

EDUARDO CAMPANA BARBOSA

CHOICE-BASED CONJOINT ANALYSIS: UM ENFOQUE BAYESIANO

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de Magister Scientiae.

VIÇOSA
MINAS GERAIS – BRASIL
2015

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

B238c
2015

Barbosa, Eduardo Campana, 1988-
Choice-Based Conjoint Analysis : Um Enfoque Bayesiano /
Eduardo Campana Barbosa. – Viçosa, MG, 2015.
x, 102f. : il. ; 29 cm.

Inclui apêndices.

Orientador: Carlos Henrique Osório da Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.83-93.

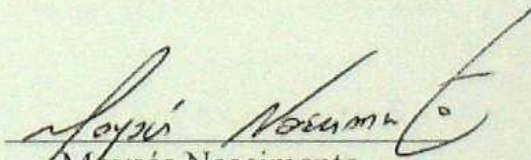
1. Inferência Bayesiana. 2. Consumidores - Preferência.
3. Marketing. 4. Probabilidade. I. Universidade Federal de
Viçosa. Departamento de Estatística. Programa de
Pós-graduação em Estatística Aplicada e Biometria. II. Título.

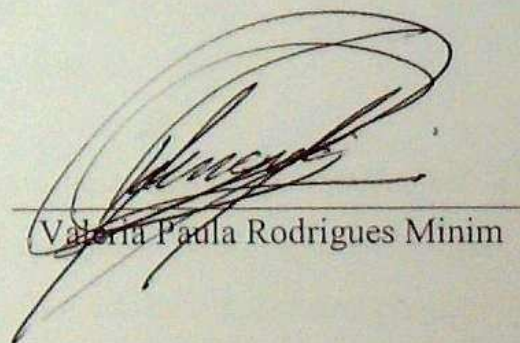
CDD 22. ed. 519.542

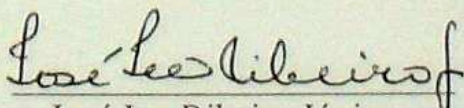
CHOICE-BASED CONJOINT ANALYSIS: UM ENFOQUE BAYESIANO


Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 25 de Fevereiro de 2015


Moysés Nascimento
(Co-orientador)


Valéria Paula Rodrigues Minim


José Ivo Ribeiro Júnior


Carlos Henrique Osório Silva
(Orientador)

Aos meus pais Jorge Luiz Resende Barbosa e Denise Gama
Campana Barbosa. A minha irmã Gabriela Campana Barbosa.

Agradecimentos

- Em primeiro lugar a Deus, por permitir que mais uma importante etapa de minha vida possa ser concluída.
- A meu pai (Jorge) e minha mãe (Denise), que desde a época de graduação, com muito esforço, permitiram com que eu pudesse me dedicar inteiramente aos estudos, o que influenciou meu interesse profissional pela área acadêmica e conseqüentemente para o ingresso no programa de mestrado em Estatística Aplicada e Biometria. Agradeço também a minha irmã (Gabriela), que sempre esteve ao meu lado, mesmo à distância, todas as vezes que precisei.
- A uma pessoa em especial, Camila de Sousa do Carmo, que teve uma participação importantíssima em minha vida, principalmente para o meu ingresso no mestrado. Jamais me esquecerei das palavras de incentivo e os dias em que você, pacientemente, sentou-se a mesa comigo para resolvermos os exercícios de probabilidade (que quase me fizeram desistir), quando estava estudando para a prova de seleção. Estará para sempre em minha memória a emoção daquela manhã em sua casa (ainda em Rodeiro), quando juntos visualizamos o resultado final do processo seletivo, o qual fui aprovado em segundo lugar. Muito obrigado por tudo, por acreditar e sempre me apoiar, mesmo nos momentos que nem eu acreditava!
- Ao meu orientador (ou como costumo dizer, meu pai acadêmico) Carlos Henrique Osório Silva (C.H.O.S.) por sua paciência, dedicação e amizade. Além disso, fica o meu agradecimento por todos os conhecimentos transmitidos e pelas diversas vezes que leu meus trabalhos, corrigindo-os e oferecendo suas sugestões, sempre com boa vontade.
- Aos meus co-orientadores, Fabyano Fonseca e Silva, Moysés Nascimento e Rosires Deliza, pela atenção e por dedicarem um tempo para me atenderem, sempre que precisei. Em especial, ao professor e amigo Moysés Nascimento, que desde o início me tratou com respeito e profissionalismo, além de acreditar e confiar no meu trabalho.

- A todos os professores e funcionários do Departamento de Estatística Aplicada e Biometria da Universidade Federal de Viçosa (DET-UFV), pelo acolhimento e ensinamentos. Em especial à secretária da pós-graduação, Carla Zinato Campos, que muitas vezes me ajudou a resolver burocracias e problemas, mostrando-se sempre atenciosa, eficiente e com extrema educação e paciência. Agradeço também a funcionária Anita Paiva, pela atenção e momentos divertidos na secretaria do DET-UFV.
- Aos meus colegas de mestrado Rômulo (Manuli), Vladimir (Vlad), Regiane, Patrícia, Filipe (Formiga), Matheus, Lidiane, Patrícia e Rafael, que dividiram comigo as dificuldades e alegrias proporcionadas pela estatística. Em especial ao amigo Rômulo, um grande companheiro nas vésperas de prova, nos trabalhos acadêmicos e para toda a vida. Jamais me esquecerei dos seminários, das nossas conversas e das vezes que me ajudou (R\$) nas inscrições dos congressos. E também ao amigo Filipe, por sua sinceridade e companheirismo, sempre procurando tornar o ambiente mais agradável e divertido. Não me esquecerei daquele SINAPE em Natal-RN, onde demos várias gargalhadas. A todos, obrigado pela amizade.
- Aos meus companheiros e amigos da República Pós, Heider (José), Júlio, Jeferson, Luciano, Ítalo, Vinícius (501), Ivan, Fernando e Rafael, com os quais tive o prazer de conviver e dividir não apenas as contas da casa, mas também experiências, momentos agradáveis e divertidos, que pretendo também jamais me esquecer.

Sumário

Lista de Figuras	vii
Lista de Tabelas	viii
Resumo	ix
Abstract	x
Introdução	1
Capítulo 1	6
Choice-Based Conjoint Analysis	6
1.1 Conceitos Introdutórios.....	6
1.2 Delineamentos de Tratamentos	11
Capítulo 2	16
Modelo Logit Multinomial	16
2.1 Fundamentos Teóricos Introdutórios	16
2.2 Desenvolvimento do Modelo	20
2.3 Aplicações Práticas	24
2.4 Estimação dos Parâmetros	25
2.5 Razão de Escolhas.....	30
2.6 Limitações do Modelo	32
Capítulo 3	33
Choice-Based Conjoint Analysis: Um enfoque Bayesiano	33
3.1 Uma abordagem Bayesiana para o Modelo Logit Multinomial.....	33
3.2 Distribuição a Priori.....	35
3.3 Função de Verossimilhança	38
3.4 Distribuição a Posteriori.....	40
3.4.1 Distribuição a Posteriori Marginal	42
3.5 Conceitos Iniciais sobre os Algoritmos MCMC	45
3.5.1 Aspectos Computacionais sobre os Algoritmos MCMC	48
3.5.2 Independence Metropolis Hasting	49
3.6 Critérios de Convergência.....	56

3.7 Medidas Adicionais	58
3.8 Principais Vantagens do Enfoque Bayesiano.....	60
Capítulo 4.....	62
Resultados.....	62
4.1 Descrição do Conjunto de Dados.....	62
4.2 Resultados Frequentistas para Choice-Based Conjoint Analysis	65
4.3 Resultados Bayesianos para Choice-Based Conjoint Analysis	71
Considerações Finais	82
Referências Bibliográficas.....	83
Apêndice A.....	94
Apêndice B.....	96
Apêndice C.....	98
Apêndice D.....	99

Lista de Figuras

Figura 1 – Histograma de Probabilidades e densidades esperadas.	73
Figura 2 - Gráficos para as distribuições a posteriori marginais dos parâmetros.	74
Figura 3 - Comportamento de Convergência das cadeias de Markov.....	75
Figura 4 - Densidades a posteriori estimadas para cada uma das probabilidades de escolha.	78
Figura 5 - Comparação entre as probabilidades de escolha para os tratamentos D, E e J.....	79
Figura 6 - Densidades a posteriori estimadas para as razões de escolha.....	81

Lista de Tabelas

Tabela 1 - Esquema com Fatores e Níveis do tratamento analisado (camisa)	7
Tabela 2 - Tratamentos segundo um delineamento Fatorial Completo	7
Tabela 3 - Fatores, níveis e codificação.....	62
Tabela 4 - Número de vezes que cada tratamento foi escolhido.....	62
Tabela 5 - Tratamentos Avaliados em um Esquema Fatorial Completo.....	63
Tabela 6 - Organização do conjunto de dados para estudos de CBCA.....	64
Tabela 7 - Comparação de modelos com e sem a presença de covariáveis.	65
Tabela 8 - Teste de hipótese simultâneo para os coeficientes do modelo.....	66
Tabela 9 - Estimação dos parâmetros do modelo pela abordagem Frequentista e teste de Wald.	66
Tabela 10 - Probabilidades observadas e Estimadas pelo modelo Logit Multinomial.	67
Tabela 11 - Resultados Iniciais do critério de Raftery-Lewis.....	71
Tabela 12 - Resultados da distribuição a posteriori dos betas para a CBCA via método de estimção Bayesiano.....	72
Tabela 13 - Verificação da convergência das cadeias pelo critério de Geweke.....	74
Tabela 14 - Resultados Bayesianos para as Probabilidades de Escolha.	77
Tabela 15 - Estimativas Bayesianas para as Razões de Escolhas.	80

Resumo

BARBOSA, Eduardo Campana, M. Sc., Universidade Federal de Viçosa, fevereiro de 2015. **CHOICE-BASED CONJOINT ANALYSIS: UM ENFOQUE BAYESIANO.** Orientador: Carlos Henrique Osório Silva. Coorientadores: Moysés Nascimento, Fabyano Fonseca e Silva e Rosires Deliza.

