem alto teor de sedimentos em suspensão e alta concentração de sais minerais, o que causa a aparência mais turbida das águas. Apresentam pH médio variando entre 6,5 e 7,2. A concentração de matéria orgânica dissolvida é pequena, predominando sedimentos argilosos como ilita e clorita (Guyot, 1993).

Por sua vez, Junk & Silva (1995), as águas *claras* possuem uma aparência mais transparente devido à menor quantidade de sedimentos em suspensão e menos matéria orgânica dissolvida. Os solos são argilosos (caulinita), podendo ser classificados como lato solos com baixos índices de conteúdos orgânicos (Leenheer, 1980) e apresentam capacidade de fixar substâncias húmicas nas partículas de argila, retendo a matéria orgânica e causando uma aparência mais clara às águas.

Os rios de águas *pretas* são originados em solos podzólicos, que são solos mais arenosos. A cor da água deve-se à dissolução de ácidos húmicos, pois esse tipo de solo não retém matéria orgânica. As águas escuras ocorrem como consequência da decomposição da matéria orgânica e de seus processos edáficos. A intensidade da cor da água varia de acordo com as áreas de captação dos rios, com a quantidade de matéria orgânica dissolvida e com a estação do ano, sendo os rios mais escuros no começo da época chuvosa devido ao acúmulo de substâncias orgânicas nas épocas secas. A densa cobertura de floresta faz com que os processos erosivos sejam menores, havendo uma baixa quantidade de sedimentos em suspensão.

Nos escritos de Karr & Dudley (1981), os rios (ou sistemas lóticos) podem ser divididos em três classes de tamanho: as cabeceiras (rios de 1ª e 3ª ordens), rios de trechos médios (4ª a 6ª ordens) e grandes rios (7ª ordem ou

superior). Os riachos estariam inseridos no grupo de rios de 1ª e 3ª ordens. Dependendo do rio, há trechos de 2ª e 3ª ordens que corresponde à região com áreas de declive acentuado.

No Acre especificamente no oeste Amazônico, compreende uma das regiões mais biodiversas do planeta e abriga provavelmente muita das espécies ainda desconhecidas cientificamente, sendo considerada uma área prioritária para o levantamento de biodiversidade (Brasil, 2004). Estas áreas são marcadas por uma ampla diversidade de habitats, particularmente em relação aos córregos, denominados localmente de igarapés. Em virtude dessa particularidade, o Estado do Acre é área prioritária para conhecimento de riqueza de espécies, especialmente em virtude das políticas públicas de preservação ambiental em detrimento do avanço de problemas de ordem global em debate na conferência de Meio Ambiente em Kopenhague no ano de 2009.

#### **2.1.4 Parque Nacional da Serra do Divisor (PNSD)**

Segundo Brasil (2002), o Parque Nacional da Serra do divisor (PNSD) é uma Unidade de Conservação de Proteção Integral, com objetivo de preservar os ecossistemas, possibilitando a realização de pesquisas científicas, além de desenvolver atividades de educação e interpretação ambiental, recreação e turismo ecológico.

Localizado no extremo oeste do Estado do Acre, na bacia do Alto Juruá, o PNSD faz fronteira com o Peru (Norte: 07°07'00"S, 73°48'20"W; Sul: 09°24'40"S, 73°12'40"W; Leste: 09°08'40"S, 72°40'00"W; Oeste: 07°32'40"S, 73°59'20"W). Foi criado em 16 de junho de 1989, pelo Decreto Federal número 97.839, possui uma área de 843.012 ha e ocupa 5,5% da área total do Estado do Acre. Ocupa áreas dos municípios de Cruzeiro do Sul (25,7%), Mâncio Lima (57,2%), Rodrigues Alves (34,1%), Porto Walter (37,2%) e Marechal Thaumaturgo (5,1%). Esses municípios formam a microrregião de Cruzeiro do

Sul, segunda em número de habitantes, equivalendo em 2000 a 18,35% da população total do Estado.

O PNSD é banhado pela bacia do rio Juruá, que funciona como a mais importante via de transporte da região, o PNSD é constituído de quatro blocos de relevos distintos denominados: (i) serra da Jaquirana; (ii) serra do Moa; (iii) serra do Juruá-Mirin e, (iv) serra do Rio Branco. Os quatro blocos são originários de processo erosivos da formação Solimões, que resultaram num planalto e numa vista aérea de depressão. Sua área apresenta estrutura montanhosa pertencente à formação Solimões, por ser uma região topograficamente acidentada apresenta complexa rede hidrográfica com igarapés de água branca, clara e escura, originadas em ambientes que são geomorfologicamente diferentes (Sioli, 1991).

Relatos dos estudos de Silveira & Daly (1997), o PNSD possui extensão territorial de 846.633 hectares coberta por uma vegetação florestal primária, com as taxas de biodiversidades mais elevada da Amazônia, destacando-se por ser uma das regiões mais ricas em espécies de palmeiras do mundo. Nos registros de Souza (2003), as características geológicas e morfológicas atípicas dessa região apontam uma alta biodiversidade em relação ao restante da Amazônia.

Segundo os estudos de Dinerstein et al. (1995), a vegetação do PNSD é caracterizada pela ampla diversidade arbórea, intimamente relacionada à transição geográfica entre a Cordilheira dos Andes e as terras baixas do sudoeste Amazônico. A flora do PNSD pode ser dividida em duas principais tipologias regionais: Floresta Ombrófila Densa (22,5%) e Floresta Ombrófila Aberta (77,2%), a maior parte da área é coberta de floresta amazônica aberta, com grande incidência de palmeira, cipó e bambus.

Conforme Souza (2003), a fauna do PNSD é diferenciada em relação ao restante do país, citando a fauna de anfíbios do PNSD é a mais rica tanto em relação ao restante do país quanto às demais áreas na Amazônia. Para Calouro

(1999), a fauna de mamíferos é descrita como uma das mais ricas em relação ao restante do mundo, sendo a quarta Unidade de Conservação na Amazônia com maior número de espécies de primatas, sendo encontradas quarenta e três espécies de mamíferos terrestres de grande e médio porte, duas aquáticas e duas espécies de pequenos mamíferos de hábitos terrestres; comportando segundo descreve Lopes & Rehg (2003), a maior riqueza de primatas do estado do Acre.

De acordo com Lima (2008), referindo ao estudo realizado na área norte do PNSD com a finalidade de responder se existe correspondência entre a comunidade de macroinvertebrados baseado sobre taxonomia e atributos bioecológicos, e a classificação tipológica de riachos de águas brancas e claras em planícies do oeste Amazônico. Os resultados do estudo indicam clara diferença na estrutura de comunidade e organização funcional entre águas claras e brancas.

## 2.2 Regressão linear e não-linear

Segundo Maroco (2007), o modelo clássico de regressão teve origem nos trabalhos de astronomia, elaborado por Gauss no período de 1809 a 1821. É a técnica mais adequada para modelar relações entre variáveis e prever o valor de uma variável dependente (resposta) a partir de um conjunto de variáveis conhecidas como preditoras ou regressoras. As relações entre duas ou mais variáveis podem ser de dependência funcional, em que a magnitude da variável dependente é função da magnitude das variáveis independentes, ou de mera associação, em que nenhuma das variáveis pode ser tida como dependente da outra, ou seja, apenas variam em conjunto.

Segundo Draper & Smith (1998), um modelo de regressão é dito linear se as derivadas parciais do modelo em relação a cada um dos parâmetros que o constituem não forem funções dos próprios parâmetros. Seja o modelo de regressão linear para uma variável regressora, dado por:

$$Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

em que:  $n$  é o número de observações;  $Y_i$  é a variável aleatória (dependente);  $X_i$  é a variável regressora;  $\theta_0$  e  $\theta_1$  são os parâmetros do modelo;  $\varepsilon_i$  é o erro do modelo, considerado aleatório, independente e com distribuição normal de média zero e variância constante.

Verifica-se que para o modelo apresentado as derivadas parciais são livres dos parâmetros e, portanto é considerado um modelo de regressão linear, ou seja,

$$\frac{\partial Y_i}{\partial \theta_0} = 1 \text{ e } \frac{\partial Y_i}{\partial \theta_1} = x_i$$

De acordo com Cordeiro & Lima Neto (2004), até o início da década de 70, as principais técnicas desenvolvidas para os modelos de regressão não-lineares se restringiam à suposição de normalidade para a variável resposta. Nelder & Wedderburn (1972) ampliaram a distribuição da variável resposta para a família exponencial de distribuições, definindo os modelos lineares generalizados, entretanto, os modelos normais não-lineares continuaram recebendo um tratamento especial, surgindo diversos trabalhos nas décadas de 70 e 80, destacando-se Ratkowsky (1983) e Bates & Watts (1988).

Segundo Seber & Wild (1989) e Bates & Watts (1988), a principal característica dos modelos não-lineares é que eles são deduzidos a partir de suposições teóricas inerentes ao fenômeno que se tem interesse em modelar, ou seja, modelos formulados dessa forma são chamados de modelos científicos ou mecanísticos, e os parâmetros resultantes são interpretáveis. Assim, aproximá-los pelos modelos normais lineares, mesmo que sejam alcançados ajustes satisfatórios, prejudicaria bastante a obtenção de estimativas mais realistas dos parâmetros de interesse.

Para Draper & Smith (1998), por definição, um modelo de regressão é não-linear se pelo menos uma das derivadas parciais da variável dependente em relação a cada parâmetro for função dos próprios parâmetros. Um modelo de regressão não-linear é considerado “intrinsecamente linear”, se este puder ser transformado em um modelo linear por meio de uma reparametrização apropriada. Pode-se, ainda, usar o termo “intrinsecamente linear” para referir-se a modelos que podem ser linearizados via alguma transformação.

Em geral, na prática, um modelo não-linear é linearizado para facilitar a obtenção das estimativas dos parâmetros. O inconveniente de uma transformação é que, além do parâmetro perder sua interpretação intrínseca, pode-se alterar a estrutura e distribuição do erro, ou seja, se os erros do modelo original satisfizerem as suposições usuais de normalidade, independência e homogeneidade da variância, os erros do novo modelo em geral, não satisfarão tais suposições (Khuri & Cornell, 1987). Caso não seja possível obter uma reparametrização ou uma transformação apropriada, que reduza o modelo à forma linear tem-se os chamados modelos “intrinsecamente não-lineares”.

Uma das técnicas para resolver o problema dos modelos não-lineares é linearizar o modelo expandindo a função resposta,  $f(X, \theta)$ , pela série de Taylor em torno do ponto  $\theta_0$ , dado por:

$$f(X, \theta) \approx f(X, \theta_0) + \mathbf{F}(\theta - \theta_0),$$

em que  $\mathbf{F}$  é a matriz de variância da derivada parcial.

Segundo Bates & Watts (1988), o princípio dos mínimos quadrados é utilizado para estimar os parâmetros do modelo não-lineares, da mesma forma que em modelos lineares e a estimativa para  $\theta$  é dada pelo vetor  $\hat{\theta}$  que minimiza a soma de quadrados do resíduo.

Nos estudos de Bates & Watts (1988) e Ratkowsky (1990) em modelos multiparamétricos, as soluções das equações normais podem ser extremamente difíceis de serem obtidas e algum método iterativo de resolução de equações normais não-lineares deve ser utilizado na maioria dos casos.

Em modelos de regressão linear, os estimadores de mínimos quadrados são não viesados, normalmente distribuídos e, ainda, possuem variância mínima entre qualquer outra classe de estimadores. Essas propriedades são aceitas como as melhores propriedades que uma classe de estimadores pode apresentar (Searle, 1971). No entanto, podem existir outros estimadores, também não viesados, mas esses são menos eficientes no sentido de suas variâncias excederem a variância dos estimadores de mínimos quadrados. Já no caso não-linear, essas propriedades somente são válidas assintoticamente, isto é, quando o tamanho da amostra é suficientemente grande.

Segundo Ratkowsky (1983), o tamanho da amostra necessário para os resultados assintóticos terem validade depende fundamentalmente do modelo em estudo. Podem existir modelos que mesmo com amostras consideradas muito grandes, apresentam os resultados assintóticos inválidos. Entretanto, Souza (1998) afirma que, à medida que o tamanho da amostra aumenta, os resultados assintóticos vão se tornando mais aplicáveis.

De acordo com Guttman & Meeter, (1965) e Seber & Wild, (1989), quando os estimadores de mínimos quadrados apresentarem um pequeno viés, distribuição próxima da normal e verdadeira variância próxima daquelas dadas pela matriz de variâncias-covariâncias assintóticas, podem-se afirmar que os estimadores de mínimos quadrados exibem um comportamento próximo do comportamento linear. Quanto mais próximo do linear for o comportamento de um modelo, mais precisos serão os resultados assintóticos e, conseqüentemente, mais confiáveis serão as inferências. A adequabilidade à aproximação linear e



seus efeitos nas inferências são avaliados por meio das medidas de não-linearidade (Souza, 2008).

### **2.3 Inferência Bayesiana**

A teoria bayesiana foi desenvolvida por Thomas Bayes em meados do século XVIII, o qual propôs uma teoria subjetiva de probabilidade, baseada, principalmente, em um conhecimento *a priori* em relação às incertezas envolvidas no estudo. No século XX, mais precisamente na década de 30, a análise bayesiana ressurgiu com estudos teóricos, como o de Jeffreys (1939), que reagindo contra a predominante posição frequentista conseguiu ressuscitar o bayesianismo, apresentando-lhe aplicações lógicas para a resolução de problemas estatísticos em diversas áreas da ciência.

No entanto, estes estudos exigiam resoluções de integrais complexas, o que fez com que os métodos bayesianos se sujeitassem a mais algumas décadas de adormecimento. Foi somente na década de 90 que o problema dos cálculos integrais foi solucionado de maneira alternativa com o trabalho de Gelfand & Smith (1990), que exploraram um recurso de simulação dinâmica, o algoritmo Gibbs Sampler, elaborado por Geman & Geman (1984), o qual faz uso da teoria das Cadeias de Markov.

A metodologia bayesiana baseia-se num mecanismo formal (Teorema de Bayes) que permite a introdução da informação anterior ou externa à experiência em questão. Essa informação, representada por uma distribuição *a priori*, conjugada com a informação consubstanciada nos dados resultantes dessa experiência (função de verossimilhança), resulta em uma descrição completa do grau de conhecimento disponível sobre a experiência no momento da análise dada pela distribuição *a posteriori*, elemento central da inferência estatística bayesiana.

Segundo Box & Tiao (1992), na abordagem bayesiana, por serem os parâmetros considerados como variáveis aleatórias e não como constantes desconhecidas, e em especial, nos modelos de regressão não-lineares, se a distribuição dos erros for supostamente conhecida, pode-se conduzir a análise sem a necessidade de envolvimento com teoria assintótica.

Em recentes estudos envolvendo modelos de regressão não-linear Silva et al. (2006), Perreira (2006) e Savian (2008), a metodologia bayesiana foi utilizada com sucesso, pois sua característica de considerar o parâmetro, ou vetor de parâmetros, como variáveis aleatórias faz com que a utilização desta metodologia requeira um número menor de observações, pois os conceitos probabilísticos envolvidos diminuem a dependência do ajuste do modelo em relação aos números de dados utilizados, uma vez que o conceito de graus de liberdade residual não é considerado e toda a informação com relação aos parâmetros é representada pela distribuição *a priori*.

O paradigma bayesiano está se tornando uma alternativa bastante atraente frente ao paradigma clássico, adquirindo respeitabilidade na comunidade científica por ter adquirido grande versatilidade na resolução de problemas estatísticos, o que vem atraindo a atenção de pesquisadores em diversas áreas de aplicação, pondo-se assim na fronteira do conhecimento.

### **2.3.1 Distribuições *a priori***

De acordo com Paulino et al. (2003) a distribuição *a priori* representa, probabilisticamente, o conhecimento que se tem sobre o parâmetro, ou vetor de parâmetros  $\theta$ , de interesse para um estado da natureza, antes que seja realizada a experiência. A questão é como obter e quantificar essa informação, de natureza subjetiva, de modo a transformá-la em uma distribuição que possa ser utilizada para prosseguir com a metodologia bayesiana.