A presente dissertação teve como objetivo principal demonstrar um enfoque Bayesiano para a metodologia Choice-Based Conjoint Analysis (CBCA). Apresenta-se no texto uma ampla revisão sobre a CBCA (Capítulo 1), sobre o modelo Logit Multinomial [desenvolvimento do modelo, procedimentos de estimação de parâmetros, probabilidades e razões de escolha (Capítulo 2)] e sobre o enfoque de estimação Bayesiano [distribuição a priori utilizada, aproximação de Laplace para a função de verossimilhança, distribuições a posteriori e detalhes sobre o algoritmo MCMC empregado (Capítulo 3)]. No Capítulo 4 apresenta-se um exemplo hipotético, no intuito de demonstrar os resultados e inferências que podem ser obtidos por meio desta recente abordagem (Bayesiana), sendo também apresentados os resultados do enfoque Frequentista. O tratamento em estudo foi um tipo de refrigerante e avaliou-se o efeito de três fatores (A, B e C) na intenção de compra de 96 consumidores, por meio de dados simulados. As análises estatísticas foram conduzidas no software livre R, cujos scripts encontram-se disponibilizados nos apêndices desta dissertação. Concluiu-se que a abordagem Bayesiana para CBCA apresentou resultados interessantes e satisfatórios, com estimativas similares às Frequentistas e mostrando-se uma alternativa metodológica viável para os estudos de CBCA. Adicionalmente, a abordagem proposta possibilitou ainda ao pesquisador construir intervalos de credibilidade (percentis das distribuições a posteriori) para as probabilidades e razões de escolha, no intuito de comparar estas quantidades ou testar hipóteses sobre estas. Quanto aos resultados práticos, a maior probabilidade de escolha estava associada ao tratamento 4, composto pelo nível a_1 do fator A, nível b_2 do fator B e nível c_1 do fator C.

Abstract

BARBOSA, Eduardo Campana, M. Sc., Universidade Federal de Viçosa, February, 2015. **CHOICE-BASED CONJOINT ANALYSIS: A BAYESIAN APPROACH.** Advisor: Carlos Henrique Osório Silva. Co-Advisors: Moysés Nascimento, Fabyano Fonseca e Silva and Rosires Deliza.

This dissertation main goal is to demonstrate the Bayesian approach to Choice-Based Conjoint Analysis (CBCA). We present a comprehensive review of the CBCA methodology (Chapter 1), on the Multinomial Logit model [model development, parameter estimation procedures, probabilities of choice ratios (Chapter 2)] and on the Bayesian estimation approach [prior distribution, Laplace approach to the likelihood function, posterior distributions and details about the MCMC algorithm we applied (Chapter 3)]. In Chapter 4 we present a hypothetical example, in order to demonstrate the results and inferences that can be obtained through this recent approach (Bayesian), and we also present the results of the frequentist approach. The treatment for the study was a type of refrigerant (soda or soft drink) and we evaluated the effect of three factors (volume, type and color) on purchase intention of 96 consumers, using simulated data. Statistical analyzes were conducted with the free software R, whose scripts are provided in the appendices of this dissertation. It was concluded that the Bayesian approach to CBCA presented interesting and satisfactory results, with estimates similar to the frequentist ones, therefore proved to be a viable alternative methodology for CBCA studies. Additionally, the proposed approach also allows the researcher to build credible intervals (percentiles of the posterior distributions) for the probabilities and choice ratios, in order to compare these quantities or test hypotheses about them. In terms of practical or applied results, the highest estimated probability of choice was obtained for treatment 4, with a_1 level of factor A, b_2 level of factor B and c_1 level of factor C.

Introdução

Diversos fatores contribuem para que a competitividade entre as empresas aumente a cada dia, dentre estes pode-se citar a evolução tecnológica, a mudança de percepção dos consumidores em termos de qualidade de produtos, serviços, conceitos etc. (aos quais denomina-se de uma forma geral como um tratamento), além da busca por satisfação pessoal. Nota-se ainda que o nível de exigência dos consumidores em relação aos tratamentos aumenta continuamente, fato que reflete diretamente na necessidade que as empresas possuem de estabelecer novas estratégias competitivas e mercadológicas para fidelizar os seus clientes.

Estudos são conduzidos na área de marketing no intuito de tentar entender a valiosa relação entre empresa, consumidor e tratamentos. Segundo Kotler e Armstrong (2008), a lucratividade e o sucesso de uma empresa são objetivos específicos que estão associados ao bom relacionamento e a uma comunicação eficiente com o seu cliente. Isso pode ser explicado devido o consumidor ser considerado como o destino final do tratamento, tornando evidente o quão importante e forte é a influência deste para que uma empresa tenha sucesso perante a agressividade do mercado.

As empresas precisam oferecer a seus clientes (consumidores) o que estes realmente desejam, com qualidade e eficiência, visando à sua satisfação (do cliente) e também maior rentabilidade (da empresa). No entanto, a principal pergunta a ser feita é: como descobrir que tipo de tratamento o cliente realmente deseja? Logo, o primeiro ponto para responder esta questão consiste em entender os gostos e conseqüentemente as preferências dos clientes.

É neste contexto que a técnica estatística denominada Conjoint Analysis (Análise Conjunta de Fatores) mostra-se uma ferramenta de fundamental importância para o cenário empresarial, seja para auxiliar em tomadas de decisões ou para o desenvolvimento de novos tratamentos. Por meio deste método é possível avaliar a preferência dos consumidores em relação a diversos tratamentos. De uma forma resumida, a Conjoint Analysis consiste em avaliar tratamentos que são formados pela combinação de algumas características específicas, buscando-se entender a importância de cada uma destas para o consumidor, no momento da compra (MOSKOWITZ et al., 2004).

Vale ressaltar que o termo Conjoint (de Conjunta) se deve ao fato de que quando as características dos tratamentos são estudadas conjuntamente, seu valor relativo pode ser mensurado com uma intensidade que não é possível de ser obtida quando estas são avaliadas separadamente (CHURCHILL JR.; NIELSEN JR., 1996). Esta é uma das principais características da Conjoint Analysis e, portanto, em situações práticas é adequado que os consumidores entendam que estão avaliando um tratamento e não as características específicas que os compõe separadamente.

Como tal método baseia-se na avaliação dos consumidores, o analista (aplicador do estudo) precisa planejar os cenários que serão apresentados a estes, visando obter resultados confiáveis. Portanto, em estudos de Conjoint Analysis, uma etapa inicial é o planejamento do experimento. A Conjoint Analysis é definida como um experimento, em função das ferramentas que utiliza em seu planejamento, e também como uma pesquisa de marketing, em função do objetivo de estimar a preferência do consumidor. Gustafsson et al. (1999) definem a Conjoint Analysis como um experimento aplicado às decisões de marketing. Mais adiante no texto serão descritos os aspectos relacionados ao planejamento de um experimento em Conjoint Analysis e da sua importância para a obtenção dos resultados.

A Conjoint Analysis engloba um conjunto de técnicas estatísticas, visto que com o passar do tempo várias modificações foram propostas no intuito de melhor compreender a complexa relação entre consumidor e tratamentos. Na abordagem inicial, denominada como Ratings-Based Conjoint Analysis (Análise Conjunta Baseada em Notas), cada consumidor avalia um determinado número de tratamentos e atribui a cada um destes uma nota dentro de uma escala. Em uma alternativa similar, denominada como Ranking-Based Conjoint Analysis (Análise Conjunta Baseada em Postos), os consumidores ordenam cada tratamento de acordo com o grau de preferência. Por exemplo, se forem apresentados quatro tratamentos, cada consumidor irá atribuir o valor 1 para o tratamento de maior preferência, 2 e 3, para respectivamente, o segundo e o terceiro, e 4 para o de menor preferência (LOCKSHIN et al., 2006 e HAIR JR., 2005). A análise estatística para estes dois casos consiste na utilização de modelos usuais de Regressão Múltipla. Maiores detalhes sobre essas duas metodologias podem ser encontrados nos textos de Artes (1991), Siqueira (2000), Bastos (2010) e Silva e Bastos (2010).

Neste trabalho focou-se em uma terceira abordagem de Conjoint Analysis, denominada como Choice-Based Conjoint Analysis (Análise Conjunta de Fatores

Baseada em Escolhas, denotada aqui como CBCA), em que os consumidores informam sua preferência por um ou mais tratamentos por meio de escolhas. Além disso, propõe-se um enfoque Bayesiano para a CBCA, no intuito de apresentar uma nova abordagem e suas principais vantagens, como os intervalos de credibilidade para os fatores, Probabilidades e Razões de Escolhas (principais resultados interpretativos da CBCA). No enfoque Bayesiano estes intervalos são facilmente obtidos tomando-se os quantis de 2,5% e 97,5% da distribuição a posteriori associada a cada uma destas quantidades. Na abordagem Clássica (Frequentista), obtêm-se intervalos de confiança que são assintóticos, ou seja, válidos para grandes amostras e confiáveis somente se a distribuição de interesse for aproximadamente normal. Vale ressaltar que na abordagem Frequentista, tais intervalos são obtidos apenas para os fatores e Razões de Escolhas. Pelo enfoque Bayesiano é possível acessar os intervalos de credibilidade para as 3 quantidades mencionadas anteriormente, inclusive para as Probabilidades de Escolha. Tal fato permite testar hipóteses, por exemplo, sobre a igualdade entre Probabilidades de Escolha de diferentes tratamentos, o que não é possível pela abordagem Frequentista.

Objetivos

- Demonstrar como se procede a CBCA pela abordagem Frequentista, desde a dedução do modelo estatístico, estimação dos parâmetros e à obtenção das Probabilidades e Razões de Escolha;
- Demonstrar o desenvolvimento do enfoque Bayesiano para CBCA, desde a especificação da distribuição a priori, função de verossimilhança e obtenção da distribuição a posteriori. Deduzir o funcionamento do algoritmo MCMC Independence Metropolis Hasting, empregado nesta análise;
- Desenvolver, por simulação de dados, um estudo de CBCA no intuito de comparar ambos os enfoques propostos (Frequentista e Bayesiano), com o objetivo de verificar se o enfoque Bayesiano pode ser empregado para estudos de CBCA.

Estrutura do Trabalho

No Capítulo 1 apresentou-se uma revisão sobre a CBCA, em que foram descritas as características do método (fatores, níveis e tratamentos), os principais delineamentos de tratamentos (delineamentos fatorial fracionado e ortogonal) e como deve ocorrer a apresentação dos tratamentos aos consumidores, neste tipo de análise.

No Capítulo 2 deduziu-se o modelo Logit Multinomial, empregado na CBCA, suas principais propriedades, aplicações e método de estimação dos parâmetros. Além disso, relatou-se sobre a obtenção e interpretação das probabilidades e razões de escolha.

O Capítulo 3 exibiu uma revisão sobre o enfoque Bayesiano para a CBCA. Apresentou-se a modelagem completa, desde a escolha da distribuição a priori, função de verossimilhança, densidade a posteriori e o funcionamento do algoritmo MCMC empregado.

No Capítulo 4 apresentou-se um estudo desenvolvido por simulação de dados para ilustrar a aplicação de ambos os enfoques (Frequentista e Bayesiano) para a CBCA. Neste estudo, foram desenvolvidos hipoteticamente 12 tratamentos. Definiu-se o número de vezes que cada um destes tratamentos foi escolhido, para compor a variável resposta.

Nas considerações finais ressaltou-se as principais contribuições do enfoque Bayesiano proposto para a CBCA, juntamente com recomendações para estudos futuros.

Capítulo 1

Choice-Based Conjoint Analysis

1.1 Conceitos Introdutórios

A Choice-Based Conjoint Analysis (CBCA) é uma técnica estatística aplicada em estudos onde os tratamentos avaliados são formados por características específicas comumente conhecidas como fatores, por isso a tradução como Análise Conjunta de Fatores. Se em um estudo “r” fatores forem avaliados, cada um com “ m_s ” níveis de variação, quando combinados em um esquema fatorial completo resultarão em um total de **J** tratamentos distintos, ou seja:

$$J = \prod_{s=1}^r m_s$$

Em outras palavras, se um dos fatores que compõe o conjunto de tratamentos for, por exemplo, o fator “marca” e dado que esta marca pode ser “conhecida” ou “desconhecida”, então tal fator é composto por dois níveis. Mais especificamente, Aaker, Kumar e Day (2001) definem que os níveis são valores ou alternativas possíveis para quantificar ou qualificar um fator, enquanto que os fatores são as variáveis independentes e manipuláveis pelo pesquisador, que podem ser qualitativas ou quantitativas.

Como um exemplo ilustrativo, suponha que seja de interesse avaliar a influencia de quatro fatores ($r = 4$) na aceitação (ou na intenção de compra) por um tipo de camisa. Admita cada um dos fatores com dois níveis ($m_1 = m_2 = m_3 = m_4 = 2$), conforme o apresentado na Tabela 1.

Tabela 1 - Esquema com Fatores e Níveis do tratamento analisado (camisa)

Nível/Fator	1	2	3	4
	Cor	Ilustração	Preço	Marca
1	Preta	Sim	15,00	Conhecida
2	Branca	Não	20,00	Desconhecida

A combinação dos níveis dos fatores quatro fatores em estudo resultam em $J = 2 \times 2 \times 2 \times 2$ ou $2^4 = 16$ configurações diferentes de camisas, sob um delineamento fatorial completo. Cada tratamento que compõe o conjunto **J** pode ser também denominado como um **estímulo** (MALHOTRA, 2012). Na Tabela 2 encontram-se os **J** = 16 tratamentos que podem ser formados sob o referido delineamento:

Tabela 2 - Tratamentos segundo um delineamento Fatorial Completo

Tratamento	Cor	Ilustração	Preço	Marca
1	Preto	Sim	15,00	Conhecida
2	Branco	Sim	15,00	Conhecida
3	Preto	Não	15,00	Conhecida
4	Branco	Não	15,00	Conhecida
5	Preto	Sim	20,00	Conhecida
6	Branco	Sim	20,00	Conhecida
7	Preto	Não	20,00	Conhecida
8	Branco	Não	20,00	Conhecida
9	Preto	Sim	15,00	Desconhecida
10	Branco	Sim	15,00	Desconhecida
11	Preto	Não	15,00	Desconhecida
12	Branco	Não	15,00	Desconhecida
13	Preto	Sim	20,00	Desconhecida
14	Branco	Sim	20,00	Desconhecida
15	Preto	Não	20,00	Desconhecida
16	Branco	Não	20,00	Desconhecida

Repare que o aumento do número de fatores e ou de seus respectivos níveis¹ ocasiona o aumento do número de tratamentos, sendo este um ponto que se deve atentar

¹ Tradicionalmente, fatores com apenas dois níveis são codificados por meio de variáveis Dummy (0 e 1) e fatores com mais de 2 níveis como 0, 1 e 2 ou 1, 2 e 3 e assim sucessivamente.

ao utilizar a CBCA ou qualquer outra variação da metodologia Conjoint Analysis. Aaker, Kumar e Day (2001) explicaram que o aumento do número de tratamentos torna a avaliação dos consumidores complexa e cansativa, sobrecarregando-os de informações, o que aumenta o risco de que estes ignorem as variações nos fatores menos importantes. Neste sentido, Hair Júnior et al. (1995) sugeriram que a escolha dos fatores e níveis seja realizada com relevância e que as características escolhidas para compor os tratamentos realmente influenciem na opção de escolha do consumidor. A metodologia Focus Group, conforme Deliza (1996) e Souza et al. (2004) é uma boa opção quando se desconhece completamente estes possíveis fatores e níveis. Uma segunda alternativa seria utilizar o método da Soma Constante, conforme detalhado em Dias (2010), para estudos pilotos ou preliminares. Uma terceira abordagem, e mais frequentemente utilizada, seria a escolha de outros delineamentos de tratamentos.

Segundo Silva e Bastos (2010), os estudos de CBCA são multidisciplinares e, portanto, devem ser conduzidos com a participação de um grupo de profissionais experientes e com conhecimentos técnicos e estatísticos, não apenas para a definição correta dos fatores e níveis do estudo, mas também para as realizações das etapas do planejamento experimental, amostragem e análise dos dados, que normalmente ocorrem com a utilização de alguns softwares específicos.

A CBCA vem recebendo grande destaque no atual cenário acadêmico e empresarial, justamente devido ao fato do avaliador informar sua preferência por meio de escolhas, o que é relativamente mais simples e prático, não exigindo a avaliação de cada um dos tratamentos e a atribuição de notas ou ordenação de acordo com a sua preferência. Vale ressaltar que em algumas situações é permitido ao consumidor escolher um ou mais tratamentos ou até mesmo nenhum destes (CARNEIRO et al., 2004). Tal metodologia permite também avaliar a interação entre os fatores em estudo, assim como nas abordagens citadas anteriormente (por notas e ranking). No entanto, tal procedimento não é muito usual, visto que segundo Siqueira (2000), o percentual de explicação da variabilidade total associada às interações é muito pequeno, em cerca de 5 a 10% e, por isso, muitas vezes, elas são negligenciadas.

Os resultados da CBCA são interpretados por meio de probabilidades de escolha de cada tratamento, calculadas após a estimação dos parâmetros do modelo Logit Multinomial. Além disso, diferentemente das abordagens por notas e ranking, não é analisado qual fator é o mais importante (como é feito por meio do índice denominado como Importância Relativa) ou quais níveis dos fatores irão impactar de forma positiva

(ajudar) ou negativa (atrapalhar) na decisão de compra do consumidor². Na referida abordagem, é possível estimar e obter inferências a respeito da significância estatística apenas de cada fator, o que é feito, respectivamente, pelo método da Máxima Verossimilhança ou Máxima Verossimilhança Parcial e pelo teste de Wald, cuja estatística Z, assintoticamente, possui uma distribuição Normal Padrão. Quanto aos níveis dos fatores, sua interpretação é feita por meio de uma taxa denominada como Razão de Escolhas, que permite avaliar o quanto o nível de um fator é mais provável de ser escolhido em relação a outro (s) nível (is) daquele mesmo fator, mantendo-se os demais constantes (DELLA LUCIA, 2008). Moore (2004) afirma que a avaliação por meio de escolhas, embora um pouco menos informativa do que as demais citadas anteriormente, expressa de forma mais realista a aceitação e conseqüentemente o comportamento do consumidor no momento da compra.

Diversos trabalhos associados ao estudo da preferência do consumidor são conduzidos utilizando a metodologia CBCA, seja para avaliar um novo tratamento que se tem interesse em desenvolver (LOUVIERE; WOODWORTH, 1983), para aperfeiçoar/otimizar tratamentos já existentes, adequando-os de acordo com a preferência dos consumidores (GREEN; KRIEGER, 1993) ou ainda para o estabelecimento de novas estratégias competitivas (por exemplo, promoções) que tenham por objetivo atender a alta demanda por um tratamento (MAHAJAN; GREEN; GOLDBERG, 1992). Outras aplicações que envolvem o comportamento de compra do consumidor podem ser verificadas em Lockshin et al. (2006), Enneking et al. (2007) e Tempesta et al. (2010). No Brasil a CBCA ainda é uma metodologia pouco explorada. Deliza et al. (2009) avaliaram a aceitação de consumidores por frutos de mamão que tem seu amadurecimento retardado devido a exposição à radiação. Della Lucia et al. (2010) avaliaram a influência de alguns fatores da embalagem de iogurte na intenção de compra do consumidor.

Como mencionado anteriormente, os softwares são ferramentas de fundamental importância quando se trata de estudos que envolvem a aplicação da metodologia CBCA. Diversos softwares podem ser utilizados, porém as suas necessidades não se justificam apenas para o ajuste de modelos estatísticos, mas também para auxiliar em

² Uma abordagem que permite aumentar o nível de informação na CBCA, embora recente e pouco utilizada, pode ser realizada alterando-se a forma de codificação dos níveis dos fatores. Logo, a análise torna-se similar a RBCA, sendo possível avaliar os níveis mais importantes e estimar Importâncias Relativas por meio das utilidades parciais, trabalhando-se também com as Probabilidades de Escolha. Mais detalhes no trabalho de Júnior et al. (2013).

etapas como a do planejamento experimental do estudo. Os principais softwares estatísticos são o: R, SAS, SPSS e o Sawtooth, este último especificamente desenvolvido para Conjoint Analysis. Alguns softwares matemáticos como o Maple e Matlab também podem ser utilizados. No entanto, exigem maiores conhecimentos de programação matemática. Neste trabalho foram descritos os procedimentos computacionais apenas relacionados aos softwares R (planejamento experimental, abordagem Frequentista e Bayesiana) e SAS (planejamento experimental e abordagem Frequentista). Para maiores detalhes quanto ao uso de softwares como SPSS, Sawtooth e Maple, deve-se verificar, respectivamente, o trabalho de Matos (2011), os diversos technical papers disponíveis na página oficial da Sawtooth³ e em Bastos (2010), que além de detalhar o ajuste do modelo Logit Multinomial pelo Maple, descreveu também os procedimentos para o ajuste no SAS.