Segundo Box & Tiao (1992), a partir do conhecimento que se tem sobre  $\theta$ , pode-se definir uma família paramétrica de densidades. Nesse caso, a distribuição *a priori* é representada, por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Estes parâmetros indexadores da família de distribuições *a priori* são chamados de hiperparâmetros para distingui-los dos parâmetros de interesse  $\theta$ .

Essa abordagem, em geral, facilita a análise e o caso mais importante é o de *prioris* conjugadas. A idéia é que as distribuições *a priori* e *a posteriori* pertençam à mesma classe de distribuições e assim a atualização do conhecimento que se tem sobre  $\theta$  envolve apenas uma mudança nos hiperparâmetros. Nesse caso, o aspecto seqüencial do método bayesiano pode ser explorado definindo-se apenas a regra de atualização dos hiperparâmetros, já que as distribuições permanecem as mesmas. Uma densidade *a priori* pode ser informativa ou não informativa. Esta última ocorre quando não se tem informação alguma a respeito dos parâmetros, tendo assim que se formular uma distribuição *a priori* tal que toda informação venha exclusivamente dos dados.

Box & Tiao (1992), consideram que, em algumas situações, em que não existam informações *a priori* palpáveis, ou em que o conhecimento *a priori* é pouco significativo relativamente à informação amostral, o pesquisador é conduzido a adotar distribuições *a priori* minimamente informativas, que não representam, necessariamente, o desconhecimento do pesquisador sobre o experimento. É importante ressaltar que ela é utilizada também como forma de viabilizar a inferência estatística bayesiana.

### **2.3.2 Função de Verossimilhança**

Segundo Basu & Ghosh (1988), a informação é a matéria-prima da estatística e está entre os conceitos menos consensuais e de mais difícil apreensão. Tendo-se a consciência das dificuldades limitar-se-á a apresentação

ao conceito de informação de Fisher. A função de verossimilhança de  $\theta$  é a função que associa a cada  $\theta$  o valor  $p(x|\theta)$ , denotar-se-á esta função por  $L(\theta|x)$  (do inglês *likelihood*)

$$L(.,x): \Theta \rightarrow \mathbb{R}^+$$

$$\theta \rightarrow L(\theta;x)$$

Segundo Bolstad (2004), a função de verossimilhança associa (para um valor fixo de  $x$ ) a probabilidade de ser observado  $x$  a cada valor de  $\theta$ . Assim, quanto maior o valor da função maior a chances atribuídas pelo particular valor de  $\theta$  considerado ao evento fixado. Portanto, ao fixar um valor de  $x$  e variar os valores de  $\theta$ , observa-se a plausibilidade de cada um dos valores de  $\theta$ . Nesse sentido, ela informa a ordem natural de preferência entre diversas possibilidades de  $\theta$ . Um conjunto de dados é mais consistente com um vetor  $\theta_1$  do que com outro  $\theta_2$  se a verossimilhança associada a  $\theta_1$  for maior do que a associada a  $\theta_2$ . A função de verossimilhança dá origem ao Princípio da Verossimilhança que postula que toda a informação contida na experiência está representada na função de verossimilhança. Esse princípio é um divisor de águas na inferência, colocando de um lado, a abordagem Bayesiana e da verossimilhança que o respeita e de outro a abordagem clássica que não o respeita, pois esta leva em conta todos os valores possíveis da experiência.

### 2.3.3 Teorema de Bayes

Segundo Simões (1981), o teorema de Bayes foi estabelecido pelo Reverendo Thomas Bayes (1702-1761) no artigo intitulado "An Essay Toward Solving a Problem in the Doctrine of Chances" em 1763 e republicado em Barnard (1958). Basicamente, o teorema de Bayes é empregado na solução do

seguinte problema: seja um evento aleatório  $E$  e sejam  $A_1, A_2, \dots, A_n$  eventos mutuamente exclusivos de um espaço amostral  $S$  tomado *a priori* como causas de  $E$ . Com a realização de um experimento de resultado  $E$ , qual a probabilidade de que tenha sido causado por  $A_i$ ?

Dando continuidade aos relatos de Simões (1981), Laplace em *Théorie Analytique des Probabilités* 1812 enunciou o teorema de Bayes explicitamente e o desenvolveu como se segue:

$$E = S \mid E = (A_1 \cup A_2 \cup \dots \cup A_n) \mid E$$

$$E = (A_1 \mid E) \cup (A_2 \mid E) \cup \dots \cup (A_n \mid E),$$

em que os  $(A_i \mid E)$  são mutuamente exclusivos conforme ilustrados na Figura 3.

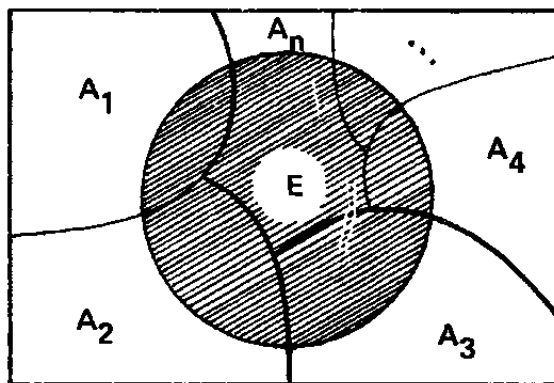


FIGURA 2 Eventos mutuamente exclusivos.

Sejam  $P(A_i)$  a probabilidade *a priori* de  $A_i$ , ou seja, desconhecendo a realização de  $E$ ;  $P(E|A_i)$  a probabilidade condicional de  $E$  quando produzido

pela causa  $A_i$ ;  $P(A_i | E)$  a probabilidade *a posteriori* de  $A_i$  desde que  $E$  aconteceu. A probabilidade do evento aleatório  $E$  ocorrer é dada por:

$$P(E) = P(A_1 | E) + P(A_2 | E) + \dots + P(A_n | E).$$

Uma vez que,

$$P(A_i | E) = \frac{P(A_i | E)}{P(E)} \quad \text{e} \quad P(A_i | E) = P(A_i) \times P(E | A_i),$$

tem-se:

$$P(E) = P(A_1) \times P(E | A_1) + P(A_2) \times P(E | A_2) + \dots + P(A_n) \times P(E | A_n),$$

logo,

$$P(A_i | E) = \frac{P(A_i) \times P(E | A_i)}{P(A_1) \times P(E | A_1) + P(A_2) \times P(E | A_2) + \dots + P(A_n) \times P(E | A_n)},$$

$$P(A_i | E) = \frac{P(A_i) \times P(E | A_i)}{\sum_{i=1}^n P(A_i) \times P(E | A_i)}.$$

O teorema de Bayes, de acordo com Simões (1981), é o mecanismo formal utilizado para tornar operacional e quantitativo o processo de modificar probabilidade, tendo em vista a incorporação de novas informações.

Segundo Bolstad (2004), as quantidades aleatórias envolvidas no estudo são do tipo discreto ou são do tipo contínuo e para o caso em que o parâmetro  $\theta$  é uma variável aleatória contínua o teorema de Bayes é dado por:

$$p(\theta | y) = \frac{L(\theta | y) p(\theta)}{\int L(\theta | y) p(\theta) d\theta}, \quad (1)$$

em que:  $p(\theta | y)$  é a distribuição *a posteriori* de  $\theta$ ,  $L(\theta | y)$  é a função de verossimilhança e  $p(\theta)$  é a distribuição *a priori* de  $\theta$ . Para o caso discreto o teorema de Bayes é enunciado por:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{L(\boldsymbol{\theta}|\mathbf{y})p(\boldsymbol{\theta})}{\sum L(\boldsymbol{\theta}|\mathbf{y})p(\boldsymbol{\theta})}. \quad (2)$$

Verifica-se que as funções, nos denominadores das expressões (1) e (2), não dependem do parâmetro  $\boldsymbol{\theta}$ , e são consideradas constantes normalizadoras. Em problemas que envolvam estimação de parâmetros, e por facilidade notacional, essas constantes podem ser removidas e o símbolo de igualdade é substituído por um de proporcionalidade ( $\propto$ ) resultando na expressão:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\boldsymbol{\theta}|\mathbf{y})p(\boldsymbol{\theta}),$$

**distribuição *a posteriori*  $\propto$  função de verossimilhança  $\times$  distribuição *a priori***

Conforme Paulino et al. (2003), o Teorema de Bayes ocupa lugar fulcral na inferência estatística bayesiana por ser um dos poucos resultados da matemática que se propõe caracterizar a aprendizagem com a experiência, isto é, a modificação da atitude inicial em relação aos antecedentes depois de ter a informação adicional de que certo acontecimento ou acontecimentos se realizaram.

### 2.3.4 Distribuições *a posteriori*

De acordo com Broemiling (1985), a distribuição *a posteriori* é a descrição mais completa do grau de conhecimento disponível sobre um parâmetro  $\boldsymbol{\theta}$  no momento da análise, será elemento central da inferência bayesiana, pois contém toda a informação probabilística a respeito deste parâmetro.

Segundo Paulino et al. (2003), no caso uniparamétrico a descrição abreviada da distribuição *a posteriori* é dada por:

- (i) um gráfico dessa função como meio de indicação qualitativa da sua forma geral é a melhor descrição do processo de inferência;
- (ii) resumos quantitativos de localização, dispersão e forma distribucional.

A escolha das estimativas a serem utilizadas depende naturalmente da forma de  $p(\boldsymbol{\theta}|\mathbf{y})$ , bem como dos objetivos do seu uso, destacar-se-á a moda *a posteriori* ( $\hat{\theta}$ ):

$$p(\hat{\theta}|\mathbf{y}) = \max_{\theta \in \Theta} p(\boldsymbol{\theta}|\mathbf{y}) = \max_{\theta \in \Theta} \{L(\boldsymbol{\theta}|x) p(\boldsymbol{\theta})\}, \quad (3)$$

essa informação restringe-se a um único valor, ou seja, toda informação presente na distribuição *a posteriori*,  $p(\boldsymbol{\theta}|\mathbf{y})$ , é condensada em uma estimacão pontual.

Rosa (1998) cita que para se inferir em relação a qualquer elemento de um vetor de parâmetros  $\mathbf{q}$ , a distribuição *a posteriori* conjunta dos parâmetros, ou somente distribuição *a posteriori*,  $p(\boldsymbol{\theta}|\mathbf{y})$ , deve ser integrada em relação a todos os outros elementos que a constituem. Assim, se o interesse do pesquisador se concentra sobre determinado conjunto de  $\mathbf{q}$ , por exemplo,  $\boldsymbol{\theta}_1$ , tem-se a necessidade da obtenção da distribuição  $p(\boldsymbol{\theta}_1|\mathbf{y})$ , denominada de distribuição marginal, a qual é dada por:

$$p(\boldsymbol{\theta}_1|\mathbf{y}) = \int_{\boldsymbol{\theta} \neq \boldsymbol{\theta}_1} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{\boldsymbol{\theta} \neq \boldsymbol{\theta}_1} \quad (4)$$



### 2.3.5 Intervalos de credibilidade e de máxima densidade *a posteriori* (HPD)

A forma mais adequada de expressar a informação que se tem sobre um parâmetro é através de sua distribuição *a posteriori*. A principal restrição da estimação pontual é que quando estimamos um parâmetro através de um único valor numérico toda a informação presente na distribuição *a posteriori* é resumida através deste número. É importante associar alguma informação sobre o quão precisa é a especificação deste número. O conceito de intervalo de credibilidade (ou intervalo de confiança Bayesiano) baseado na distribuição *a posteriori* é dado por:

Define-se  $C = [a, b]$  como um intervalo de credibilidade de  $100(1 - \alpha)\%$  para  $q$  por:

$$\int_a^b p(q | y) dq = 1 - \alpha, \quad 0 \leq a \leq b \leq 1$$

nota-se que a definição expressa, de forma probabilística, a pertinência ou não de  $q$  ao intervalo. Assim, quanto menor for o comprimento do intervalo, mais concentrada é a distribuição do parâmetro, ou seja, o tamanho do intervalo informa sobre a dispersão do parâmetro  $q$  (Ehlers, 2004).

É possível construir uma infinidade de intervalos utilizando a definição acima, mas é de principal interesse aquele com menor comprimento possível. Os intervalos de comprimento mínimo são obtidos tomando-se os valores de  $q$  com maior densidade *a posteriori*, denominados intervalos de máxima densidade *a posteriori* ou intervalos HPD. Segundo Ehlers (2007), um intervalo de credibilidade  $C$  de  $100(1 - \alpha)\%$  para  $q$  é de máxima densidade *a posteriori* se  $C = \{q \in Q : p(q | y) \geq k(\alpha)\}$ , em que  $k(\alpha)$  é a maior constante tal que  $P(q \in C) \geq 1 - \alpha$ .

Usando esta definição, todos os pontos dentro do intervalo de máxima densidade a *posteriori* (intervalo HPD) terão densidade maior do que qualquer ponto fora do intervalo.

## 2.4 Simulação Estocástica

Segundo Ehlers (2004) a Estatística como ciência empírica de tratamento e análise de dados teve um desenvolvimento sem precedentes nos últimos anos em sua capacidade de tratar problemas mais complexos. Isto se deveu principalmente à redescoberta de técnicas de simulação relativamente simples, mas extremamente poderosas, que puderam ser implementados graças ao avanço nas capacidades computacionais e a área que talvez tenha se beneficiado mais com este avanço foi a de inferência Bayesiana.

De acordo com Gamerman & Lopes (2006), a inferência bayesiana baseada em técnicas de simulação utiliza amostras da distribuição *a posteriori* para extrair informação a seu respeito. Obviamente, como uma amostra é sempre um substituto parcial da informação contida em uma densidade, método baseado em simulação são aproximados e devem apenas ser utilizados quando for constatada a impossibilidade de extração analítica de informação da distribuição *a posteriori*. Os métodos baseados em simulações não dependem de quantas observações foram tomadas e fornecem aproximações que serão tão melhores quanto maior for o número de valores gerados. Isso faz com que embora seja mais difícil controlar os erros cometidos sob o ponto de vista teórico, torne-se cada vez mais atraente do ponto de vista de aplicações práticas a utilização de métodos em simulações.

### 2.4.1 Método de Monte Carlo Simples

Os métodos de Monte Carlo têm sido utilizados como forma de obter aproximações numéricas de funções complexas, eles tipicamente envolvem a geração de observações de alguma distribuição de probabilidade e o uso da amostra obtida para aproximar a função de interesse. Neste texto será considerado somente o método de Monte Carlo Simples para o entendimento de conceitos posteriores.

As aplicações mais comuns dos métodos de Monte Carlo em computação numérica são para avaliar integrais. A idéia do método é justamente escrever a integral que se deseja calcular como um valor esperado. Para introduzir o método considere o problema de calcular a integral de uma função  $g(y)$  no intervalo  $(a, b)$ ,  $I = \int_a^b g(y) dy$ , esta integral pode ser reescrita como:

$$I = \int_a^b (b-a) g(y) \frac{1}{b-a} dy = (b-a) E[g(Y)]$$

identificando  $Y$  como uma variável aleatória com distribuição  $U(a, b)$ .

Assim, transformamos o problema de avaliar a integral no problema estatístico de estimar uma média,  $E[g(y)]$ . Se dispusermos de uma amostra aleatória de tamanho  $n$ ,  $y_1, \dots, y_n$  da distribuição uniforme no intervalo  $(a, b)$  teremos também uma amostra de valores  $g(y_1), \dots, g(y_n)$  da função  $g(y)$  e a integral acima pode ser estimada pela média amostral,

$$\hat{I} = (b-a) \frac{1}{n} \sum_{i=1}^n g(y_i).$$

não é difícil verificar que esta estimativa é não viesada, já que

$$E[g(Y_i)] = E[g(Y)]$$

para todo  $i$  e, portanto,

$$E(I) = \frac{b-a}{n} \sum_{i=1}^n E[g(Y_i)] = (b-a)E[g(Y)] \int_a^b g(y) dy.$$

#### 2.4.2 Métodos de Monte Carlo via cadeias de Markov (MCMC)

De acordo com Boldrini et al. (1980), muitos dos processos que ocorrem na natureza e na sociedade podem ser estudados (pelo menos em primeira aproximação) como se o fenômeno estudado passasse, a partir de um estado inicial, por uma seqüência de estados, onde a transição de um determinado estado para o seguinte ocorresse segundo uma certa probabilidade.