³ <http://www.sawtoothsoftware.com/support/technical-papers>

1.2 Delineamentos de Tratamentos

A coleta de dados para CBCA é um procedimento muito importante e que determina diretamente a qualidade dos resultados obtidos. Neste sentido, o condutor do estudo deve conhecer os principais tipos de delineamentos que podem ser aplicados, no intuito de reduzir o número de tratamentos a serem avaliados pelos consumidores. Os tratamentos apresentados na Tabela 2 estão dispostos de acordo com um delineamento fatorial completo, que engloba todas as possíveis combinações entre os níveis dos fatores. Cabe ressaltar que a principal vantagem deste delineamento é sua permissividade em relação à estimação de todos os efeitos principais dos fatores, além das interações de segunda ordem e de ordem superior, de modo que tais termos não sejam correlacionados. Mais especificamente, um delineamento fatorial completo é denominado como ortogonal e balanceado. Entende-se que um delineamento é ortogonal quando todos os efeitos principais são estimados de forma independente dos demais, e por balanceado, quando cada nível ocorre ou aparece com a mesma frequência dentro de cada fator (KUHFIELD et al., 1994). Devido à restrição matemática de ortogonalidade, os coeficientes estimados possuem variância mínima e consequentemente, precisão máxima (KUHFIELD, 1997).

No entanto, se um estudo for composto por $r = 5$ fatores, cada um com $m_s = 2$ níveis, um delineamento fatorial completo resultaria em $\mathbf{J} = 2 \times 2 \times 2 \times 2 \times 2 = 2^5 = 32$ tratamentos, o que mesmo com a praticidade de avaliação oferecida pela CBCA, seria uma tarefa tediosa (para os consumidores) e que poderia gerar altos custos para os aplicadores (com a forma de apresentação dos tratamentos). Uma alternativa para solucionar este problema é a utilização de um delineamento fatorial fracionado, que consiste na apresentação de apenas uma parte da totalidade dos tratamentos aos consumidores, por exemplo, $\frac{1}{4}$ destes tratamentos, o que resultaria em oito tratamentos (SILVA; BASTOS, 2010). O uso deste procedimento implica que alguns efeitos serão confundidos⁴ e possuirão uma estrutura de correlação, não respeitando a restrição de ortogonalidade (KUHFIELD, 1997). Apesar desta desvantagem, o delineamento fatorial fracionado é utilizado com muita frequência em estudos de CBCA.

Kuhfeld (1997) descreve ainda um tipo de delineamento fatorial fracionado especial, denominado como matriz ortogonal ou simplesmente ortogonal. Como o

⁴ Entende-se que dois efeitos serão confundidos quando eles não se distinguem um do outro.

próprio nome sugere, este é um tipo de delineamento fatorial fracionado, em que os efeitos principais são estimados de forma independente um dos outros e, portanto, não são correlacionados. No entanto, tem-se como uma desvantagem que nestes delineamentos não se escolhe o número de tratamentos e a seleção destes é baseada em alguma resolução, como por exemplo, a Resolução III, frequentemente utilizada em estudos de marketing e, conseqüentemente, nos delineamentos ortogonais produzidos por softwares estatísticos quando se trata de Conjoint Analysis. Nesta resolução, todos os efeitos principais são estimados de forma independente dos outros. No entanto, alguns destes são confundidos com as interações de segunda ordem. É importante ressaltar que os delineamentos fatoriais fracionados podem coincidir com os delineamentos ortogonais, em determinadas situações.

A obtenção dos referidos delineamentos fatoriais é realizada segundo algoritmos computacionais disponíveis em softwares estatísticos, que buscam a criação de projetos ou delineamentos eficientes. Alguns algoritmos são mais simples e rápidos, porém, menos confiáveis, como é o caso do algoritmo de Dykstra's (1971). Outros, como os de Mitchell e Miller (1970) e Mitchell (1974) são mais confiáveis que o primeiro. No entanto, exigem maior tempo computacional. Kuhfeld (1997) explica que a confiabilidade de um algoritmo esta associada à sua capacidade de gerar um delineamento ótimo, o que pode ser feito segundo a maximização de um determinado tipo de critério, como por exemplo, o D-efficiency, A-efficiency ou G-efficiency [ver detalhes sobre estes critérios em Kuhfeld (1997)]. Além disso, alguns algoritmos são mais confiáveis que outros, pois permitem a geração de repetidos e diferentes delineamentos e posteriormente, selecionam o mais eficiente. Por isso, exigem um maior tempo computacional. No algoritmo de Dykstra's (1971), apenas um único delineamento é gerado e por isso a baixa confiabilidade em relação aos demais. No entanto, os dois algoritmos mais empregados em CBCA são os de Federov (1972) e o de Cook e Nachtsheim (1980), que é uma modificação do algoritmo de Federov (modified Federov). O primeiro busca um delineamento ótimo realizando todas as possíveis combinações de tratamentos⁵ em pares, adicionando-os e removendo-os, simultaneamente, no intuito de aumentar a eficiência do delineamento. Devido a este tipo de abordagem, tal algoritmo exige maior tempo computacional, porém, é muito eficiente em encontrar delineamentos ótimos. O algoritmo de Federov modificado é similar ao primeiro. No entanto, a modificação proposta consiste na não necessidade de

⁵ No contexto de delineamento de experimentos o termo “tratamento” pode ser encontrado na literatura como um “ponto” que compõe o arranjo experimental.

se realizar todas as possíveis combinações em pares de tratamentos, mas apenas às que irão aumentar a eficiência do delineamento. Os dois últimos algoritmos citados apresentam resultados similares em termos de eficiência, porém, o algoritmo modificado executa o procedimento com aproximadamente metade do tempo gasto pelo primeiro (KUHFIELD, 1997).

Embora não se encontrem descritos neste trabalho, existem diversos outros tipos de delineamentos de tratamentos que podem ser utilizados em estudos de CBCA, dentre os quais destacam-se os não ortogonais e os quase ortogonais (KUHFIELD, 1997 e KUHFIELD et al., 1994).

No software livre R é possível obter delineamentos de tratamentos pelos pacotes **AlgDesign** e **Conjoint**. Além dos manuais disponíveis na rede, maiores detalhes podem ser verificados no artigo de Aizaki e Nishimura (2008). No software SAS, pode-se utilizar o procedimento **PROC OPTEX**. Nos pacotes **AlgDesign** e **Conjoint** do R, o algoritmo utilizado é o de Federov e no procedimento **PROC OPTEX** do SAS, pode-se escolher entre o de Federov e de Federov Modificado, entre outros menos eficientes.

Para finalizar, é citado um trecho, que esclarece a importância de tais delineamentos dentro de um contexto mais amplo e real, respaldando a importância do conhecimento teórico necessário para a correta aplicação da metodologia CBCA:

Kuhfeld (1997),

“A good metaphor for discussing experimental designs is a raft. A raft is a flat boat that you hope will support your weight and keep you from getting wet. An experimental design forms the basis of a conjoint study, and you hope it will provide you with good information to support your marketing decisions. If your raft is not properly constructed, you will fall in the water and get eaten by alligators. If your experimental design is nonoptimal, you will have less information to use to make important decisions, and if your decisions are wrong, you will be eaten alive by your competitors.”

1.3 Formas de Apresentação dos Tratamentos

A etapa final para aplicação da CBCA consiste em definir como cada tratamento será apresentado aos consumidores, para que a avaliação seja realizada. Neste contexto, três pontos devem ser entendidos: como os estímulos serão apresentados, a ordem de apresentação e o número de consumidores (tamanho da amostra).

Segundo Silva e Bastos (2010), o aplicador do estudo deve escolher a forma mais adequada para apresentar os tratamentos aos consumidores, seja por meio de instruções, questionários, objetos reais, bens de consumo, protótipos de produtos ou fotografias. Além disso, a avaliação deverá ocorrer em um local que seja tranquilo e confortável, com um intervalo de tempo adequado entre a apresentação de cada tratamento aos consumidores.

A ordem de apresentação dos tratamentos é de extrema importância, visto que existem implícitos os efeitos da ordem de apresentação e da influência de uma amostra na avaliação subsequente (DELLA LÚCIA et al., 2007). Em outras palavras, pode-se pensar, por exemplo, que se os consumidores avaliarem inicialmente o melhor tratamento, conseqüentemente o irão escolher ou atribuir a maior nota, desfavorecendo os demais. Este comportamento está associado à “psicologia do consumidor” e conseqüentemente com a expectativa do mesmo em relação aos tratamentos. É esta expectativa que influencia para que um consumidor melhore ou piore sua percepção em relação a um dado tratamento, sem mesmo a este ter sido apresentado (tocado, degustado etc.). Segundo Deliza (1996), “uma alta expectativa geralmente leva à escolha do produto, enquanto uma baixa expectativa certamente produz a rejeição do mesmo”.

Por este motivo, uma alternativa adotada para minimizar estes problemas, é a apresentação dos tratamentos conforme os delineamentos experimentais (não confundir com delineamentos de tratamentos) propostos por MacFIE et al. (1989). Nestes delineamentos, a ordem de apresentação de cada tratamento é diferente para cada consumidor. Além disso, cada tratamento aparece precedido e sucedido de todos os outros tratamentos o mesmo número de vezes. Este procedimento garante a eliminação do efeito da ordem de apresentação e do efeito residual caracterizado pela influência de uma amostra na avaliação da subsequente (DELLA LÚCIA et al., 2007).

O número de consumidores deve ser suficiente no intuito de permitir que os efeitos principais dos fatores em estudo possam ser estimados de modo não-viesado ou não tendencioso. Os delineamentos experimentais propostos por MacFIE et al. (1989) oferecem esta informação, ou seja, o número mínimo de consumidores que são necessários para se avaliar um total de J tratamentos. No artigo original é possível verificar os delineamentos para até $J = 16$ tratamentos. Por isso essa vem sendo uma prática muito comum em estudos de CBCA. Caso haja interesse do pesquisador em aumentar o tamanho da amostra, além do sugerido, é possível estabelecer repetições para o delineamento.

Capítulo 2

Modelo Logit Multinomial

2.1 Fundamentos Teóricos Introdutórios

Nos estudos em que a metodologia CBCA é empregada, o modelo estatístico comumente utilizado é o Logit Multinomial (Multinomial Logit Model). Este é um dos mais populares modelos probabilísticos que compõe a classe dos modelos não ordenados⁶ de escolha discreta (discrete choice models) (LOUVIERE, 1984 e GOLDSTEIN, 1997). Seu desenvolvimento ocorreu entre o final da década de 1960 e o início da década de 1970, com os estudos de Theil (1969). O referido autor foi o responsável pela generalização do modelo de escolha binária ou Logit Binomial, que é caracterizado por apresentar variável dependente de natureza qualitativa e não mensurável, com apenas duas opções de escolha, em modelos de escolha múltipla ou Logit Multinomial, que é similar ao primeiro, porém, a variável dependente exibe mais de duas opções de escolhas. No entanto, foi McFadden (1974) que atribuiu a devida consistência teórica ao modelo Logit Multinomial, demonstrando suas propriedades, limitações e popularizando sua aplicação, principalmente no setor de transportes, pesquisas de marketing e estudos econométricos.