No caso em que esta probabilidade de transição dependa apenas do estado em que o fenômeno se encontra e do estado a seguir, o processo será chamado processo de Markov e uma seqüência de estados seguindo este processo será denominada uma cadeia de Markov. Evidentemente, ao se supor tal restrição estaremos simplificando, talvez demasiadamente, uma vez que as probabilidades podem se modificar com o tempo.

##### 2.4.2.1 Amostrador de Gibbs

Segundo Santos (2000), o amostrador de Gibbs é uma técnica de simulação de observações de uma distribuição marginal, sem que tenhamos que calcular analiticamente a densidade e sim utilizando as densidades condicionais completas. Para Gamerman & Hedibert (2006), o amostrador de Gibbs consiste numa classe de esquemas de simulação estocástica utilizando cadeias de Markov. O amostrador foi originalmente proposto dentro do contexto de reconstrução de imagens.

O algoritmo Gibbs é baseado no fato de que se a distribuição *a posteriori* conjunta  $p(\theta|y)$  for positiva em  $\Theta_1 \times \dots \times \Theta_k$ , com  $\Theta_i$  suporte da distribuição de  $\theta_i$  para  $i = 1, \dots, k$ , então ela é unicamente determinada pelas distribuições

condicionais completas  $p(\theta_i|y, \theta(-i))$ ,  $i = 1, \dots, k$ , em que  $\theta = (\theta_1, \dots, \theta_k)$  e  $\theta(-i) = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$ , ou seja, é o vetor  $\theta$  sem a  $i$ -ésima componente (Paulino et al., 2003).

O mesmo autor ainda descreve o algoritmo Gibbs da seguinte maneira. Seja  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$  um valor arbitrário inicial para o vetor  $\theta$ .

De modo iterativo procede-se da seguinte forma:

- 1) obtém-se  $\theta_1^{(1)}$  de  $p(\theta_1|y, \theta_2^{(0)}, \dots, \theta_k^{(0)})$ ,  
 obtém-se  $\theta_2^{(1)}$  de  $p(\theta_2|y, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$ ,  
 obtém-se  $\theta_3^{(1)}$  de  $p(\theta_3|y, \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$ ,  
 ...  
 obtém-se  $\theta_k^{(1)}$  de  $p(\theta_k|y, \theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \dots, \theta_{k-1}^{(1)})$ .

Completa-se assim uma iteração do esquema e uma transição de  $\theta^0$  para  $\theta^1 = (\theta_1^{(1)}, \dots, \theta_k^{(1)})$ .

- 2) O esquema anterior é repetido com  $\theta^1$ , obtido anteriormente, como vetor inicial para obter um novo vetor  $\theta^2$  e assim haver uma transição de  $\theta^1$  para  $\theta^2 = (\theta_1^{(2)}, \dots, \theta_k^{(2)})$ ,
- 3) Repete-se  $t$  vezes este ciclo de geração de observações aleatórias de cada uma das distribuições condicionais, produzindo assim  $\theta^0, \dots, \theta^t$ .

A sucessão  $\theta^0, \dots, \theta^t, \dots$  é uma realização de uma cadeia de Markov com espaço de estados  $\Theta$ . Resultados teóricos permitem concluir que, quanto  $t \rightarrow \infty$ ,  $(\theta_1^{(t)}, \dots, \theta_k^{(t)})$  tende em distribuição para um vetor aleatório cuja função densidade de probabilidade conjunta é  $p(\theta|y)$ . Em particular,  $\theta_i^{(t)}$  tende em distribuição para uma quantidade aleatória cuja densidade é  $p(\theta_i|y)$ , que é a densidade marginal *a posteriori* de  $\theta_i$  e  $\frac{1}{t} \sum_{i=1}^t g(\theta^i) \rightarrow E[g(\theta)|y]$  converge quase certamente, para qualquer função  $g(\cdot)$ , em que  $E[g(\theta)|y]$  representa o valor esperado de  $g(\theta)$  em relação à distribuição a posteriori  $p(\theta|y)$ .

De acordo com a teoria exposta, o método de amostragem Gibbs permite, através de um processo iterativo, gerar realizações de uma cadeia de Markov de tal modo que, à medida que o número de iterações aumenta, a cadeia se aproxima da sua condição de equilíbrio. Se num determinado instante  $t$  a cadeia já se encontra no estado de equilíbrio então o vetor  $\theta^t$  gerado nesse instante pode ser considerado como uma realização da distribuição  $p(\theta|y)$ . No entanto, sucessivas realizações de uma mesma cadeia ao longo do tempo, não constituem uma amostra aleatória da distribuição que se pretende gerar. Com efeito, os vetores  $\theta^t$  que vão sendo gerados são correlacionados.

A questão que se põe é como utilizar as realizações da cadeia, após o equilíbrio, para fazer inferências sobre  $\theta$ , ou qualquer outra função de  $\theta$ . Há essencialmente dois métodos em disputa. A teoria sugere que as inferências devem ser feitas com base nos resultados obtidos de uma única realização da cadeia durante um longo período. Por outro lado, gerações múltiplas de cadeias independentes também podem ser utilizadas para obter uma amostra

independente e identicamente distribuída de  $p(\theta|y)$ , utilizando as diferentes realizações, num determinado instante  $t$ , das várias cadeias geradas (Paulino et al., 2003).

#### 2.4.2.2 Algoritmo de Metropolis-Hasting

Conforme Nascimento (2009), o algoritmo de Metropolis-Hastings é, sem dúvida, o mais fundamental dos métodos MCMC, pois todos os outros são derivações dele. O algoritmo originalmente foi proposto por Metropolis et al. (1953) e generalizado por Hastings em 1970. O algoritmo usa a mesma idéia dos métodos de aceitação rejeição, isto é, um valor é gerado de uma distribuição auxiliar e aceito com uma dada probabilidade (Robert & Casella 1999).

Para Santos (2000), no amostrador de Gibbs, amostras aleatórias das distribuições conjuntas desejadas são simuladas por meio de densidades condicionais completas. Em Inferência Bayesiana, quando se faz necessário gerar uma amostra aleatória da distribuição *a posteriori*, as densidades condicionais completas são comumente disponíveis sob forma padronizada, principalmente quando a *priori* é conhecida. Entretanto, às vezes as densidades condicionais não apresentam formas conhecidas como Normais, Gamas, etc. Assim, fará o uso do algoritmo de Metropolis-Hastings para gerar tais amostras.

O algoritmo de Metropolis-Hastings usa a idéia de que um valor é gerado de uma distribuição auxiliar e aceito com uma dada probabilidade. Este mecanismo de correção garante a convergência da cadeia para a distribuição de equilíbrio, que neste caso é a distribuição *a posteriori* (Ehlers, 2007).

Suponha que a cadeia esteja no estado  $\theta$  e um valor  $\theta'$  é gerado de uma distribuição proposta  $q(\cdot|\theta)$ . Note que a distribuição proposta pode depender do estado atual da cadeia,  $\theta$ . O novo valor  $\theta'$  é aceito com probabilidade

$$\alpha(\theta, \theta') = \min \left( 1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right),$$

em que:  $\pi$  é a distribuição de interesse.

Em termos práticos, o algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos,

- 1) Inicialize o contador de iterações  $t=0$  e especifique um valor inicial  $\theta^{(0)}$ ;
- 2) Gere um novo valor  $\theta'$  da distribuição  $q(\cdot|\theta)$ ;
- 3) Calcule a probabilidade de aceitação  $\alpha(\theta, \theta')$  e gere  $u \sim U(0,1)$ ;
- 4) Se  $u < \alpha$  então aceite o novo valor,  $\theta'$ , e faça  $\theta^{(t+1)} = \theta'$ , caso contrário rejeite e faça  $\theta^{(t+1)} = \theta$ ;
- 5) Incremente o contador de  $t$  para  $t+1$  e volte ao passo 2.

Um caso particular é quando a distribuição proposta não depende do estado atual da cadeia, ou seja,  $q(\theta'|\theta) = q(\theta')$  para todos os valores de  $\theta$  e  $\theta'$ . Neste caso a probabilidade de aceitação se reduz para:

$$\alpha(\theta, \theta') = \min \left( 1, \frac{\pi(\theta')}{\pi(\theta)} \right).$$

### 2.4.3 Erro de Monte Carlo

No contexto de inferência bayesiana, o método de Monte Carlo é utilizado para o cálculo de integrais definidas, utilizando-se aproximações. De acordo com Bustos & Orgambide (1992), o estimador de Monte Carlo é dado por:



$$\hat{I} = \frac{1}{n} \sum_{i=1}^n g(Y_i),$$

em que:  $\hat{I}$  é uma aproximação para a integral que se deseja calcular, portanto, precisamos estudar o erro  $\hat{I} - I$ . Uma vez que as gerações são independentes, então, pela Lei Forte dos Grandes Números, segue que  $\hat{I}$  converge quase certamente para  $I$ , isto é,

$$\frac{1}{n} \sum_{i=1}^n g(Y_i) \xrightarrow{\text{q.t.}} E[g(Y)].$$

Além disso, definindo  $s^2 = \text{Var}[g(Y)]$  e assumindo que esta variância existe, o erro padrão de Monte Carlo é uma estimativa consistente de  $\sigma$ , dada por:

$$\hat{s} = \sqrt{\frac{1}{n} \sum_{i=1}^n (g(Y_i) - \bar{g})^2},$$

isto é, a aproximação pode ser tão acurada quanto se deseje, bastando aumentar o valor de  $n$  (número de iterações).

Este método consiste em gerar vários números aleatórios uniformemente distribuídos no intervalo de interesse, avaliando-se o valor da função a ser integrada em cada um destes pontos e, finalmente, calculando-se a média desses valores obtidos para se obter a aproximação da integral definida. Como o método de Monte Carlo é uma aproximação, não se pode afirmar que o valor encontrado por ele é o valor exato da integral. Entretanto, quanto mais pontos forem gerados, maior será a confiabilidade na aproximação obtida.

#### 2.4.4 Avaliação da convergência

Os algoritmos Gibbs Sampler e Metropolis-Hastings são iterativos, ou seja, necessitam da constatação de convergência, para que realmente se possa inferir sobre seus resultados como sendo valores das distribuições marginais dos parâmetros do modelo considerado. Portanto, a monitoração dessa convergência é importante, evitando um número excessivo ou insuficiente de iterações no processo de amostragem. Segundo Silva (2006), o número de iterações necessárias para que a convergência seja verificada nunca terá uma resposta definitiva, pois na prática a distribuição estacionária será desconhecida, mas pode-se sempre avaliar a convergência das cadeias, detectando problemas relacionados com a sua instabilidade.

Para Buzolin (2005), é comum descartar as primeiras iterações, porque esses valores iniciais influenciam a convergência, este período constituído pelas primeiras iterações designa-se por período de aquecimento (*burn-in*), outro aspecto importante relacionado com a avaliação da convergência diz respeito à dependência entre as observações subsequentes da cadeia. Para eliminar a autocorrelação entre as consecutivas iterações as observações devem ser espaçadas por um determinado número de iterações, ou seja, considerar saltos (*thin*) de tamanho  $k$ , usando para compor a amostra, os valores a cada  $k$  iterações.

De acordo com Nascimento (2009), existem duas grandes formas de abordar o estudo de convergência. A primeira é a mais teórica e visa o estabelecimento de cotas para funções de distribuições geradas em uma cadeia e distâncias entre elas, a segunda forma de estudar convergência é mais empírica e procura estudar as propriedades estatísticas da série observada.

Embora as duas formas sejam válidas e se complementem, os resultados obtidos através de estudos teóricos se mostram bem mais limitados e de difícil acesso que os baseados em um tratamento empírico (Cowles & Carlin, 1994).

#### 2.4.4.1 Monitoração formal da convergência

O critério proposto por Raftery & Lewis (1992) fornece a estimativa do número de iterações necessárias para se obter a convergência, bem como as estimativas do número de iterações que devem ser descartadas (*burn-in*) e a distância de uma iteração à outra (*thin*) para se obter uma amostra independente, ou seja, o método fornece o tamanho do descarte que deve ser dado para extrair o efeito dos valores iniciais e o salto de uma iteração para a outra, para retirar o efeito de dependência entre os valores amostrados.

A regra de decisão do critério de Raftery & Lewis (1992), é baseada no fator de dependência. O fator de dependência é responsável pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar a convergência, devido à autocorrelação dentro da seqüência. Quando o fator de dependência for maior que 5, pode-se concluir que a cadeia não atingiu convergência.

Gelman & Rubin (1992) sugerem como critério de convergência a utilização de várias cadeias em paralelo, começando de valores iniciais distintos. Para cada parâmetro de interesse, compara-se a variabilidade dentro e entre as cadeias amostradas, sendo cada um desses parâmetros denotados por  $\theta$ ; e uma vez que a estacionariedade tenha sido atingida, por exemplo, na  $i$ -ésima iteração, consideram-se as realizações  $\theta_j, \theta_{j+h}, \dots, \theta_{j+Nh}, j \geq i$  como uma amostra aleatória da distribuição desejada. A convergência é monitorada através do fator  $\hat{R}$ . Se o valor de  $\hat{R}$  for grande, será necessário considerar mais iterações para obter uma melhor estimativa dos parâmetros. Quando  $\hat{R} \approx 1$ , a convergência ocorre e a amostra selecionada das iterações é iid. A lentidão na convergência pode estar relacionada com a alta correlação entre as componentes de  $\theta$ .

O critério de Geweke (1992) é fundamentado em técnicas de análise espectral, e fornece um diagnóstico para a ausência de convergência. propõe uma avaliação de convergência baseada no teste de igualdade de médias da primeira e da última parte da cadeia, geralmente, dos primeiros 10% e dos últimos 50%. Quando a maioria dos dados estiver entre os limites de uma distribuição normal padrão e se a diferença padronizada entre as médias for muito grande, existe indicação de ausência de convergência. A regra de decisão do critério de Geweke é baseada na estimativa do p-valor. Se o p-valor for menor que um nível de significância adotado pelo pesquisador, conclui-se que a cadeia não atingiu convergência.

Heidelberger & Welch (1983) sugerem o uso de testes estatísticos para avaliar a hipótese nula de estacionariedade da amostra gerada. Se a hipótese nula for rejeitada para um dado valor, o teste é repetido após serem descartados os 10% dos valores iniciais da amostra. Caso a hipótese seja rejeitada novamente, descarta-se mais 10% dos valores iniciais, e assim sucessivamente, até serem descartados 50% dos valores iniciais. Se a hipótese nula for rejeitada novamente, isso indica que é necessário um número maior de iterações. Caso contrário, o número de iterações descartadas é indicado como o valor do *burn-in* (Nogueira et al., 2004).

#### **2.4.5 Autocorrelação**

Segundo Reis (2008), autocorrelações altas dentro das cadeias indicam cadeias lentas e, usualmente, a sua convergência também será lenta. Uma forma de fácil visualização é por meio do gráfico de autocorrelação, sendo possível tirar conclusões a respeito da convergência e também da autocorrelação da cadeia ocasionada pelo uso das cadeias de Markov. Com o uso da função de autocorrelação (fac), pode-se ter uma idéia do salto que se deve tomar na cadeia para que a mesma apresente uma forma estacionária.

## 2.5 Modelo não-linear de Michaelis-Menten

Segundo Machado (2006) descreveu, pouco mais de 90 anos a equação de Michaelis-Menten parece ter sua maior consolidação atuando como ponto de partida para tratamentos matemáticos que definem o comportamento cinético das reações catalisadas por enzimas, bem como para descrição teórica dos mecanismos enzimáticos.

O modelo teórico de Michaelis-Menten (1913) atravessou os limites da aplicação bioquímica e alcançou as áreas agronômicas a partir do estudo da dinâmica existente na relação entre a fertilidade do solo e produtividade de plantas, adequado à ecologia é um dos modelos mais utilizados para a estimação em espécies.

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i \quad (5)$$

$y_i$ : Quantidade acumulada de espécies

$x_i$ : Número de amostras

$\theta_1$ : Quantidade máxima de espécies

$\theta_2$ : O número de amostras na qual a quantidade acumulada de espécies é metade da quantidade máxima de espécies

$i = 1, \dots, n$

$\mathbf{y} = \{y_1, \dots, y_n\}$

$\varepsilon_i$  é o resíduo na  $i$ -ésima amostra considerado independente e igualmente distribuído, com distribuição  $N(0, \sigma_\varepsilon^2)$

### **3 MATERIAL E MÉTODO**

#### **3.1 Material**

##### **3.1.1 Dados analisados no estudo**

A análise dos dados utilizados neste estudo teve como suporte Lima (2008), em sua dissertação intitulada “Macroinvertebrados em igarapés de águas claras e brancas no oeste Amazônico (Acre-Brasil): implicações para elaboração de estratégias de biomonitoramento”, obtidos no Parque Nacional da Serra do Divisor (PNSD), no município de Mâncio Lima – Acre - Brasil (figura 3).

O período de realização da campanha amostral ocorreu no ano de 2006, no PNSD, o qual apresenta duas principais divisões: área norte e área sul. Foi escolhida a área norte do parque para realização das coletas em virtude do apelo à maior variação geográfica nesta região o que permite postular maior número de espécies do que em áreas menos acidentadas geograficamente. Isto está intimamente ligado a influência da declividade sobre a diversidade de espécies. As coletas foram realizadas em dez igarapés (IG) de primeira e segunda ordem em águas brancas e claras da bacia do rio Moa e azul (figura 4).

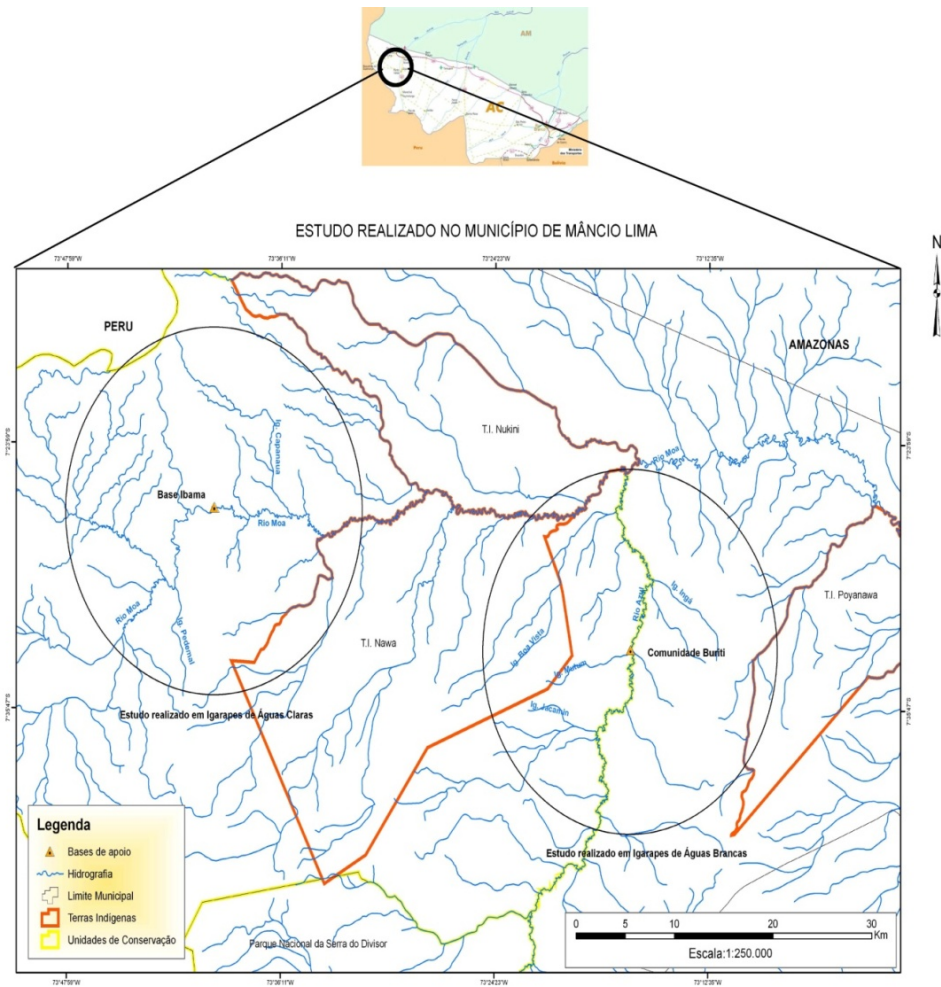


FIGURA 3 Mapa do PNSD, área norte, contendo as duas bacias de coleta. Bacia do rio Moa para coletas de águas claras e bacia do rio Azul para coleta de águas brancas



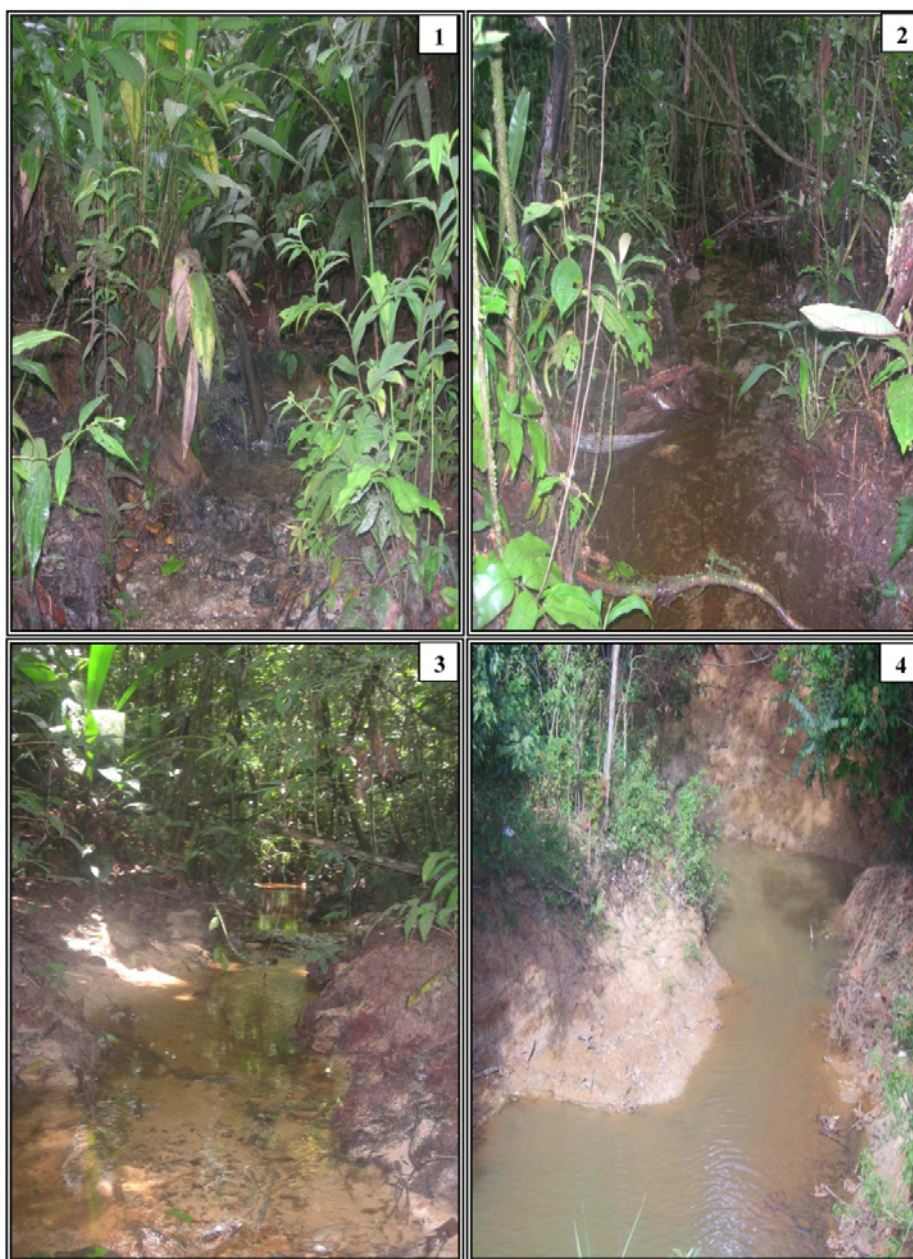


FIGURA 4 Imagem dos riachos (igarapés) de águas claras (1, 2 e 3) e águas brancas (4)

Os macroinvertebrados após as coletas foram identificados e os dados tabulados, conforme Tabela 1.

TABELA 1 Representação da estrutura dos dados coletados em dez igarapés (IG) de águas claras

<b>Táxons</b>	<b>IG1</b>	<b>IG2</b>	<b>IG3</b>	<b>IG4</b>	<b>IG5</b>	<b>IG6</b>	<b>IG7</b>	<b>IG8</b>	<b>IG9</b>	<b>IG10</b>
<i>Curculionidae gênero ind.</i>	0	0	0	0	0	1	0	0	0	0
<i>Dytiscidae gênero ind.</i>	0	0	0	0	0	0	1	0	0	0
<i>Neollmis</i>	0	2	6	0	1	11	31	48	33	13
...	.	.	.	.	.	.	.	.	.	.
...	.	.	.	.	.	.	.	.	.	.
...	.	.	.	.	.	.	.	.	.	.
<i>Polypetropus</i>	0	0	0	0	0	0	0	3	0	0
<i>Cyrnellus</i>	0	0	0	1	2	0	0	0	0	0
<i>Seracostomatidae Gênero I</i>	1	0	0	0	0	0	0	0	0	0
<i>Xiphocentron</i>	6	0	0	0	0	3	0	2	3	0

Fonte: Lima (2008)

Da amostra obtida foram detectados 111 táxons distintos independentes de sua abundância, e baseado em fundamentos ecológicos relacionados à ecologia de riachos, assumiram-se as seguintes premissas:

- (i) A riqueza de espécies em águas claras é superior em relação às águas brancas;

- (ii) O incremento de novas espécies a cada amostra para águas claras é significativamente maior do que em águas brancas;
- (iii) Águas claras apresentam maior número de táxons exclusivos da tipologia do que águas brancas;
- (iv) Águas claras apresentam ambientes naturalmente menos perturbados e por isso com maior incidência de espécies diferentes;
- (v) A magnitude dos tipos de microhabitat é mais evidente em águas claras e por isso promove maior número de espécies diferentes convivendo no local;
- (vi) Chironomidae é esperado ser o táxon de maior dominância em águas claras.

Aceitando-se as premissas tornar-se-á razoável admitir que a informação obtida da amostragem em águas claras será assumida como uma população teórica de referência bastante próxima da população real.

Foram realizadas três amostragens a partir da população teórica de referência e cada amostra foi composta de seis unidades amostrais representada pelos igarapés sorteados aleatoriamente entre os dez existentes, sendo que em cada igarapé (unidade amostral) a contagem utilizada foi o número de táxons distintos, ou seja, presença ou ausência independente da abundância deles no igarapé, implicando em três amostras distintas denotadas por amostra 1, amostra 2 e amostra 3 conforme Tabelas 2, 3 e 4.

TABELA 2 Representação da estrutura da amostra 1 coletada em seis igarapés (IG) de águas claras

<b>Taxons</b>	<b>IG10</b>	<b>IG8</b>	<b>IG1</b>	<b>IG3</b>	<b>IG5</b>	<b>IG2</b>
<i>Neollmis</i>	13	48	0	6	1	2
<i>Heterelmis</i>	0	0	0	0	0	2
<i>Prionocyphon</i>	0	0	0	0	0	1
<i>Haliplidae gênero ind.</i>	0	11	0	0	0	0
...	.	.	.	.	.	.
...	.	.	.	.	.	.
...	.	.	.	.	.	.
<i>Polypectropus</i>	0	3	0	0	0	0
<i>Cyrnellus</i>	0	0	0	0	2	0
<i>Seracostomatidae Gênero 1</i>	0	0	1	0	0	0
<i>Xiphocentron</i>	0	2	6	0	0	0

Fonte: Lima (2008)

TABELA 3 Representação da estrutura da amostra 2 coletada em seis igarapés (IG) de águas claras

<b>Taxons</b>	<b>IG2</b>	<b>IG8</b>	<b>IG1</b>	<b>IG7</b>	<b>IG3</b>	<b>IG9</b>
<i>Dytiscidae gênero ind.</i>	0	0	0	1	0	0
<i>Neollmis</i>	2	48	0	31	6	33
<i>Heterelmis</i>	2	0	0	0	0	0
<i>Prionocyphon</i>	1	0	0	0	0	0
...	.	.	.	.	.	.
...	.	.	.	.	.	.
...	.	.	.	.	.	.
<i>Polycentropodidae gênero ind.</i>	0	0	0	0	3	0
<i>Polypectropus</i>	0	3	0	0	0	0
<i>Seracostomatidae Gênero 1</i>	0	0	1	0	0	0
<i>Xiphocentron</i>	0	2	6	0	0	3

Fonte: Lima (2008)

TABELA 4 Representação da estrutura da amostra 3 coletada em seis igarapés (IG) de águas claras

<b>Taxons</b>	<b>IG3</b>	<b>IG5</b>	<b>IG9</b>	<b>IG6</b>	<b>IG4</b>	<b>IG7</b>
<i>Curculionidae gênero ind.</i>	0	0	0	1	0	0
<i>Dytiscidae gênero ind.</i>	0	0	0	0	0	1
<i>Neollmis</i>	6	1	33	11	0	31
<i>Cyphon</i>	1	0	0	0	0	1
...	.	.	.	.	.	.
...	.	.	.	.	.	.
...	.	.	.	.	.	.
<i>Chimarra</i>	0	3	0	0	0	0
<i>Polycentropodidae gênero ind.</i>	3	0	0	1	0	0
<i>Cyrnellus</i>	0	2	0	0	1	0
<i>Xiphocentron</i>	0	0	3	3	0	0

Fonte: Lima (2008)

A partir das amostras 1, 2 e 3 serão geradas três massas de dados denominadas *dados1*, *dados2* e *dados3* conforme procedimentos adotados e que serão utilizados para *dados1* descrito pelas seguintes etapas:

- 1) A partir da amostra 1 (Tabela 2), observa-se a primeira unidade amostral (IG 10) e conta-se o número de táxons presentes, sem considerar a sua abundância, obtendo-se o seguinte resultado:

unidade amostral 1  $\Rightarrow$  38 táxons;

- 2) Repete-se o procedimento adotado na etapa 1 para a segunda unidade amostral (IG 8), contando-se o número de táxons presente mais distintos da primeira unidade amostral, ou seja, sem repetição e obteve-se o seguinte resultado:

unidade amostral 2  $\Rightarrow$  14 táxons;

3) O procedimento utilizado na etapa 2 será o mesmo para as unidades amostrais 3, 4, 5 e 6 tendo-se como resultado:

unidade amostral 3  $\Rightarrow$  12 táxons

unidade amostral 4  $\Rightarrow$  13 táxons

unidade amostral 5  $\Rightarrow$  8 táxons

unidade amostral 6  $\Rightarrow$  8 táxons

4) Utilizando-se dos resultados obtidos nas etapas anteriores gera-se *dados1* que será composto do número de táxons acumulados por unidades amostrais (Igarapés), visualizados na Tabela 5.

TABELA 5 Táxons acumulados por unidades amostrais (IG) relacionados a *dados1*

<b>Unidades Amostrais</b>	<b>Táxons acumulados por unidades amostrais</b>
1	38
2	52
3	64
4	77
5	85
6	93

Por analogia as tabelas 6 e 7 estão relacionadas a *dados2* e *dados3*.