Segundo Greene (2003), um dos motivos que contribuíram para a popularização do modelo Logit Multinomial e, conseqüentemente, para a sua ampla utilização em diversas áreas do conhecimento, é a menor complexidade de estimação de seus parâmetros, devido, principalmente, ao uso da distribuição Gumbel em sua formulação.

Os modelos probabilísticos de escolha discreta são baseados na teoria econométrica de utilidade aleatória, que consiste em uma função para a utilidade de cada alternativa (tratamento), conforme em (1). A definição de utilidade pode ser entendida como o benefício ou satisfação percebida por um consumidor em relação a um determinado tratamento (VARIAN; REPCHECK, 2010).

⁶ Um modelo de escolha discreta é não ordenado quando o entrevistado efetua uma única escolha dentre um conjunto de alternativas e ordenado quando o mesmo efetua uma ordenação total ou parcial do conjunto de escolhas (BASTOS, 2010).

$$U_{nj}^* = \mathbf{X}_j \boldsymbol{\beta}^* + \varepsilon_{nj}^* \quad (1)$$

- U_{nj}^* é a utilidade atribuída pelo n-ésimo consumidor ao j-ésimo tratamento;
- $\mathbf{X}_j = [x_1^j, x_2^j, \dots, x_r^j]'$ é o vetor 1 x r com a codificação dos níveis dos r fatores presentes no tratamento j, onde cada x_i^j representa o nível do fator i presente no tratamento j, em que $i = 1, 2, \dots, r$ e $j \in \{1, 2, \dots, J\}$.
- $\boldsymbol{\beta}^* = [\beta_1^*, \beta_2^*, \dots, \beta_r^*]$ é o vetor de parâmetros r x 1 a ser estimado. Destaca-se que na metodologia CBCA o número de parâmetros é igual ao número de fatores em estudo;
- ε_{nj}^* é o erro aleatório não observável associado à utilidade atribuída pelo n-ésimo consumidor ao j-ésimo tratamento.

Para o desenvolvimento do modelo Logit Multinomial, segundo Bem-Akiva e Lerman (1985), o termo aleatório ε_{nj}^* assume uma distribuição de Valores Extremos Tipo I (Extreme Value Type I) ou Gumbel⁷. Em (2) e (3) tem-se, respectivamente, a função densidade de probabilidade (f.d.p) e a função de distribuição acumulada (f.d.a) de tal distribuição:

$$f(\varepsilon_{nj}^*) = \frac{1}{\sigma} \left[e^{\left(\frac{-\varepsilon_{nj}^* - \mu}{\sigma}\right)} \right] \left[e^{-e^{\left(\frac{\varepsilon_{nj}^* - \mu}{\sigma}\right)}} \right] \quad (2)$$

$$F(\varepsilon_{nj}^*) = e^{-e^{\left(\frac{\varepsilon_{nj}^* - \mu}{\sigma}\right)}} \quad (3)$$

O modelo conhecido como Logit Multinomial Padrão (ou Standard Multinomial Logit), referido apenas como Logit Multinomial, é obtido assumindo-se que o componente aleatório possui distribuição Gumbel com parâmetros $\mu = 0$ e $\sigma^2 = \frac{\pi^2}{6}$. A escolha do valor $\frac{\pi^2}{6}$ para a variância do termo aleatório é por conveniência e segundo Train (2002), de forma implícita, consiste em normalizar a escala da utilidade, procedimento que é feito considerando-se $VAR(\varepsilon_{nj}^*) = \sigma^2 \left(\frac{\pi^2}{6}\right)$ e dividindo os termos de (1) por σ :

$$\frac{U_{nj}^*}{\sigma} = \frac{\mathbf{X}_j \boldsymbol{\beta}^*}{\sigma} + \frac{\varepsilon_{nj}^*}{\sigma} \Rightarrow U_{nj} = \mathbf{X}_j \left(\frac{\boldsymbol{\beta}^*}{\sigma}\right) + \varepsilon_{nj} \Rightarrow U_{nj} = \mathbf{X}_j \boldsymbol{\beta} + \varepsilon_{nj}$$

⁷ Uma variável aleatória com distribuição Gumbel pode ser obtida a partir do logaritmo natural de uma variável aleatória com distribuição Weibull.

Logo, a variância do componente aleatório é obtida como:

$$VAR(\varepsilon_{nj}) = VAR\left(\frac{\varepsilon_{nj}^*}{\sigma}\right) = \frac{1}{\sigma^2} VAR(\varepsilon_{nj}^*) = \frac{1}{\sigma^2} \sigma^2 \left(\frac{\pi^2}{6}\right) = \frac{\pi^2}{6}$$

Portanto, em (4) e (5), tem-se respectivamente, a f.d.p. e f.d.a. reescritas, sendo que a f.d.a. em (5) é uma forma fechada⁸, que como mencionado anteriormente, retrata a principal vantagem da utilização do modelo Logit Multinomial:

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \quad (4)$$

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}} \quad (5)$$

Ao realizar tal procedimento e normalizar a escala da utilidade, os parâmetros estimados referem-se na verdade aos parâmetros originais divididos pelo parâmetro de escala, que é o desvio padrão da distribuição do componente aleatório. Devido a esta padronização, a interpretação de tais coeficientes está associada ao impacto de cada variável observada em relação aos fatores não observados no estudo. Conforme Train (2002):

“A larger variance in unobserved factors leads to lower coefficients, even if the observed factors have the same impact on utility (i.e., higher σ means lower β even if β^* is the same.)”

É importante destacar que o erro aleatório é assumido como independente e identicamente distribuído (i.i.d.) em relação às alternativas de escolha. Esta pressuposição é fundamental para o cumprimento da propriedade citada na literatura como Independence of Irrelevant Alternatives (ou apenas IIA) (JOHNSON; KOTZ, 1970 e TRAIN, 2002), uma característica específica do modelo Logit Multinomial.

Cabe ressaltar que em meados da década de 1950, quando surgiram as primeiras aplicações práticas no campo da econometria, os principais modelos de escolha discreta utilizados eram o Logit Binomial e o Probit Multinomial. Para o segundo, admitia-se que o componente aleatório da função utilidade possuía distribuição normal (gaussiana), sendo esta a principal vantagem de sua utilização em relação ao Logit Binomial, que exige a manipulação da distribuição acumulada logística, o que naquela época ainda era uma tarefa complexa e com menor confiabilidade, devido à inconsistência teórica a respeito desta distribuição (SILVA, 2004).

⁸ Diz-se que uma distribuição possui forma fechada quando as integrais que envolvem a obtenção de seus principais momentos (por exemplo, média e variância) possuem solução analítica.

No entanto, em pesquisas de marketing, mesmo com a flexibilidade da distribuição normal, por permitir a existência de uma estrutura de autocorrelação entre os erros aleatórios, o uso do modelo Probit Multinomial apresenta algumas desvantagens quanto à estimação de seus parâmetros (complexidade e alto custo computacional), principalmente se número de alternativas do conjunto de escolhas for elevado. A dificuldade de se trabalhar com este modelo está associada à necessidade de se avaliar integrais múltiplas da distribuição normal, o que implica em algumas limitações, visto que tal distribuição não possui uma forma fechada⁹ (CHANDUKAL et al., 2007). Neste sentido, devido às vantagens do modelo Logit Multinomial, os pesquisadores priorizaram sua utilização em relação aos demais modelos concorrentes.

⁹ A escolha da distribuição normal para os termos aleatórios da função utilidade torna a integral que permite calcular as probabilidades de escolha para o modelo Probit Multinomial complexa e, portanto, sem solução analítica, possível de solução apenas por métodos numéricos ou simulação (TRAIN, 2002).

2.2 Desenvolvimento do Modelo

No processo de modelagem das escolhas, de acordo com a teoria da utilidade aleatória, considera-se que em um conjunto com \mathbf{J} alternativas (tratamentos), a j -ésima alternativa será escolhida pelo n -ésimo consumidor, se e somente se:

$$U_{nj} > U_{nk}, \forall j \neq k \text{ e } j, k \in \{1, 2, \dots, J\} \quad (6)$$

Considere agora a probabilidade de que o n -ésimo consumidor escolha o j -ésimo tratamento, dado o conjunto de tratamentos \mathbf{X} , $P(Y_n = j|\mathbf{X})$. É comum na literatura encontrar o termo market share ou share of preference para representar tais probabilidades, cujo significado refere-se à participação ou quota de mercado associada a cada um dos j tratamentos. Escrevendo (6) em termos probabilísticos, tem-se:

$$P(Y_n = j|\mathbf{X}) = P(U_{nj} > U_{nk})$$

$$P(Y_n = j|\mathbf{X}) = P(\mathbf{X}_j\boldsymbol{\beta} + \varepsilon_{nj} > \mathbf{X}_k\boldsymbol{\beta} + \varepsilon_{nk})$$

$$P(Y_n = j|\mathbf{X}) = P(\varepsilon_{nk} < \varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})$$

$$P(Y_n = j|\mathbf{X}) = P(\varepsilon_{nk} < \Delta_{jk})$$

Em que $\Delta_{jk} = \varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta}$. Expandindo k para cada um dos j tratamentos e n consumidores, têm-se:

$$P(Y_n = j|\mathbf{X}) = P(\varepsilon_{n1} \leq \Delta_{j1}; \varepsilon_{n2} \leq \Delta_{j2}; \dots; \varepsilon_{nJ} \leq \Delta_{jJ}) \forall k \neq j$$