TABELA 6 Táxons acumulados por unidades amostrais (IG) relacionados a *dados2*

Unidades amostrais	Táxons acumulados por unidades amostrais
1	37
2	59
3	68
4	79
5	90
6	96

TABELA 7 Táxons acumulados por unidades amostrais (IG) relacionados a *dados3*

Unidades Amostrais	Táxons acumulados por unidades amostrais
1	43
2	61
3	70
4	80
5	85
6	94

### 3.2 Método

#### 3.2.1 Obtenção da distribuição *a posteriori* conjunta para os parâmetros do modelo de Michaelis-Menten relacionado a *dados1*

Seja a função densidade de probabilidades para os dados observacionais dada por:

$$f(y_i | \theta_1, \theta_2, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right],$$

a função de verossimilhança para os *dados1*, considerando-se uma amostra aleatória de tamanho  $n$  é:

$$L_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2\sigma^2} \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \right\}$$

$$L_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (6)$$

Uma densidade a priori para  $\theta = (\theta_1, \theta_2)$  e  $\sigma^2$ , assumindo  $\theta$  e  $\sigma^2$  independentes é dada por:

$$p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2) = p_{\text{dados1}}(\theta_1, \theta_2) p_{\text{dados1}}(\sigma^2)$$

Assumiu-se que o pesquisador não tinha informação prévia a respeito dos parâmetros, e para a obtenção de uma priori não informativa para  $\theta_1$  e  $\theta_2$  utilizou a regra multiparamétrica de Jeffreys (Apêndice A), dada por:

$$p_{\text{dados1}}(\theta_1, \theta_2) \propto |\theta_1^2 q(\theta_2)|^{\frac{1}{2}} \Rightarrow$$

$$p_{\text{dados1}}(\theta_1, \theta_2) \propto \theta_1^2 \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{\frac{1}{2}} \quad (7)$$

em que:



$$q(\theta_2) = \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}$$

para a variância residual considerou-se a distribuição *a priori* não informativa para  $\sigma^2$  dada por:

$$p(\sigma^2) \propto \sigma^{-1}$$

obtendo-se então a distribuição *a priori* conjunta para os parâmetros

$$0 < \theta_1, \theta_2 < \infty \text{ e } \sigma^2 > 0$$

dada por:

$$\begin{aligned} p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2) &\propto \sigma^{-1} |\theta_1^2 q(\theta_2)|^{\frac{1}{2}} \Rightarrow \\ p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2) &\propto \sigma^{-1} \theta_1 |q(\theta_2)|^{\frac{1}{2}} \end{aligned} \quad (8)$$

A distribuição *a posteriori* conjunta para os parâmetros  $\theta_1$ ,  $\theta_2$  e  $\sigma^2$  do modelo, considerando a distribuição *a priori* (8) e a função de verossimilhança (6) é dada por:

$$\begin{aligned} p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) &\propto p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2) L_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \\ &\propto \left\{ \sigma^{-1} \theta_1 |q(\theta_2)|^{1/2} \right\} \left\{ (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \right\} \end{aligned}$$

$$\begin{aligned}
& \propto \sigma^{-1} \theta_1 |q(\theta_2)|^{1/2} (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \\
& \propto (\sigma^2)^{-\frac{1}{2}} (\sigma^2)^{-\frac{n}{2}} \theta_1 \left| \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\} \right|^{1/2} \times \\
& \quad \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \\
& \propto (\sigma^2)^{-\frac{1}{2}} (\sigma^2)^{-\frac{n}{2}} (\sigma^2)^{-1} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\
& \quad \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]
\end{aligned}$$

efetuando o produto e substituindo

$$(\sigma^2)^{-\frac{1}{2}} (\sigma^2)^{-\frac{n}{2}} (\sigma^2)^{-1} = (\sigma^2)^{-\frac{n}{2}} (\sigma^2)^{-\frac{1}{2}} (\sigma^2)^{-1} = (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right]}$$

tem-se:

$$\begin{aligned}
p_{\text{dados}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) & \propto (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\
& \quad \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (9)
\end{aligned}$$

A partir de (9), ou seja, da distribuição *a posteriori* conjunta dos parâmetros obter-se-á as distribuições condicionais completas para os parâmetros  $\theta_1$ ,  $\theta_2$  e  $\sigma^2$  do modelo.

Para  $\theta_1$  tem-se:

$$p_{\text{dados1}}(\theta_1 | \mathbf{y}, \theta_2, \sigma^2) \propto \theta_1 \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (10)$$

para  $\theta_2$  tem-se:

$$p_{\text{dados1}}(\theta_2 | \mathbf{y}, \sigma^2, \theta_1) \propto \left\{ \left[ \sum_{i=1}^n \frac{x_i}{\theta_2 + x_i} \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\ \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (11)$$

para  $\sigma^2$  tem-se:

$$p_{\text{dados1}}(\sigma^2 | \mathbf{y}, \theta_1, \theta_2) \propto (\sigma^2)^{-\left[ \frac{(n+1)}{2} + 1 \right]} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (12)$$

que é o núcleo de uma gama inversa cuja a função de densidade é da forma

$$p(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\beta}{x}}, \quad x > 0 \quad \mathbf{X} : GI(\alpha, \beta)$$

de parâmetros dados por:

$$\alpha = \frac{(n+1)}{2}$$

$$\beta = \frac{1}{2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2$$

### 3.2.2 Obtenção da distribuição *a posteriori* conjunta para os parâmetros do modelo de Michaelis-Menten relacionado a *dados2*

A distribuição *a priori* conjunta a ser utilizada será a distribuição *a posteriori* conjunta (9) que foi obtida em *dados1* dada por:

$$P_{\text{dados2}}(\theta_1, \theta_2, \sigma^2) \propto (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (13)$$

e a função de verossimilhança utilizada eq. (6), considerando-se que o modelo e o número de unidades amostrais são os mesmos é:

$$L_{\text{dados2}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (14)$$

obtendo-se a distribuição *a posteriori* por meio da combinação de (13) e (14) dada por:

$$p_{\text{dados2}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = p_{\text{dados2}}(\theta_1, \theta_2, \sigma^2) L_{\text{dados2}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y})$$

implicando em,

$$p_{\text{dados2}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\ \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (15)$$

simplificando (15) tem-se:

$$\propto (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right]} (\sigma^2)^{-\frac{n}{2}} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\ \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \\ \propto (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right] - \frac{n}{2}} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\ \times \exp \left[ (2) \times \left( -\frac{1}{2\sigma^2} \right) \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]$$

$$\begin{aligned}
p_{\text{dados2}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left[\frac{(2n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \frac{x_i}{\theta_2 + x_i} \right]^2 \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\
\times \exp \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (16)
\end{aligned}$$

A partir de (16), ou seja, da distribuição *a posteriori* conjunta dos parâmetros obter-se-á as distribuições condicionais completas para os parâmetros  $\theta_1$ ,  $\theta_2$  e  $\sigma^2$  do modelo:

Para  $\theta_1$  tem-se:

$$p_{\text{dados2}}(\theta_1 | \mathbf{y}, \theta_2, \sigma^2) \propto \theta_1 \exp \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (17)$$

para  $\theta_2$  tem-se:

$$\begin{aligned}
p_{\text{dados2}}(\theta_2 | \mathbf{y}, \theta_1, \sigma^2) \propto \left\{ \left[ \sum_{i=1}^n \frac{x_i}{\theta_2 + x_i} \right]^2 \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \\
\times \exp \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (18)
\end{aligned}$$

Para  $\sigma^2$  tem-se:

$$p_{\text{dados2}}(\sigma^2 | \mathbf{y}, \theta_1, \theta_2) \propto (\sigma^2)^{-\left[\frac{(2n+1)}{2} + 1\right]} \exp \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (19)$$

que é o núcleo de uma gama inversa:

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\beta}{x}}, \quad x > 0 \quad X : GI(\alpha, \beta)$$

de parâmetros dados por:

$$\alpha = \frac{(2n+1)}{2}$$

$$\beta = \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2$$

### 3.2.3 Obtenção da distribuição *a posteriori* conjunta para os parâmetros do modelo de Michaelis-Menten relacionado a *dados3*

A distribuição *a priori* conjunta a ser utilizada será a distribuição *a posteriori* conjunta (16) que foi obtida em *dados2* dada por:

$$p_{\text{dados3}}(\theta_1, \theta_2, \sigma^2) \propto (\sigma^2)^{-\left[\frac{(2n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times$$

$$\times \exp \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (20)$$

e a função de verossimilhança que será utilizada é idêntica a utilizada em *dados2* eq (14) considerando-se o mesmo modelo, tamanho amostral e a massa de *dados3*,

$$L_{\text{dados3}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right] \quad (21)$$

Obtendo-se a distribuição *a posteriori* por meio da combinação de (20) e (21) dada por:

$$p_{\text{dados3}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = p_{\text{dados3}}(\theta_1, \theta_2, \sigma^2) L_{\text{dados3}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y})$$

implicando em,

$$p_{\text{dados3}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{\left[\frac{(2n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\ \times \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right] (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right] \quad (22)$$

simplificando tem-se:

$$\propto (\sigma^2)^{\left[\frac{(2n+1)}{2} + 1\right]} (\sigma^2)^{-\frac{n}{2}} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\ \times \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right] \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right]$$



$$\begin{aligned}
p_{\text{dados3}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left[\frac{(3n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\
\times \exp \left[ -\frac{3}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (23)
\end{aligned}$$

A partir de (23), ou seja, da distribuição *a posteriori* conjunta dos parâmetros obter-se-á as distribuições condicionais completas para os parâmetros  $\theta_1$ ,  $\theta_2$  e  $\sigma^2$  do modelo:

Para  $\theta_1$  tem-se:

$$p_{\text{dados3}}(\theta_1 | \mathbf{y}, \theta_2, \sigma^2) \propto \theta_1 \exp \left[ -\frac{3}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (24)$$

Para  $\theta_2$  tem-se:

$$\begin{aligned}
p_{\text{dados3}}(\theta_2 | \mathbf{y}, \theta_1, \sigma^2) \propto \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \\
\times \exp \left[ -\frac{3}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (25)
\end{aligned}$$

Para  $\sigma^2$  tem-se:

$$p_{\text{dados3}}(\sigma^2 | \mathbf{y}, \theta_1, \theta_2) \propto (\sigma^2)^{-\left[\frac{(3n+1)}{2} + 1\right]} \exp \left[ -\frac{3}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \quad (26)$$

que é o núcleo de uma gama inversa:

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\beta}{x}}, \quad x > 0 \quad X: GI(\alpha, \beta)$$

de parâmetros dados por:

$$\alpha = \frac{(3n+1)}{2}$$

$$\beta = \frac{3}{2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2$$

### 3.3 Implementação das análises

Utilizar-se-á o software R (R Development Core Team, 2010) versão 2.10.0 construindo-se rotinas (algoritmos), ou seja, o Amostrador de Gibbs e o Algoritmo de Metropolis Hastings, para estimar os parâmetros do modelo via MCMC do modelo de Michaelis-Menten (apêndice B), a convergência dessas cadeias foi verificada de acordo com os seguintes critérios (Raftery & Lewis (1992), Geweke (1992), Gelman & Rubin (1992) utilizando o pacote do R denominado BOA (Bayesian Output Analysis) e para a análise dos resultados foram implementadas rotinas específicas (Apêndice C).

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Ajuste do modelo de Michaelis-Menten aos *dados1*, *dados2* e *dados3*

Para  $\sigma^2$  por meio dos núcleos das densidades *a posteriori* condicionais completas foram identificadas distribuições gamas inversas respectivamente relacionadas à *dados1*, *dados2* e *dados3*. No entanto, devido à dificuldade em reconhecer o núcleo das *posteriors* condicionais completas para  $\theta_1$  e  $\theta_2$  buscou-se distribuições candidatas normais que pudessem ser utilizadas na geração de cadeias de valores por meio do MCMC, obtendo-se amostras dos parâmetros.

As amostras foram compostas de 5 000 valores, tamanho padronizado para todas as massas de dados. Para descartar os valores iniciais foi utilizado um *Burn-in* de 1 000 e o *Thin* utilizado foi o obtido pela função de autocorrelação (fac), para cada massa de dados da seguinte forma:

- a) Para  $\theta_1$  e  $\theta_2$  (*dados1*) *Thin* = 400
- b) Para  $\theta_1$  e  $\theta_2$  (*dados2*) *Thin* = 200
- c) Para  $\theta_1$  e  $\theta_2$  (*dados3*) *Thin* = 250

O tamanho das cadeias para os parâmetros  $\theta_1$  e  $\theta_2$  é dado genericamente por:

$$\text{Número Iterações} = \text{Tamanho amostra} \times \text{Thin} + \text{Burn-in}$$

A convergência das cadeias de todos os parâmetros do modelo foi monitorada por meio da visualização gráfica do traço e dos critérios disponíveis no pacote BOA do software livre R (R Development Core Team, 2010), de acordo com Figuras 5, 6 e 7 e Tabelas 8, 9 e 10.

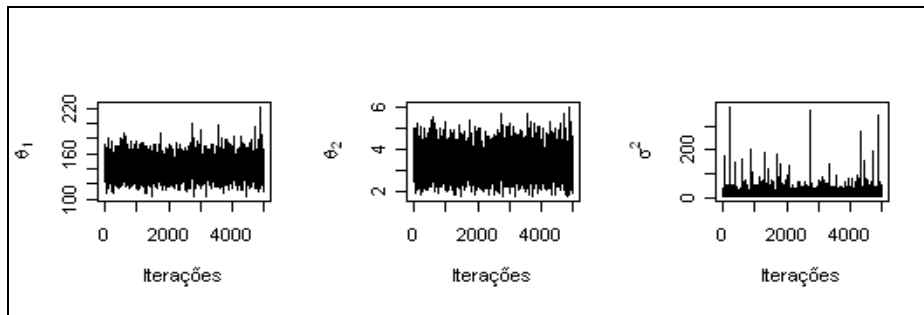


FIGURA 5 Gráfico de traços dos parâmetros da cadeia *dados1*

TABELA 8 Critérios de convergência relacionados a *dados1*

Parâmetros	Raftery & Lewis FD	Geweke p-valor	Gelman & Rubin $\hat{R}$
$\theta_1$	2,2216	0,5618	1,0330
$\theta_2$	2,1949	0,4866	1,0339
$\sigma^2$	1,0320	0,6761	1,0596

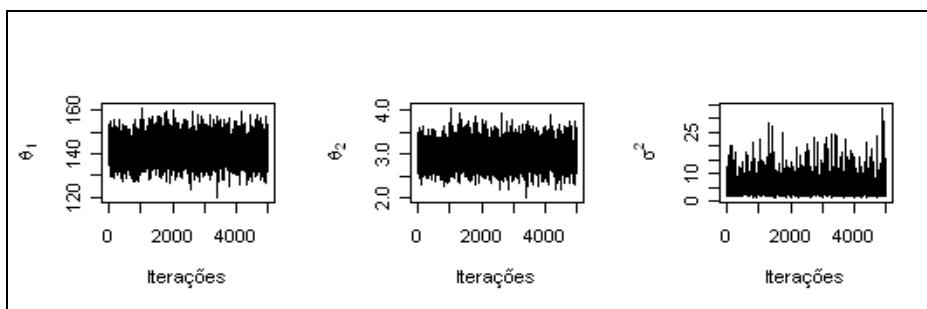


FIGURA 6 Gráfico de traços dos parâmetros da cadeia *dados2*

TABELA 9 Critérios de convergência relacionados a *dados2*

Parâmetros	Raftery & Lewis FD	Geweke p-valor	Gelman & Rubin $\hat{R}$
$\theta_1$	2,1554	0,9763	1,0003
$\theta_2$	2,2103	0,8934	1,0003
$\sigma^2$	0,9824	0,8355	1,0005

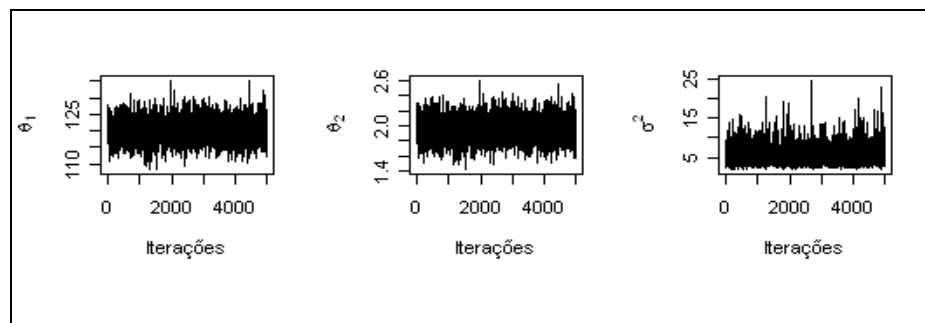


FIGURA 7 Gráfico de traços dos parâmetros da cadeia *dados3*

TABELA 10 Critérios de convergência relacionados a *dados3*

Parâmetros	Raftery & Lewis FD	Geweke p-valor	Gelman & Rubin $\hat{R}$
$\theta_1$	1,0320	0,0720	1,0001
$\theta_2$	1,0320	0,0853	1,0002
$\sigma^2$	1,0075	0,5985	0,9999

Por meio da visualização gráfica percebe-se um comportamento que denota a estabilização da cadeia evidenciando a convergência da mesma para os parâmetros do modelo conforme Figuras 5, 6 e 7 relacionadas a *dados1*, *dados2* e *dados3*.

Observando-se as Tabelas 8, 9 e 10, nota-se que o fator de dependência de Raftery & Lewis (1992), não forneceu evidências contrárias a convergência de todos os parâmetros para os dados, pois seu valor foi menor que 5.

O critério de Geweke (1992) apresentado nas Tabelas 8, 9 e 10 possibilita postular a não existência de evidências contra a convergência porque o  $p$ -valor foi menor do que o nível de significância determinado pelo pesquisador que nesse estudo considerou um nível de significância de 5% e todos os parâmetros analisados nas três massas de dados mostrou não haver nenhuma evidência contra a convergência dos mesmos.

Os valores de  $\hat{R}$  estão em todas as tabelas muito próximos ao valor unitário, situação que indica a convergência segundo Gelman & Rubin (1992).

As estimativas dos parâmetros, ou seja, moda, mediana e média acompanhadas dos seus respectivos desvios padrões, erros de Monte Carlo e o HPD encontram-se nas Tabelas 11,12 e 13.

TABELA 11 Estimativas da moda, mediana, média, *a posterior* dos parâmetros do modelo de Michaelis-Menten, bem como seus respectivos desvios padrões (DP), erros de Monte Carlo (EMC) e intervalos HPD 95% relacionados a *dados1*

Parâmetros	Moda	Mediana	Média	DP	EMC	HPD
$\theta_1$	137,92	139,26	140,96	13,80	0,5790	[113,90;168,3]
$\theta_2$	3,16	3,19	3,28	0,70	0,0299	[1,96; 4,66]
$\sigma^2$	6,79	9,57	13,76	16,86	0,3334	[1,99; 35,92]

TABELA 12 Estimativas da moda, mediana, média, *a posterior* dos parâmetros do modelo de Michaelis-Menten, bem como seus respectivos desvios padrões (DP), erros de Monte Carlo (EMC) e intervalos HPD 95% relacionados a *dados2*

Parâmetros	Moda	Mediana	Média	DP	EMC	HPD
$\theta_1$	140,78	141,16	141,41	5,47	0,0948	[131,00;152,5]
$\theta_2$	2,93	2,95	2,97	0,26	0,0046	[2,47; 3,49]
$\sigma^2$	3,95	3,60	5,29	2,80	0,0416	[1,67; 10,3]

TABELA 13 Estimativas da moda, mediana, média, *a posterior* dos parâmetros do modelo de Michaelis-Menten, bem como seus respectivos desvios padrões (DP), erros de Monte Carlo (EMC) e intervalos HPD 95% relacionados a *dados3*

Parâmetros	Moda	Mediana	Média	DP	EMC	HPD
$\theta_1$	120,08	120,17	120,26	3,12	0,0505	[114,04;126,38]
$\theta_2$	1,95	1,95	1,95	0,14	0,0025	[1,67; 2,22]
$\sigma^2$	4,42	4,93	5,36	2,09	0,0282	[2,22; 9,28]

Observa-se que os erros de Monte Carlo apresentaram valores baixos, confirmando a boa qualidade das estimativas dos parâmetros.

Os HPD para todos os parâmetros foram consideravelmente reduzidos em seus tamanhos proporcionando alta precisão, em especial o parâmetro  $\theta_1$  que tem uma interpretação ecológica das mais importantes por significar o número de espécies que pode ser encontrada no ambiente onde o estudo foi realizado.

A estimativa pontual dos parâmetros acompanhado dos seus HPD com seus desvios padrões faz com que a estimativas se aproxime da quantidade de espécies contidas na população teórica considerada na massa de *dados3*.

As Figuras 8, 9 e 10 representam por meio de gráficos as densidades das distribuições *a posteriori* de *dados1*, *dados2* e *dados3*.

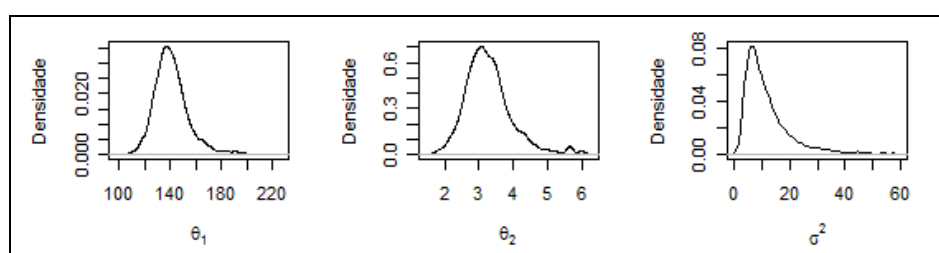


FIGURA 8 Gráfico de densidade da distribuição *a posteriori* dos parâmetros *dados1*

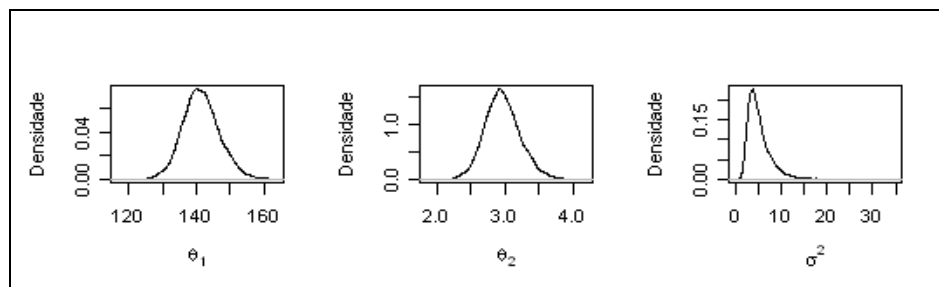


FIGURA 9 Gráfico de densidade da distribuição *a posteriori* dos parâmetros *dados2*

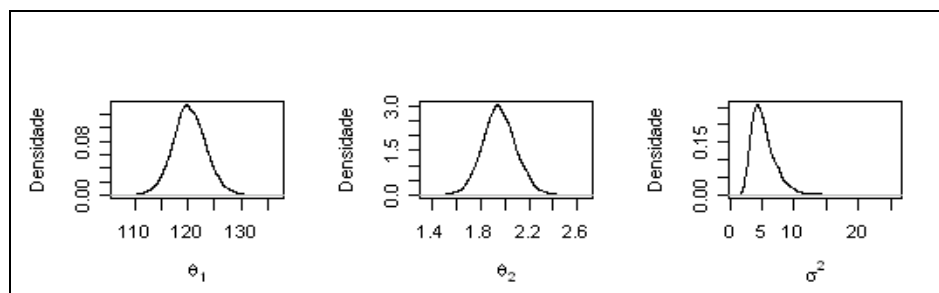


FIGURA 10 Gráfico de densidade da distribuição *a posteriori* dos parâmetros *dados3*



A Figura 11 representa as distribuições de densidades *a posteriori* dos parâmetros do modelo agrupadas em *dados1*, *dados2* e *dados3*.

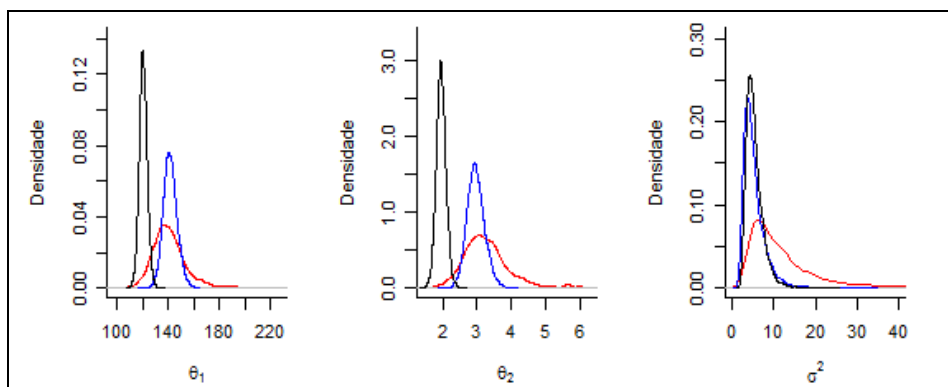


FIGURA 11 Gráfico de densidade da distribuição *a posteriori* dos parâmetros. As curvas em vermelho, azul e preto são referentes aos *dados1*, *dados2* e *dados3*

O significado dos resultados dessa sobreposição de *posteriors* é um indicador que a atualização da informação aplicada sucessivamente, evidencia que todos os parâmetros do modelo aumentam a precisão das estimativas, em especial  $\theta_1$  que fica próximo do número de espécies da população teórica.

O ajuste do modelo aos *dados1*, *dados2* e *dados3* está representado nas Figuras 12, 13 e 14.

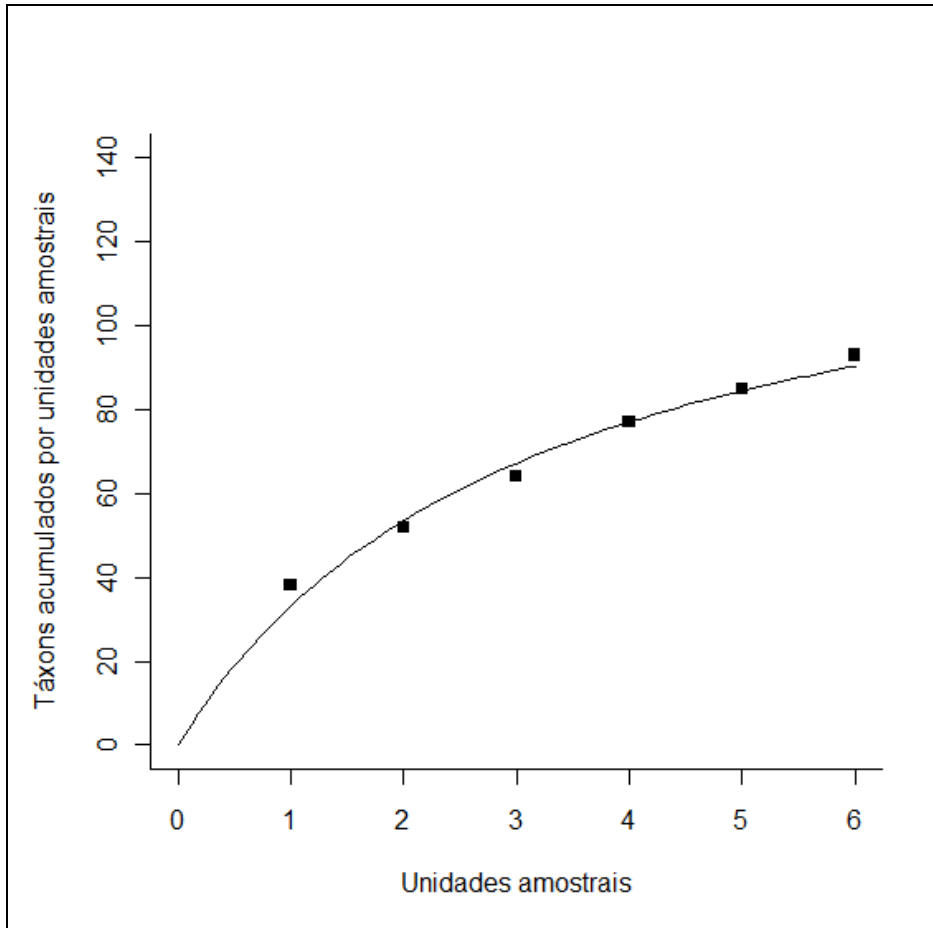


FIGURA 12 Gráfico de ajuste ao modelo relacionado à *dados1* utilizando como parâmetro para o modelo a moda *a posteriori*

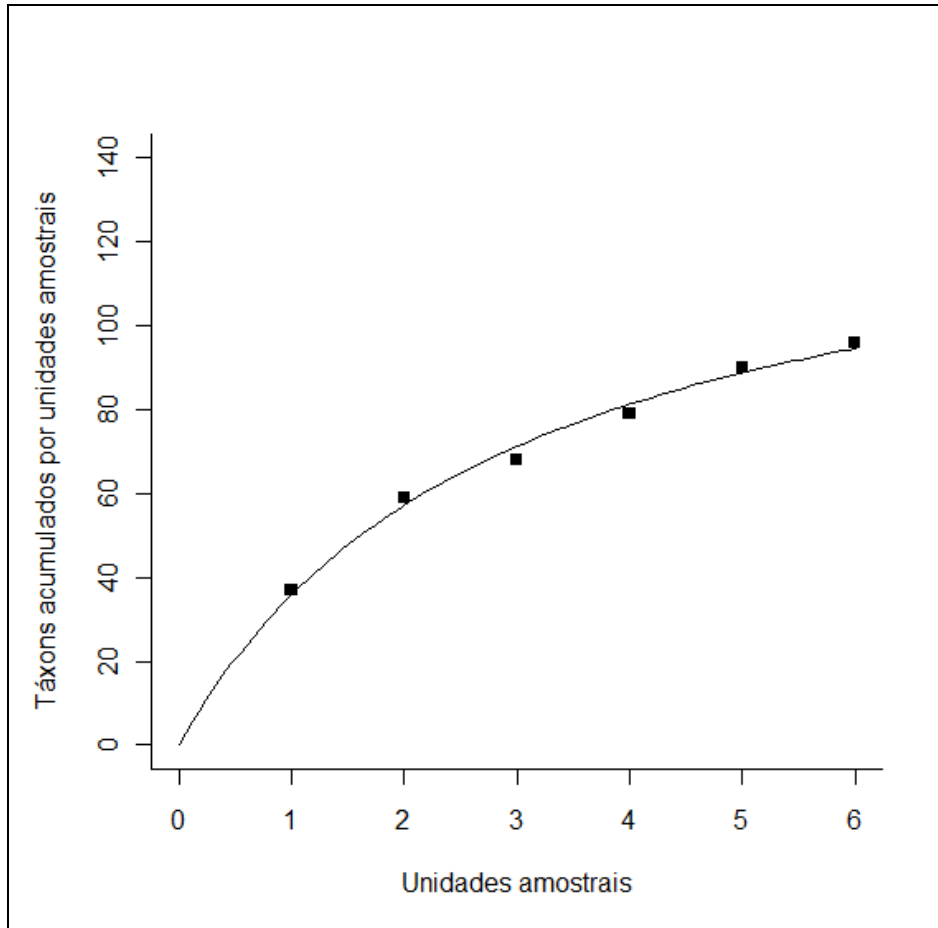


FIGURA 13 Gráfico de ajuste ao modelo relacionado à *dados2* utilizando como parâmetro para o modelo a moda *a posteriori*

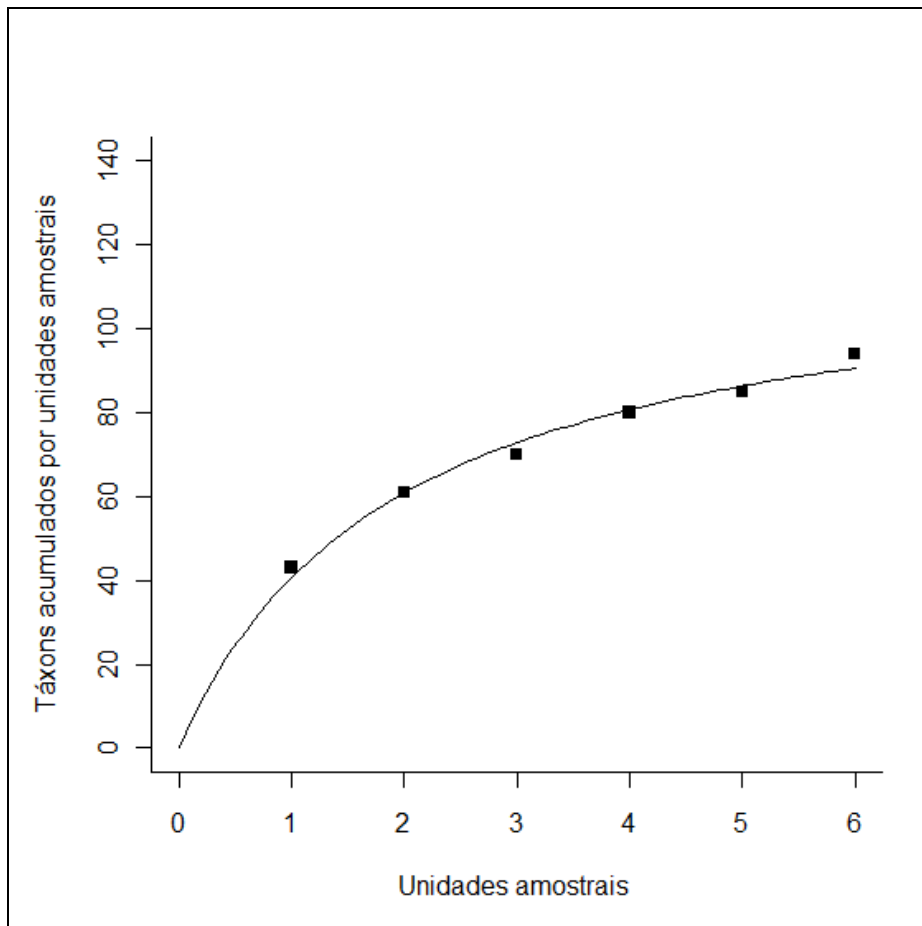


FIGURA 14 Gráfico de ajuste ao modelo relacionado à *dados3* utilizando como parâmetro para o modelo a moda *a posteriori*

O ajuste dos modelos aos dados foi bastante satisfatório em *dados1*, *dados2* e *dados3*.