O que diferencia os modelos de escolha discreta é a especificação da distribuição de probabilidades para o componente aleatório da função utilidade. Logo, diferentes distribuições conduzem a diferentes modelos. No desenvolvimento do modelo Logit Multinomial, os termos aleatórios assumem distribuição Gumbel e são independentes e identicamente distribuídos. Portanto:

$$P(Y_n = j|\mathbf{X}) = P(\varepsilon_{n1} \leq \Delta_{j1}) P(\varepsilon_{n2} \leq \Delta_{j2}) \dots P(\varepsilon_{nJ} \leq \Delta_{jJ})$$

$$P(Y_n = j|\mathbf{X}) = \prod_{\substack{k=1 \\ k \neq j}}^J P(\varepsilon_{nk} \leq \Delta_{jk})$$

$$P(Y_n = j|\mathbf{X}) = \prod_{\substack{k=1 \\ k \neq j}}^J P(\varepsilon_{nk} \leq \varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})$$

$$^{10}P(Y_n = j|\mathbf{X}) = E_{\varepsilon_{nj}} \left\{ \prod_{\substack{k=1 \\ k \neq j}}^J P[\varepsilon_{nk} \leq (\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta}) | \varepsilon_{nj}] \right\}$$

$$P(Y_n = j|\mathbf{X}) = E_{\varepsilon_{nj}} \left[\prod_{\substack{k=1 \\ k \neq j}}^J F_{\varepsilon_{nk}}(\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta}) \right]$$

$$P(Y_n = j|\mathbf{X}) = \int_{\Omega_{\varepsilon_{nj}}} f(\varepsilon_{nj}) \prod_{\substack{k=1 \\ k \neq j}}^J F_{\varepsilon_{nk}}(\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta}) d\varepsilon_{nj}$$

Se $\varepsilon_{nj} \sim \text{Gumbel}\left(0, \frac{\pi^2}{6}\right)$ então o suporte de $\varepsilon_{nj} \in (-\infty, \infty)$, logo:

$$P(Y_n = j|\mathbf{X}) = \int_{-\infty}^{\infty} f(\varepsilon_{nj}) \prod_{\substack{k=1 \\ k \neq j}}^J F_{\varepsilon_{nk}}(\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta}) d\varepsilon_{nj}$$

Recorrendo a expressão (4), que é a f.d.p. da distribuição Gumbel:

$$P(Y_n = j|\mathbf{X}) = \int_{-\infty}^{\infty} e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \prod_{\substack{k=1 \\ k \neq j}}^J F_{\varepsilon_{nk}}(\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta}) d\varepsilon_{nj}$$

Analisando a função acumulada da distribuição Gumbel (5) verifica-se que o segundo termo da f.d.p. (4) pode ser escrito como a função acumulada quando $k = j$, ou seja:

$$e^{-e^{-\varepsilon_{nj}}} = P(\varepsilon_{nk} \leq \varepsilon_{nj}) = F_{\varepsilon_{nk}}(\varepsilon_{nj})$$

$$P(Y_n = j|\mathbf{X}) = \int_{-\infty}^{\infty} e^{-\varepsilon_{nj}} F_{\varepsilon_{nk}}(\varepsilon_{nj}) \prod_{\substack{k=1 \\ k \neq j}}^J F_{\varepsilon_{nk}}(\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta}) d\varepsilon_{nj}$$

¹⁰ Para W e L, duas variáveis aleatórias, tem-se a seguinte propriedade estatística:

$$P(W \leq L) = E_L[P_W(W \leq L|L)]$$

$$P_W(W \leq L) = \int_{\Omega_T} f_L(t)P_W(W \leq t)dt$$

$$P_W(W \leq L) = \int_{\Omega_T} f_L(t)F_W(t)dt$$

Devido à independência dos erros aleatórios, este termo pode ser incorporado dentro da função produtório. Isso implica que a restrição de $k \neq j$ não mais existirá:

$$P(Y_n = j|\mathbf{X}) = \int_{-\infty}^{\infty} e^{-\varepsilon_{nj}} \prod_{k=1}^J F_{\varepsilon_{nk}}(\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta}) d\varepsilon_{nj}$$

$$P(Y_n = j|\mathbf{X}) = \int_{-\infty}^{\infty} e^{-\varepsilon_{nj}} \prod_{k=1}^J e^{-e^{-(\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}} d\varepsilon_{nj}$$

$$P(Y_n = j|\mathbf{X}) = \int_{-\infty}^{\infty} e^{-\varepsilon_{nj}} e^{[-\sum_{k=1}^J e^{-(\varepsilon_{nj} + \mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}]} d\varepsilon_{nj}$$

$$P(Y_n = j|\mathbf{X}) = \int_{-\infty}^{\infty} e^{-\varepsilon_{nj}} e^{[-\sum_{k=1}^J e^{-\varepsilon_{nj}} e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}]} d\varepsilon_{nj}$$

$$P(Y_n = j|\mathbf{X}) = \int_{-\infty}^{\infty} e^{-\varepsilon_{nj}} e^{[-e^{-\varepsilon_{nj}} \sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}]} d\varepsilon_{nj}$$

Para resolver a integral, basta proceder à substituição $t = e^{-\varepsilon_{nj}}$. Consequentemente, derivando em relação à ε_{nj} , tem-se que $dt = -e^{-\varepsilon_{nj}} d\varepsilon_{nj}$. Automaticamente, quando $\varepsilon_{nj} \rightarrow -\infty \Rightarrow t \rightarrow \infty$ e $\varepsilon_{nj} \rightarrow \infty \Rightarrow t \rightarrow 0$:

$$P(Y_n = j|\mathbf{X}) = - \int_{\infty}^0 e^{[-t \sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}]} dt$$

$$P(Y_n = j|\mathbf{X}) = \int_0^{\infty} e^{[-t \sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}]} dt$$

$$P(Y_n = j|\mathbf{X}) = - \frac{e^{[-t \sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}]}{\sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}}$$

Substituindo os limites da integral:

$$P(Y_n = j|\mathbf{X}) = - \left\{ \frac{e^{[-(\infty) \sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}]}{\sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}} - \frac{e^{[-(0) \sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}]}{\sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}} \right\}$$

$$P(Y_n = j|\mathbf{X}) = - \left\{ \frac{e^{[-\infty]}}{\sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}} - \frac{e^{[0]}}{\sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}} \right\}$$

$$P(Y_n = j|\mathbf{X}) = - \left\{ 0 - \frac{1}{\sum_{k=1}^J e^{-(\mathbf{X}_j\boldsymbol{\beta} - \mathbf{X}_k\boldsymbol{\beta})}} \right\}$$

$$P(Y_n = j|\mathbf{X}) = \frac{1}{\sum_{k=1}^J e^{-(x_j\beta - x_k\beta)}}$$

$$P(Y_n = j|\mathbf{X}) = \frac{1}{\sum_{k=1}^J e^{-x_j\beta} e^{x_k\beta}}$$

$$P(Y_n = j|\mathbf{X}) = \frac{1}{e^{-x_j\beta} \sum_{k=1}^J e^{x_k\beta}}$$

$$P(Y_n = j|\mathbf{X}) = \frac{e^{x_j\beta}}{\sum_{k=1}^J e^{x_k\beta}} \tag{7}$$

2.3 Aplicações Práticas

O modelo Logit Multinomial (7) permite calcular a probabilidade de que o n -ésimo consumidor escolha o j -ésimo tratamento em função das características que compõem cada um dos \mathbf{J} tratamentos, ou seja, condicionadas ao conjunto de fatores \mathbf{X} [uma matriz com dimensão $(n \times \mathbf{J}) \times r$]. A análise desenvolvida por (7) é também conhecida como “Agregada” devido à estimação de um único vetor de parâmetros e , conseqüentemente, da obtenção de probabilidades que representam a preferência média ou geral da amostra em estudo, ou seja, considera-se que esta é homogênea em termos dos gostos e preferências.

As probabilidades de escolha estimadas com o modelo Logit Multinomial podem também ser obtidas por meio de covariáveis específicas, referentes a cada indivíduo e comumente denominadas como variáveis demográficas (demographics variables). Como por exemplo: renda, idade, sexo, nível de escolaridade etc. Neste caso a análise não é mais agregada, pois estima-se um vetor de parâmetros para cada indivíduo e conseqüentemente probabilidades de escolhas a nível individual (o que é conhecido como uma análise desagregada). Embora ainda não haja um consenso na literatura, em alguns artigos científicos e softwares estatísticos (como o R e o SAS) cita-se uma distinção entre os modelos como sendo do tipo **Logit Multinomial** (Multinomial Logit Model) ou do tipo **Logit Condicional** (Conditional Logit Model), onde a primeira refere-se à utilização do modelo (7) com as covariáveis sendo variáveis demográficas (análise desagregada) e a segunda, quando as covariáveis são expressas pelas combinações dos níveis de cada tratamento (análise agregada), que é o objetivo deste trabalho. Uma revisão sobre este assunto, que explica as principais diferenças entre estes dois modelos, pode ser encontrada em Hoffman e Duncan (1988).

2.4 Estimação dos Parâmetros

Considere uma amostra aleatória que contem as escolhas de n consumidores em relação a um conjunto de J tratamentos, definida como $Y = (Y_1 = i, Y_2 = j, \dots, Y_n = k) \{i, j, \dots, k \in 1, 2, \dots, J\}$. Logo, a função de verossimilhança, que é a probabilidade conjunta desta amostra, sobre a pressuposição de independência entre as escolhas dos indivíduos, é obtida como:

$$L(\beta) = \prod_{n=1}^N \left(\frac{e^{X_{ny_n} \beta}}{\sum_{k=1}^J e^{X_{nk} \beta}} \right)$$

Train (2002) explica que para a aplicação do Modelo Logit Multinomial, três pressuposições fundamentais sobre o conjunto de tratamentos precisam ser atendidas; sendo que estes devem ser: i) mutuamente exclusivos, ii) exaustivos e iii) finitos. Os tratamentos são mutuamente exclusivos quando a escolha de um dos J tratamentos implicar que os demais $J - 1$ não poderão ser escolhidos pelo mesmo consumidor. O conjunto de tratamentos é exaustivo se pelo menos um tratamento é escolhido dentre o conjunto finito de J tratamentos. No entanto, conforme mencionado anteriormente, na CBCA um consumidor pode escolher mais de um tratamento ou não escolher nenhum deles. Para solucionar o problema de dois tratamentos A e B que não são mutuamente exclusivos, pode-se estabelecer uma nova alternativa de escolha do tipo “Ambos A e B” e para o problema de tratamentos exaustivos, uma opção de escolha do tipo “nenhum dos tratamentos apresentados”. Note que o problema de tratamentos exaustivos é facilmente resolvido. No entanto, para o primeiro (tratamentos mutuamente exclusivos), conforme o número de tratamentos aumenta, torna-se necessário estabelecer muitas opções com as combinações entre eles, o que impossibilita o uso dessa abordagem.

Neste sentido, So e Kuhfeld (2010) apresentam uma solução simples e prática para este problema, que consiste em utilizar conceitos e métodos estatísticos de estimação empregados em análise de sobrevivência. Por isso, a estimação dos parâmetros do modelo Logit Multinomial em softwares estatísticos como o R e o SAS ocorre via funções ou procedimentos aplicados à análise de sobrevivência. No R, estima-se β por meio da função `clogit` (Conditional logistic regression) do pacote `survival` e no SAS pelo procedimento `PROC PHREG` (Proportional Hazards Regression). Segundo So e Kuhfeld (2010), isso pode ser explicado devido à função de verossimilhança do modelo Logit Multinomial ter aproximadamente a mesma forma da

função de verossimilhança parcial de um modelo aplicado em análise de sobrevivência, que é o modelo de Cox. Neste contexto, faz-se a seguinte analogia: a escolha de um determinado tratamento é codificada como uma falha (que recebe o valor 1) e as alternativas não escolhidas são consideradas como não observadas (censuras). Portanto, censuras recebem o valor 0. Logo, se mais de um tratamento é escolhido por um mesmo consumidor, repete-se a codificação de escolha (1) para os tratamentos selecionados. O mesmo ocorre se nenhum tratamento é escolhido, codificando todos os tratamentos com o valor zero (0).

Normalmente obtêm-se $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_r)'$ por Máxima Verossimilhança ¹¹ (TRAIN, 2002). No entanto, com o uso dos procedimentos de análise de sobrevivência, o método empregado é o de Máxima Verossimilhança Parcial, que encontra os estimadores maximizando a função de verossimilhança parcial apresentada em (8) (COLOSIMO; GIOLO, 2006). Admita I_{nj} como a variável indicadora de escolha, onde $I_{nj} = 1$ caso o j -ésimo tratamento tenha sido escolhido pelo n -ésimo consumidor e $I_{nj} = 0$ caso contrário.

$$\begin{aligned}
 L_n(\boldsymbol{\beta}) &= \prod_{j=1}^J P(Y_n = j | \mathbf{X})^{I_{nj}} \\
 L(\boldsymbol{\beta}) &= \prod_{n=1}^N L_n(\boldsymbol{\beta}) \\
 L(\boldsymbol{\beta}) &= \prod_{n=1}^N \prod_{j=1}^J \left(\frac{e^{X_j \boldsymbol{\beta}}}{\sum_{k=1}^J e^{X_k \boldsymbol{\beta}}} \right)^{I_{nj}} \tag{8}
 \end{aligned}$$

No contexto da análise de sobrevivência, o método da Máxima Verossimilhança Parcial é utilizado para separar a componente não paramétrica da componente paramétrica do modelo de Cox, visto que é a presença da primeira inviabiliza a aplicação, por exemplo, do método tradicional da Máxima Verossimilhança para a estimação do vetor de parâmetros $\boldsymbol{\beta}$.

¹¹ As Estimativas de Máxima Verossimilhança para os parâmetros do Modelo Logit Multinomial podem ser obtidas pelo pacote e função **mlogit** do software livre R. Retirando o termo intercepto da análise (ver detalhes no Help do pacote) os EMV irão convergir para os Estimadores de Máxima Verossimilhança Parcial APENAS na situação padrão, ou seja, em que cada consumidor escolhe apenas um único tratamento.

Destaca-se que quando um consumidor escolhe mais de um tratamento, deve-se empregar uma aproximação para a função de Verossimilhança Parcial, capaz de suportar ou indicar o número de escolhas feitas por este indivíduo, o que no contexto da análise de sobrevivência é conhecido como um empate. Em geral, a aproximação de Breslow para a função de Verossimilhança Parcial é a mais indicada, pois com uma pequena modificação¹² (a inclusão do termo d_{nj}), permite indicar o número de escolhas feitas (ou falhas) por cada consumidor. Naturalmente, tal expressão se reduz à forma em (8) quando não ocorrer empates. A aproximação também é válida quando nenhum tratamento é escolhido. Existem outros tipos de aproximações, tais como a Exata (Exact) e a de Efron. No entanto, conforme Kuhfeld (2010), a aproximação de Breslow é a mais apropriada para representar o modelo Logit Multinomial e ainda possui melhor eficiência computacional no contexto da CBCA. Maiores detalhes sobre as aproximações citadas podem ser visualizados em Kalbfleisch e Prentice (1980), para a aproximação exata, e Efron (1977) para a aproximação de Efron. Em Colosimo e Giolo (2006) é possível encontrar detalhes sobre o procedimento de estimação e também sobre estas aproximações e suas principais diferenças.

No contexto matemático, dado a situação básica da CBCA, em que cada consumidor escolhe um único tratamento, para a estimação do vetor de parâmetros, o procedimento inicial consiste em obter o logaritmo da função de verossimilhança $l(\boldsymbol{\beta})$. Maximizar tal função não implica em perda de generalidades, como segue:

$$\begin{aligned}
 l(\boldsymbol{\beta}) &= \ln[L(\boldsymbol{\beta})] = \ln \left[\prod_{n=1}^N \prod_{j=1}^J \left(\frac{e^{X_j \boldsymbol{\beta}}}{\sum_{k=1}^J e^{X_k \boldsymbol{\beta}}} \right)^{I_{nj}} \right] \\
 l(\boldsymbol{\beta}) &= \sum_{n=1}^N \sum_{j=1}^J I_{nj} \ln \left(\frac{e^{X_j \boldsymbol{\beta}}}{\sum_{k=1}^J e^{X_k \boldsymbol{\beta}}} \right) \\
 l(\boldsymbol{\beta}) &= \sum_{n=1}^N \sum_{j=1}^J I_{nj} \ln(e^{X_j \boldsymbol{\beta}}) - \sum_{n=1}^N \sum_{j=1}^J I_{nj} \ln \left(\sum_{k=1}^J e^{X_k \boldsymbol{\beta}} \right) \\
 l(\boldsymbol{\beta}) &= \sum_{n=1}^N \sum_{j=1}^J I_{nj} X_j \boldsymbol{\beta} - \sum_{n=1}^N \sum_{j=1}^J I_{nj} \ln \left(\sum_{k=1}^J e^{X_k \boldsymbol{\beta}} \right)
 \end{aligned}$$

¹² $L(\boldsymbol{\beta}) = \prod_{n=1}^N \prod_{j=1}^J \left\{ \frac{e^{X_j \boldsymbol{\beta}}}{\left[\sum_{k=1}^J e^{X_k \boldsymbol{\beta}} \right]^{d_{nj}}} \right\}^{I_{nj}}$

Feito isso, para obter $\hat{\boldsymbol{\beta}}$ deve-se encontrar a derivada primeira de $l(\boldsymbol{\beta})$ em relação ao vetor de parâmetros $\boldsymbol{\beta}$ e igualá-la a zero:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} [l(\boldsymbol{\beta})] &= \sum_{n=1}^N \sum_{j=1}^J I_{n_j} \mathbf{X}_j - \sum_{n=1}^N \sum_{j=1}^J n_j \frac{\sum_{k=1}^J e^{X_k \boldsymbol{\beta}} \mathbf{X}_k}{\sum_{k=1}^J e^{X_k \boldsymbol{\beta}}} \\ \frac{\partial}{\partial \boldsymbol{\beta}} [l(\boldsymbol{\beta})] &= 0 \\ \sum_{n=1}^N \sum_{j=1}^J I_{n_j} \mathbf{X}_j &= \sum_{n=1}^N \sum_{j=1}^J I_{n_j} \frac{\sum_{k=1}^J e^{X_k \boldsymbol{\beta}} \mathbf{X}_k}{\sum_{k=1}^J e^{X_k \boldsymbol{\beta}}} \end{aligned} \quad (9)$$

Como a derivada de $l(\boldsymbol{\beta})$ apresenta-se como uma função do vetor de parâmetros, isso implica que $L(\boldsymbol{\beta})$ é não linear em $\boldsymbol{\beta}$. Portanto, não é possível obter solução analítica em (9) e para encontrar os valores de $\hat{\boldsymbol{\beta}}$ que maximizem a função de verossimilhança, torna-se necessário a utilização de métodos iterativos, como por exemplo, o método de Newton-Raphson.

A execução do método de Newton-Raphson pode ser realizada em softwares matemáticos ou estatísticos. Entretanto, uma exigência é que se determine os valores iniciais $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \dots, \beta_{0r})$ para o vetor de parâmetros. Um importante detalhe é que a convergência do método está diretamente associada à escolha destes valores, ou seja, valores iniciais mais próximos dos reais implicam em menor tempo computacional e maior probabilidade de convergência (GALLANT, 1984). Define-se $\mathbf{g}_{\boldsymbol{\beta}_0}$, o vetor gradiente ou vetor score, que contém as derivadas parciais da função de verossimilhança em relação a cada parâmetro, avaliadas em $\boldsymbol{\beta}_0$. Denomina-se $\mathbf{H}_{\boldsymbol{\beta}_0}$, a matriz Hessiana ou a matriz de derivadas parciais de segunda ordem da função de verossimilhança, em relação a cada parâmetro e também avaliada em $\boldsymbol{\beta}_0$:

$$\mathbf{g}_{\boldsymbol{\beta}_0} = \begin{bmatrix} g_{10} \\ g_{20} \\ \dots \\ g_{r0} \end{bmatrix} \text{ onde } g_{i0} = \left(\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_i} \Big|_{\boldsymbol{\beta}_0} \right)$$

$$\mathbf{H}_{\boldsymbol{\beta}_0} = \frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}} = \left(\begin{array}{c} \left[\begin{array}{ccc} \frac{\partial g_{10}}{\partial \beta_1} & \frac{\partial g_{10}}{\partial \beta_2} & \dots & \frac{\partial g_{10}}{\partial \beta_r} \\ \frac{\partial g_{20}}{\partial \beta_1} & \frac{\partial g_{20}}{\partial \beta_2} & \dots & \frac{\partial g_{20}}{\partial \beta_r} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_{r0}}{\partial \beta_1} & \frac{\partial g_{r0}}{\partial \beta_2} & \dots & \frac{\partial g_{r0}}{\partial \beta_r} \end{array} \right] \Big|_{\boldsymbol{\beta}_0} \end{array} \right)$$

O método de Newton-Raphson consiste essencialmente em uma aproximação para $\mathbf{g}_{\hat{\beta}}$ por meio de uma série de Taylor de primeira ordem, que igualada à zero, fornece o seguinte esquema iterativo:

$$\hat{\beta}_1 = \beta_0 + (-H_{\beta_0}^{-1}) \mathbf{g}_{\beta_0}$$

$$\hat{\beta}_{m+1} = \beta_m + (-H_{\beta_m}^{-1}) \mathbf{g}_{\beta_m}$$

Como todo método iterativo, para obter o valor final é necessário estabelecer a priori um número máximo de iterações ou um critério de paragem. Entende-se como critério de paragem uma medida que se refere ao erro ou à tolerância máxima admissível entre o valor de $\hat{\beta}$ obtido na iteração $m + 1$ e na iteração m . Se tal diferença for inferior ao valor estabelecido (erro), o processo continua. Caso contrário, repete-se novamente o passo m , porém, tomando-se como aproximação inicial o valor obtido no passo $m + 1$. Ver possíveis critérios em Bastos (2010).