## 5 CONCLUSÕES

Diante dos resultados obtidos, pode-se concluir que:

1. A metodologia Bayesiana mostrou-se eficiente para estudar a estimação em espécies por meio do ajuste do modelo de Michaelis-Menten, pois produziu estimativas condizentes com a realidade ecológica;
2. A essência deste estudo foi à utilização de um conjunto de distribuições de densidades *a posteriori* para as massas de dados observadas em que se percebeu que as estimativas dos parâmetros de interesse relacionadas ao modelo não-linear apresentaram nesta sequência um melhoramento em relação à precisão e variabilidade.

Em estudos futuros, utilizar coletas de dados amostrais em etapas sucessivas e estimar a riqueza em espécies por meio do modelo não-linear utilizado e outros modelos disponíveis na literatura, aprofundando a idéia em *prioris* mais informativas, com base no atual estudo.

A partir de parcerias a serem firmadas com pesquisadores pertencentes ao programa de Pós-Graduação em ecologia da Universidade Federal do Acre – UFAC, inventariar a comunidade macroinvertebrados bentônicos no PNSD, que é uma unidade de conservação e de proteção integral destinada à preservação de ecossistemas.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANDRÉA, M. M. **Bioindicadores ecotoxicológicos de agrotóxicos**. São Paulo: Centro de Pesquisa e Desenvolvimento de Proteção Ambiental, 2008. (Comunicado Técnico, 83).

BASU, D.; GHOSH, J. K. **Statistical information and likelihood: a collection of critical essays**. New York: Springer-Verlag, 365p. 1988.

BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its application**. New York: J. Wiley, 1988. 392p.

BOLDRINI, J. L.; COSTA, S.I.R.; FIGUEIREDO, V. L.; WETZLER, H. G. **Álgebra linear**. 3. ed. ampl. e rev. São Paulo: Harper & Row do Brasil, 1986.

BOLSTAD, W. M. **Introduction to Bayesian statistics**. Hoboken: J. Wiley, 2004. 354p.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: J. Wiley, 1992. 588p.

BRASIL. Ministério do Meio Ambiente. **Avaliação e identificação de áreas e ações prioritária para a conservação, utilização sustentável e repartição dos benefícios da biodiversidade dos biomas brasileiros**. Brasília, 2002. 404 p.

BRASIL. Ministério do Meio Ambiente. Instruções Normativas n. 5, de 21 de maio de 2004. Lista de espécies da fauna brasileira ameaçadas de extinção. **Diário Oficial da República Federativa do Brasil**, Brasília, 28 maio 2004.

BROEMILING, L. D. **Bayesian analysis of linear models**. New York: M. Dekker, 1985. 454p.

BUSTOS, O. H.; ORGAMBIDE, A. C. F. Simulação estocástica teoria e algoritmos. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 10., 1992, Rio de Janeiro. **Anais...** Rio de Janeiro: UFRJ, 1992.

BUZOLIN, P. G. C. **Uma abordagem clássica e bayesiana para os modelos de Gompertz e de Richards heteroscedásticos**. 2005. 128p. . Dissertação (Mestrado em Estatística) – Universidade Federal de São Carlos, São Carlos.

CALLISTO, M.; FERREIRA, W.; MORENO, P.; GOULART, M.D.C.; PETRUCIO, M. Aplicação de um protocolo de avaliação rápida da diversidade de habitats em atividades de ensino e pesquisa (MG-RJ). **Acta Limnologica Brasiliensia**, São Paulo, v. 14, n. 1, p. 91-98, 2002.

CALLISTO, M.; GONÇALVES JUNIOR, J. F. Bioindicadores bentônicos. In: ROLAND, F.; CÉSAR, D.; MARINHO, M. (Ed.). **Lições de limnologia**. São Paulo: USP, 2005. 379p.

CALLISTO, M.; MORETTI, M.; GOULART, M. D. C. Macroinvertebrados bentônicos como ferramenta para avaliar a saúde de riachos. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v. 6, n. 1, p. 71-82, jan./mar. 2001.

CALOURO, A. M. Riqueza de mamíferos de grande e médio porte do Parque Nacional da Serra do Divisor (Acre, Brasil). **Revista Brasileira de Zoologia**, Curitiba, v. 16, n. 2, p.195-213, 1999. Suplemento.

CARO, M. T.; O'DOHERTY, G. On the use of surrogates species in conservation. **Conservation Biology**, Hoboken, v. 13, n. 4, p. 805-814, 1998.

CODDINGTON, J. A.; YOUNG, L. H.; COYLE, F. A. Estimating spider species richness in a Southern Appalachian cove hardwood cove hardwood Forest. **Journal of Arachnology, New York** v. 24, n. 2, 111-128 p. 111-118, 1996.

COLWELL, R. K. **Estimates**: statistical estimation of species richness and shared species from samples: versão 8.2.0. Disponível em: <<http://viceroy.eeb.uconn.edu/estimates>>. Acesso em: 2 set. 2009.

COLWELL, R. K.; CODDINGTON, J. A. Estimating terrestrial biodiversity through extrapolation. **Philosophical transactions Royal Society of London: series B**, London, v. 345, n. 1344, p. 101-118, July 1994.

CORDEIRO, G. M.; LIMA NETO, E. A. Modelos Paramétricos. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 16., 2004, Recife. **Anais...** São Paulo: Associação Brasileira de Estatística, 246p.2004.

COWLES, M. K.; CARLIN, B. P. **Markov chain Monte Carlo convergence diagnostics**: a comparative review. Minnesota: Universidade de Minnesota, 1994. (Relatório Técnico).

DINERSTEIN, E. D.; OLSON, M.; GRAHAM, D. J.; WEBSTER, A. L.; RIM, A. A.; BOOKBINDER, M. P.; LEDEC, G. **A conservation assessment of the terrestrial ecoregions of Latin America and Caribbean**. Washington: World Wildlife Fund – The World Bank, 1995. 135p.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3.ed. New York: J. Wiley, 1998. 706p.

EHLERS, R. **Introdução à inferência Bayesiana**. 2007. 68p. Disponível em: <<http://leg.ufpr.br/~ehlers/bayes>>. Acesso em 10/09/2009.

EHLERS, R. S. **Métodos computacionalmente extensivos em estatística**. Versão n. 2. 2004. Disponível em: <<http://leg.ufpr.br/~ehlers/notas/mci.pdf>>. Acesso em: 2 jul. 2009.

ESBÉRARD, C. E. L.; BERALLO, H. G. Influência do esforço amostral na riqueza de espécies de morcegos no sudeste do Brasil. **Revista Brasileira de Zoologia**, Curitiba, v. 25, n. 1, p. 67–73, mar. 2008.

ESTEVEES, F. A. **Fundamentos de limnologia**. Rio de Janeiro: Interciência, 1998. 602p.

FLATHER, C. H. Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. **Journal of biogeography**, Oxford, v. 23, n. 2, p.155-168, Mar. 1996.

GALDEAN, N.; CALLISTO, M.; BARBOSA, F. A. R. Lotic ecosystems of Serra do Cipo, southeast Brasil: water quality and a tentative classification based on the benthic macroinvertebrate community. **Aquatic Ecosystem Health and Management**, v. 3, n. 4, p. 545-552, 2000.

GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo: stochastic simulation for Bayesian inference**. 2.ed. New York: Chapman & Hall/CRC, 2006. 323p.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-511, Nov. 1992.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. **IEEE Transaction on Pattern Analysis and Machine Intelligence**, Los Alamitos, v. 6, n. 6, p. 721-741, 1984.



GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; SMITH, A. F. M. (Ed.). **Bayesian statistics 4**. New York: Oxford University, 1992. 526p.

GOULART, M.; CALLISTO, M. Bioindicadores de qualidade de água como ferramenta em estudos de impacto ambiental. **Revista FAPAM**, Pará de Minas v. 2, n. 2, p. 153-164, 2003.

GUTTMAN, I.; MEETER, D. A. On the use of measures of non-linearity. **Thechnometrics**, Washington, v. 7, n. 4, p. 623-637, 1965.

GUYOT, J. L. Hydrogéochimie des fleuves de L'Amazonie Bolivienne. In: COLLECTION Etudes & Thèses. Paris: ORSTOM, 1993. p. 261.

HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, Landing, v. 31, n. 6, p. 1109-1144, 1983.

JEFFREYS, H. The times of P, S, and SKS, and the velocities Of P and S; Monthly Notices R. A. S. Geophys. Helsinki, v.5, p.498-533, 1939. Suppl.

JUNK, W. J. ; SILVA, C. J. Neotropical floodplains: a comparison between the pantanal of Mato Grosso and the large Amazonian river floodplains. In: TUNDISI, J. G.; BICUDO, C. E. M.; TUNDISI, T. M. (Ed.). **Limnology in Brazil**. Rio de Janeiro: ABC/SBL, 1995. p. 195-217.

JUNQUEIRA, V. M.; AMARANTE, M. C.; DIAS, C. F. S. Biomonitoramento da qualidade das águas da bacia do Alto Rio das Velhas através de macroinvertebrados. **Acta Limnológica Brasiliensia**, São Paulo, v. 12, n. 1 p. 73-87, 2000.

KARR, J. R.; DUDLEY, D. R. Ecological perspective on water quality goals. **Environmental Management**, New York, v. 5, n. 1, p. 55-68, 1981.

KHURI, A. I.; CORNELL, J. A. **Response surface: designs and analyses**. New York: Dekker, 1987.

LEENHEER, J. A. Origin and nature of humic substances in the waters of the Amazon River Basin. **Acta Amazônica**, Manaus, v. 10, n. 3, p. 513-526, set. 1980.

LÉVÊQUE, C.; BALIAN, E. V.; MARTENS, K. An assessment of animal species diversity in continental waters. **Hydrobiologia**, Baarn, v. 542, n. 1, p. 39-63, 2005.

LIMA, D. V. M. L. **Macroinvertebrados em igarapés de águas claras e brancas no oeste Amazônico (Acre, Brasil):** implicações para elaboração de estratégias de biomonitoramento. 2008. Dissertação (Mestrado em Ecologia e Manejo de Recursos Naturais) – Universidade Federal do Acre, Rio Branco.

LONGINO, J. T.; COLWELL, R. K. Biodiversity assessment using structured inventory: capturing the ant fauna of a tropical rain Forest. **Ecological Applications**, Ithaca, v. 7, n. 4, p. 1263-1277, 1997.

LOPES, M. A. de O. A.; REHG, J. A. Observation of callimico goeldii with Saguinus imperator in the Serra do Divisor National Park, Acre – Brasil. **Neotropical Primates**, Bogotá, v. 11, n. 3, p. 181-183, 2003.

MACHADO, E. J. **Uso do ‘Bootstrap’ na estimação de parâmetros em modelos não lineares:** uma aplicação em mecanismos cinéticos de Michaelis-Menten. 2006. 131p. Dissertação (Mestrado em Agronomia/Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras.

MAENPAA, K. A.; SORMUNEN, A. J.; KUKKONEN, J. V. K. Bioaccumulation and toxicity of sediment associated herbicides (ioxynil, pendimethalin and bentazone) in Lumbriculus variegates (Oligochaeta) and chironomus riparius (Insecta). **Ecotoxicology and Environmental Safety**, Bethesda, v. 56, n. 3, p. 398-410, Nov. 2003.

MAROCO, J. **Análise estatística:** com utilização do SPSS. 3.ed. Lisboa: Sílabo, 2007. 822p.

MAZZINI, F. **Efeitos da resolução taxonômica de invertebrados bentônicos no diagnóstico da qualidade de ecossistemas lóticos.** 2007. Dissertação (Mestrado em Ecologia) – Universidade de São Paulo, São Paulo.

MELO, A. S. Diversidade de macroinvertebrados em riachos. In: RUDRAN, R.; PÀDUA, C.; CULLEN JÚNIOR, L. **Método de estudos em biologia da conservação manejo da vida silvestre.** Curitiba: UFPR, 2004. cap. 3, p. 69 - 90.

METCALFE, J. L. Biological water quality assessment of running Waters based on macroinvertebrates communities: history and present status in Europe. **Environmental Pollution**, Bethesda, v. 60, n. 1-2, p. 101-139, 1989.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, Woodbury, v. 21, n. 6, p. 1087-1092, 1953.

MICHAELIS, L.; MENTEN, M.L. Die Kinetik der invertinwirkung. **Brochen, Stuttgart**, v. 49, p. 335-369, 1913.

MOULTON, T. P. Saúde e integridade do ecossistema e o papel dos insetos aquáticos. In: NESSIMIAN J. L.; CARVALHO A. L. (Ed.). **Ecologia de insetos aquáticos**. Rio de Janeiro: UFRJ, 1998. p. 281–298. (Oecologia Brasiliensis, 5).

NASCIMENTO, M. **O uso de simulação de Monte Carlos via cadeia de Markov no melhoramento genético**. 2009. 45p. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, Viçosa, MG.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society: statistics in society: series A**, London, v. 135, n. 3, p. 370-384, Aug. 1972.

NOGUEIRA, D. A.; SÁFADI, T.; FERREIRA, D. F. Avaliação de critérios de convergência para o método de Monte Carlo via Cadeia de Markov. **Revista Brasileira de Estatística**, São Paulo, v. 5, n. 1, p. 59-88, 2004.

NONATO, E. A.; VIOLA, Z. G. G.; ALMEIDA, K. C. B. Tratamento estatístico dos parâmetros da qualidade das águas da bacia do alto curso do Rio das Velhas. **Química Nova**, Belo Horizonte, v. 30, n. 4, p. 797-804, 2007.

PAULINO, C. D. ; TURKMAN, M. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 446p. 2003.

PEARSON, D. L. Seleying indicator taxa for the quantitative assessment of biodiversity. **Philosophical transaction of the Royal Society of London: biological science**, London, v. 1311, n. 345, p. 75-79, July 1994.

PEREIRA, J. M. **Comparação entre modelos para predição do nitrogênio mineralizado: uma abordagem bayesiana**. 2006.79 p. Tese (Doutorado em Agronomia/Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

R DEVELOPMENT CORE TEAM. **The R manuals**: current version: 2.9 2010. Disponível em: <<http://www.r-project.org>>. Acesso em: 15 jan. 2010.

RAAIJMAKERS, J. G. W. Statistical analysis of the Michaelis-Menten equation. **Biometrics**, Washington, v. 43, n. 4, p.793-803, Nov. 1987.

RAFTERY, A. L.; LEWIS, S. comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. **Statistical Science**, Hayward, v. 7, n. 4, p. 493-497, Nov. 1992.

RATKOWSKY, D. A. **Nonlinear regression modeling**. A unified practical approach. New York: Marcel Dekker, 1983. 276p.

REIS, R. L. dos. **Comparação Bayesiana de modelos para o desequilíbrio de Hardy-Weinberg**. 2008. 86p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras.

RIBEIRO, L. O.; UIEDA, V. S. Estrutura da comunidade de macroinvertebrados bentônicos de um riacho de serra em Ipatinga, São Paulo, Brasil. **Revista Brasileira de Zoologia**, Curitiba, v. 22, n. 3, p. 613-618, 2005.

ROBERT, C.; CASELLA, G. **Monte Carlo statistical methods**. New York: Springer-Verlag, 1999. 645p.

ROSA, G. J. M. **Análise Bayesiana de modelos mistos robustos via amostrador de Gibbs**. 1998. 57p. Tese (Doutorado em Estatística e Experimentação Agrônômica) – Universidade de São Paulo, Piracicaba.

SANTOS, A. J. dos. Estimativas de riqueza em espécies. In: RUDRAN, R.; PÀDUA, C.; CULLEN JÚNIOR, L. **Método de estudos em biologia da conservação manejo da vida Silvestre**. Curitiba: UFPR, 2004. Cap.1, p. 19-41.

SANTOS, L. A. dos. **Inferência Bayesiana para o número de espécies de uma população via processo de Poisson não homogêneos**. 2000. 110p. Dissertação (Mestrado Ciências da Computação e Matemática computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade São Paulo, São Paulo.

SCHMIT, J. P.; MURPHY, J. F.; MUELLER, G. M. Macrofungal diversity of a temperate oak forest: a test of species richness estimators. **Canadian journal of botany**, Ottawa v. 77, n. 7, p. 1014-1027, July 1999.

SEARLE, S. R. **Linear models**. New York: J. Wiley, 560p. 1971.

SEBER, G. A. F.; WILD, C. J. **Nonlinear regression**. New York: J. Wiley, 800p.1989.

SILVA, N. M.; MUNIZ, J. A.; SILVA, F. F.; AQUINO, L. H. Estudo de parâmetro de crescimento de bezerro Nelore por meio de um modelo de regressão linear: uma abordagem Bayesiana. **Ciência Animal Brasileira**, Goiânia, v. 7, n. 1, p. 57-65, jan./mar. 2006.

SILVEIRA, M.; DALY, D. **Florística e Botânica econômica do Acre**. Rio Branco: Universidade Federal do Acre/The New York Botanical Garden/CNPq/NSF, 1997. 24p.

SIMÕES, N. R. **Estimadores de Bayes, aplicação ao modelo de regressão linear simples**. 1981. 138p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba.

SIOLI, H. **Amazônia: fundamentos da ecologia da maior região de floresta tropical**. 3.ed. Rio de Janeiro: Vozes, 1991. 72p.

SOBERÓN, J.; LLORENTE, J. The use of species accumulation function for the prediction of species richness. **Conservation Biology**, Hoboken, v. 7, n. 3, p. 480-488, Sept. 1993.

SOUZA, E. M. de. **Modelagem não linear da extração de zinco em um solo tratado com lodo de esgoto**. 2008. 95p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras.

SOUZA, M. B. **Diversidade de anfíbios nas unidades de conservação ambiental: reserva extrativista do Alto Juruá (REAJ) e Parque Nacional da Serra do Divisor (PNSD), Acre - Brasil**. 2003. 152p. Tese (Doutorado em Ciências Biológicas/Zoologia) – Instituto de Biociências, Universidade Estadual Paulista, Rio Claro.

SAVIAN, T. V. **Análise bayesiana para modelos de degradabilidade ruminal**. 2008. 81p. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

WILSON, E. O. **A situação atual da atual da diversidade biológica**. Rio de Janeiro: Nova Fronteira, 1997.

## APÊNDICE A

### Regra multiparamétrica de Jeffreys

#### Distribuição *a priori* conjunta para os parâmetros $\theta_1$ e $\theta_2$

Seja a função de verossimilhança (10), ou seja,

$$L_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right]$$

que será descrita por:

$$L(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right]$$

omitindo-se o índice dados1 para facilitar a notação tendo como objetivo simplificar a escrita no desenvolvimento algébrico, proceder-se-á a logaritmação da função de verossimilhança obtendo-se a função de log-verossimilhança:

$$\begin{aligned} \log[L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})] &= \log\left\{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right]\right\} \\ &= \log(2\pi\sigma^2)^{-\frac{n}{2}} + \log\left\{\exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right]\right\} \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2 \\ \log[L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})] &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2 \end{aligned}$$

Aplicam-se derivadas parciais na função de log-verossimilhança em relação aos parâmetros  $\theta_1$ ,  $\theta_2$ ,  $\theta_1^2$ ,  $\theta_2^2$ ,  $\theta_1\theta_2$  e calculam-se as esperanças

( $E[\cdot]$ ) destas derivadas parciais tendo como objetivo construir a matriz de informação de Fisher  $\mathbf{I}(\boldsymbol{\theta})$  multiparamétrica.

$$\begin{aligned} \frac{\partial(\log[L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1} &= \frac{\partial \left[ -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]}{\partial \theta_1} \\ &= -\frac{2}{2\sigma^2} \sum_{i=1}^n \left[ \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \frac{\partial \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)}{\partial \theta_1} \right] \\ &= -\frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) (-1) \left( \frac{x_i}{\theta_2 + x_i} \right) \right] \\ \frac{\partial(\log[L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1} &= \frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{x_i}{\theta_2 + x_i} \right) \right] \\ \frac{\partial^2(\log[L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1^2} &= \frac{\partial \left\{ \frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{x_i}{\theta_2 + x_i} \right) \right] \right\}}{\partial \theta_1} \\ \frac{\partial^2(\log[L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1^2} &= -\frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \\ \frac{\partial(\log[L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2} &= \frac{\partial \left[ -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]}{\partial \theta_2} \end{aligned}$$

$$\begin{aligned}
&= -\frac{2}{2\sigma^2} \sum_{i=1}^n \left[ \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \frac{\partial \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)}{\partial \theta_2} \right] \\
\frac{\partial (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2} &= -\frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right) \right] \\
\frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2^2} &= \frac{\partial \left\{ -\frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right) \right] \right\}}{\partial \theta_2} \\
&= -\frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right) \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right) - 2 \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^3} \right) \right] \\
\frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2^2} &= -\frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( \frac{(\theta_1 x_i)^2}{(\theta_2 + x_i)^4} \right) - 2 \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^3} \right) \right] \\
\frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2 \partial \theta_1} &= \frac{\partial}{\partial \theta_2} \left( \frac{\partial (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1} \right) \\
&= \frac{\partial \left\{ \frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{x_i}{\theta_2 + x_i} \right) \right] \right\}}{\partial \theta_2}
\end{aligned}$$



$$\frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2 \partial \theta_1} = \frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} \right) - \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{x_i}{(\theta_2 + x_i)^2} \right) \right]$$

$$-E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1^2} \right] = -E \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right]$$

$$-E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1^2} \right] = \frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2$$

$$-E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2^2} \right] = -E \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( \frac{(\theta_1 x_i)^2}{(\theta_2 + x_i)^4} \right) - 2 \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^3} \right) \right] \right]$$

$$= -E \left[ \frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{(\theta_1 x_i)^2}{(\theta_2 + x_i)^4} \right) \right] - E \left[ 2 \frac{1}{\sigma^2} \sum_{i=1}^n \left( \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^3} \right) \right) \right]$$

$$-E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2^2} \right] = \frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{(\theta_1 x_i)^2}{(\theta_2 + x_i)^4} \right)$$

$$\begin{aligned}
-E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2 \partial \theta_1} \right] &= -E \left[ \frac{1}{\sigma^2} \sum_{i=1}^n \left[ \left( \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right) \left( \frac{x_i}{\theta_2 + x_i} \right) - \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{x_i}{(\theta_2 + x_i)^2} \right) \right] \right] \\
&= -E \left[ \frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} \right) \right] - E \left[ \frac{1}{\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \left( \frac{x_i}{(\theta_2 + x_i)^2} \right) \right] \\
-E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2 \partial \theta_1} \right] &= -\frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} \right)
\end{aligned}$$

A partir das  $(E[\cdot])$  das derivadas parciais em relação aos parâmetros forma-se a matriz de informação de Fisher ( ),

$$( ) = \begin{bmatrix} -E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1^2} \right] & -E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_1 \partial \theta_2} \right] \\ -E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2 \partial \theta_1} \right] & -E \left[ \frac{\partial^2 (\log [L(\theta_1, \theta_2, \sigma^2 | \mathbf{y})])}{\partial \theta_2^2} \right] \end{bmatrix}$$

$$( ) = \begin{bmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 & -\frac{1}{\sigma^2} \sum_{i=1}^n \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} \\ -\frac{1}{\sigma^2} \sum_{i=1}^n \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} & \frac{1}{\sigma^2} \sum_{i=1}^n \frac{(\theta_1 x_i)^2}{(\theta_2 + x_i)^4} \end{bmatrix}$$

calcula-se o determinante desta matriz, denotado por  $|(\cdot)|$

$$\begin{aligned}
 |(\cdot)| &= \left\{ \left[ \frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \frac{1}{\sigma^2} \sum_{i=1}^n \frac{(\theta_1 x_i)^2}{(\theta_2 + x_i)^4} \right] - \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} \right] \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} \right] \right\} \\
 |(\cdot)| &= \left\{ \frac{\theta_1^2}{(\sigma^2)^2} \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \frac{\theta_1^2}{(\sigma^2)^2} \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\} \\
 |(\cdot)| &= \frac{\theta_1^2}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\} \\
 |(\cdot)| &= \theta_1^2 \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\} \\
 &\qquad\qquad\qquad q(\theta_2) \\
 |(\cdot)| &= \theta_1^2 q(\theta_2)
 \end{aligned}$$

em que,

$$q(\theta_2) = \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}$$

portanto *a priori* de Jeffreys para  $\theta = (\theta_1, \theta_2)$  é dada por:

$$\begin{aligned}
 P_{\text{dados}11}(\theta_1, \theta_2) &\propto |\theta_1^2 q(\theta_2)|^{\frac{1}{2}} \Rightarrow \\
 P_{\text{dados}1}(\theta_1, \theta_2) &\propto \theta_1^2 \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{\frac{1}{2}} \quad (11)
 \end{aligned}$$

## APÊNDICE B

### Rotina R para gerar as cadeias via MCMC do modelo de Michaelis-Menten

#Obs.: na rotina "a" representa theta1, "b" representa theta2 e "c" representa sigma2

# Amostra 1

# Entrando com os dados

x=c(1,2,3,4,5,6)

y=c(38,52,64,77,85,93)

# definindo o tamanho da amostra

n=length(y)

# definindo contador para a taxa de aceitação

cont1=0 # contador para o theta1

cont2=0 # contador para o theta2

# definindo a distribuição *a posteriori* da amostra 1 (Eq. 9)

```
p=function(a,b,c) (c^-((n+1)/2+1))*a*  
  (abs((sum((x/(b+x))^2))*sum((x^2)/((b+x)^4))-  
  ((sum((x^2)/((b+x)^3))^2))^0.5))*  
  exp((-1/(2*c))*(sum((y-(a*x)/(b+x))^2)))
```

# algoritmo com *Burn-in* e jump

*Burn-in*=0

jump = 1

iter=jump\*5000+*Burn-in*

theta1=rep(0,iter) # armazenar o theta1

```

theta2=rep(0,iter) # armazenar o theta2
sigma=rep(0,iter) # armazenar o sigma

# chutes iniciais

theta1[1]=139
theta2[1]=3.2
sigma[1]=10

for(i in 2:iter)
{

#utilizando o gibbs para o sigma

alfa=(n+1)/2
beta=0.5*sum((y-(theta1[i-1]*x))/(theta2[i-1]+x))^2
sigmanovo=1/rgamma(1,alfa,beta)
sigma[i]=sigmanovo

#utilizando o Metr polis para o theta 1

cand1=abs(rnorm(1,139,20)) # candidata
qcand1=dnorm(cand1,139,20)# candidata em candidato dado theta1
qthet1=dnorm(theta1[i-1],139,20)# candidata em theta1 dado candidato
aux1=p(cand1,theta2[i-1],sigma[i])*qthet1
aux2=p(theta1[i-1],theta2[i-1],sigma[i])*qcand1

prob=min(1,aux1/aux2)
u = runif(1) # valor u para ser comparado com alfa
if (u<=prob){
  theta1[i]=cand1
  cont1 = cont1 + 1 # conta quantas vezes a cadeia recebe um novo valor
}
else {
  theta1[i]=theta1[i-1]
}
}
}

```

```

# utilizando o Metrópolis para o theta 2

cand2=abs(rnorm(1,3.5,0.5)) # candidata
qcand2=dnorm(cand2,3.5,0.5)# candidata em candidato dado theta2
qthet2=dnorm(theta2[i-1],3.5,0.5)# candidata em theta2 dado candidato
aux3=p(theta1[i],cand2,sigma[i])*qthet2
aux4=p(theta1[i],theta2[i-1],sigma[i])*qcand2

prob2=min(1,aux3/aux4)
u2 = runif(1) # valor u2 para ser comparado com alfa
if (u2<=prob2){
  theta2[i]=cand2
  cont2 = cont2 + 1 # conta quantas vezes a cadeia recebe um novo valor
}
else {
  theta2[i]=theta2[i-1]
}
if(i > Burn-in && i%%jump == 0)
write(t(c(theta1[i],theta2[i],sigma[i])), "cadeiapre.txt",3,append=TRUE)

}

# Taxa de Aceitação para theta1

cont1/iter

# Taxa de Aceitação para theta2

cont2/iter

dados=read.table("cadeiapre.txt")
attach(dados)

mean(dados)

acf(V1,500)

```

```
library(coda)  
raftery.diag(dados)
```

```
library(boa)  
boa.menu()
```

```
acf(V1,lag=600)
```

```
reg=nls(y~a*x/(b+x),start=c(a=100,b=1))  
summary(reg)
```

## APÊNDICE C

### Rotinas específicas dos resultados

**# Amostra 1**

**# Lendo os dados**

**dados1=as.matrix(read.table("cadeia1.txt"))**

**dados2=as.matrix(read.table("cadeia2.txt"))**

**#####**

**# Calculando a moda**

**library(modeest) # pacote necessário para o cálculo da moda**

**# Calculando a Moda para os três primeiros parâmetros**

**moda1=mlv(dados1[,1])**

**round(modas1\$M,2) # moda para o theta1**

**moda2=mlv(dados1[,2])**

**round(modas2\$M,2) # moda para o theta2**

**moda3=mlv(dados1[,3])**

**round(modas3\$M,2) # moda para o sigma**

**#####**

**library(coda)**

**# Gelman & Rubin para theta1**

**rbcad1 <- mcmc.list(mcmc(dados1[,1]),mcmc(dados2[,1]))**

**gel1=gelman.diag(rbcad1)**

**round(gel1\$psrf,4)**



```

# Gelman & Rubin para theta2

rbcad2 = mcmc.list(mcmc(dados1[,2]),mcmc(dados2[,2]))
gel2=gelman.diag(rbcad2)
round(gel2$psrf,4)

# Gelman & Rubin para variância

rbcad3 = mcmc.list(mcmc(dados1[,3]),mcmc(dados2[,3]))
gel3=gelman.diag(rbcad3)
round(gel3$psrf,4)

#####
# Resultados Gráficos

par(mfrow=c(3,3))
plot(dados1[,1],xlab="Iterações",ylab=expression(theta[1]),type="l")
plot(dados1[,2],xlab="Iterações",ylab=expression(theta[2]),type="l")
plot(dados1[,3],xlab="Iterações",ylab=expression(sigma^2),type="l")
plot(density(dados1[,1]),xlab=expression(theta[1]),ylab="Densidade",main
="")
plot(density(dados1[,2]),xlab=expression(theta[2]),ylab="Densidade",main
="")
plot(density(dados1[,3]),xlab=expression(sigma^2),ylab="Densidade",main
="")
hist(dados1[,1],xlab=expression(theta[1]),ylab="Frequencia",main="")
hist(dados1[,2],xlab=expression(theta[2]),ylab="Frequencia",main="")
hist(dados1[,3],xlab=expression(sigma^2),ylab="Frequencia",main="")

#####

```

```
# Gráficos de Gelman & Rubin
```

```
gelman.plot(rbcad1, xlab="Últimas iterações na cadeia",ylab="Fator  
Shrink",main=expression(theta[1]))  
gelman.plot(rbcad2,xlab="Últimas iterações na cadeia",ylab="Fator  
Shrink",main=expression(theta[2]))  
gelman.plot(rbcad3,xlab="Últimas iterações na cadeia",ylab="Fator  
Shrink",main=expression(sigma^2))
```

```
#####
```

```
# resultados usando o BOA
```

```
library(boa)  
boa.menu()
```