Uma observação feita por McFadden (1974), é que a função do log da verossimilhança do modelo Logit Multinomial é globalmente côncava, o que implica na existência de apenas um único ponto de máximo (máximo relativo = máximo global). Portanto, garante-se uma rápida e eficiente otimização computacional.

Note ainda que o algoritmo de Newton-Raphson trabalha com o negativo da Matriz Hessiana $(-H_{\beta_0})$, pois desta forma tem-se uma matriz positiva definida¹³ e, conseqüentemente, o mesmo ocorre para sua inversa $(-H_{\beta_0}^{-1})$. Tal fato garante que o log da função de verossimilhança aumente a cada iteração, conduzindo o processo de estimação para o único ponto de máximo global. Se a matriz Hessiana avaliada não for positiva definida, essa condição não será assegurada (TRAIN, 2002).

¹³ Por definição uma matriz simétrica M é positiva definida se $x^T M x > 0$ para qualquer $x \neq 0$.

2.5 Razão de Escolhas

A Razão de Escolhas (R.E.) fornece um resultado inferencial muito utilizado em estudos de CBCA, pois permite comparar a probabilidade de escolha entre dois tratamentos, por exemplo, \mathbf{p} e \mathbf{q} , onde o tratamento \mathbf{q} é obtido do tratamento \mathbf{p} quando são fixados $r - 1$ níveis dos r fatores que compõem a configuração do primeiro, ou seja, alterando-se apenas o nível de um fator (SILVA; BASTOS, 2010). Especificamente, no numerador da Razão de Escolhas tem-se a probabilidade de escolha do tratamento \mathbf{p} , com configuração composta por $r - 1$ níveis (fixos) e o nível 1 do fator variável. No denominador tem-se a probabilidade de escolha do tratamento \mathbf{q} , com os mesmos $r - 1$ níveis (fixos) e o nível 2 do fator variável, conforme em (10):

$$RE(X_p, X_q) = \frac{P(Y_n = p|\mathbf{X})}{P(Y_n = q|\mathbf{X})} = \frac{\left(\frac{e^{X_p\beta}}{\sum_{k=1}^J e^{X_k\beta}} \right)}{\left(\frac{e^{X_q\beta}}{\sum_{k=1}^J e^{X_k\beta}} \right)} = \frac{e^{X_p\beta}}{e^{X_q\beta}} = e^{(X_p - X_q)\beta} \quad (10)$$

Em geral, pode-se concluir que se $RE(X_p, X_q) > 1$, o tratamento \mathbf{p} é mais provável de ser escolhido que o tratamento \mathbf{q} . O inverso ocorre se $RE(X_p, X_q) < 1$. Automaticamente, $RE(X_p, X_q) = 1$ implica que os dois tratamentos são igualmente prováveis de serem escolhidos. Destaca-se que o número de razões de escolha é definido como o número de níveis de um determinado fator combinados 2 a 2.

Uma informação importante é que a Razão de Escolhas depende apenas das características dos tratamentos \mathbf{p} e \mathbf{q} . Isto pode ser explicado devido à propriedade IIA (Independence of Irrelevant Alternatives), que é assegurada principalmente devido à independência dos erros aleatórios do modelo Logit Multinomial. Essa propriedade garante que tal proporção (R.E.) seja sempre constante, por isso é possível retirar ou incluir outros tratamentos do modelo sem alterar esta relação. Isso também é válido para a mudança de nível de um determinado fator (TRAIN, 2002). Em outras palavras, a probabilidade de escolha de um determinado tratamento não é afetada por tais mudanças ou inclusões.

Pode-se ilustrar tal fato com um simples exemplo relativo à escolha de três tipos de meios de transportes: **carro**, **avião** e **trem**. Neste exemplo, admita, respectivamente, para carro, avião e trem, as seguintes probabilidades de escolha: 0,20; 0,20 e 0,60. Neste caso, a Razão de Escolhas entre o tratamento trem e o tratamento carro é de

$R. E. (\text{Trem}, \text{Carro}) = \frac{0,60}{0,20} = 3$. Suponha que os fatores analisados para compor cada um dos respectivos tratamentos sejam o preço da passagem (R\$50, R\$150, R\$80) e o tempo de viagem (4 horas, 1 hora e 2 horas). Se devido a uma promoção o preço da passagem de avião diminuísse de R\$150 para R\$90 e sabendo-se que a viagem de avião é a mais rápida dentre as três opções, possivelmente a probabilidade de escolha do tratamento avião aumentaria. Portanto, as novas probabilidades estimadas de escolhas, poderiam ser, por exemplo, $P_{\text{Avião}} = 0,60$, $P_{\text{Carro}} = 0,10$, $P_{\text{Trem}} = 0,30$. No entanto, o valor da Razão de Escolhas anteriormente mencionada permanece igual a três, visto que agora, a probabilidade de escolha do tratamento trem é de 0,30 e do tratamento carro é de 0,10. Logo, $R. E. (\text{Trem}, \text{Carro}) = \frac{0,30}{0,10} = 3$.

2.6 Limitações do Modelo

Uma limitação do modelo Logit Multinomial é que ele não considera a heterogeneidade entre as escolhas dos consumidores, ou seja, pressupõe-se que a amostra em estudo possui, em geral, a mesma preferência. Portanto, conforme mencionado anteriormente, ao se estimar o vetor de parâmetros e as probabilidades de escolha, as estimativas serão únicas e tenderão a representar o comportamento médio de escolha dos consumidores. Neste sentido, algumas abordagens ainda no contexto de estimação Frequentista tem sido desenvolvidas, dentre as quais se destacam o modelo Logit com Heterogeneidade Ajustada (Heterogeneity Adjusted Logit ou HAL), conforme citado por Chesher e Santos Silva (2002), e o modelo Logit Misto (Mixed Logit model) ou com coeficientes aleatórios, relatado por McFadden e Train (2000). Além destas, em pesquisas recentes, tem sido abordada a modelagem da heterogeneidade não observável na amostra por meio de métodos de estimação Bayesianos, que procedem a Conjoint Analysis via modelos hierárquicos, denotados como Choice-Based Conjoint Hierarchical Bayes (CBC-HB), conforme apresentados em Rossi et al. (2005). Neste tipo de abordagem, estimam-se os parâmetros do modelo Logit Multinomial em nível individual (com variáveis demográficas) e admite-se a incorporação de informações a priori sobre estes parâmetros, por meio de modelos de probabilidade [ver também os trabalhos de Solgaard e Hansen (2003); Wellman e Vidican (2008) e Sawtooth (2009), entre outros].

Capítulo 3

Choice-Based Conjoint Analysis: Um enfoque Bayesiano

3.1 Uma abordagem Bayesiana para o Modelo Logit Multinomial

Com o avanço dos métodos computacionais e, conseqüentemente, com a maior flexibilidade e rapidez dos softwares, uma área da estatística que vem recebendo grande destaque (a partir da década de 1990) e apresentando alternativas práticas e viáveis para a solução de problemas complexos é a inferência Bayesiana, que emprega métodos iterativos conhecidos como Monte Carlo via Cadeias de Markov (Markov Chain Monte Carlo ou simplesmente MCMC). A aplicação da inferência Bayesiana na CBCA ainda é recente e escassa. Rossi et al. (2005) são os propulsores desta nova tendência em modelagem da preferência do consumidor e desenvolveram uma abordagem Bayesiana para a estimação dos parâmetros do Modelo Logit Multinomial, conhecido como Conditional Logit Model.

No contexto da inferência Bayesiana, os parâmetros não são tratados como quantidades fixas e desconhecidas, e sim como variáveis aleatórias associadas a modelos de probabilidades. Esta é a principal diferença entre tal abordagem e a Frequentista (ou Clássica), em que na segunda, o estimador do parâmetro é que possui uma distribuição de probabilidades, denominada de distribuição amostral do estimador, e as inferências são realizadas a partir do conhecimento desta distribuição, em geral válida para grandes amostras.

Os métodos de inferência Bayesiana originam-se no Teorema de Bayes (BAYES, 1763). Por este teorema a distribuição a posteriori conjunta dos parâmetros, denominada por $\pi(\boldsymbol{\beta})$, é obtida combinando-se distribuição a priori com informações provenientes da amostra aleatória, introduzidas por meio da função de verossimilhança.

A informação a priori, como o próprio nome sugere, é uma informação que o pesquisador possui antes da coleta dos dados e deve ser incorporada ao estudo por meio de modelos de probabilidades, denotados por $p(\boldsymbol{\beta})$. A função de verossimilhança é a

função de probabilidade ou de densidade de probabilidade conjunta da amostra, $L(\boldsymbol{\beta})$, obtida após se observar os dados. Pelo Teorema de Bayes tem-se que:

$$P(\boldsymbol{\beta}|\mathbf{y}) = \frac{P(\boldsymbol{\beta} \cap \mathbf{y})}{P(\mathbf{y})}^{14}$$

$$P(\boldsymbol{\beta}|\mathbf{y}) \propto P(\boldsymbol{\beta} \cap \mathbf{y})$$

$$P(\boldsymbol{\beta}|\mathbf{y}) \propto P(\mathbf{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta})$$

A expressão final em (11), conforme a notação mencionada anteriormente é a base das inferências segundo o enfoque Bayesiano:

$$\pi(\boldsymbol{\beta}) \propto L(\boldsymbol{\beta}) p(\boldsymbol{\beta}) \quad (11)$$

Para uma revisão sobre os respeito dos conceitos básicos da inferência Bayesiana, sugere-se ao leitor os textos de Gianola e Fernando (1986), Box e Tiao (1992), Gamerman (1997), Ehlers (2003), Gamerman e Lopes (2006), Migon et al. (2008).

¹⁴ Como $P(\mathbf{y})$ independe de $\boldsymbol{\beta}$ este termo atua como uma constante normalizadora e pode ser incorporado a proporção.

3.2 Distribuição a Priori

A metodologia Bayesiana exige que se atribua uma distribuição de probabilidades a priori para o parâmetro (uma distribuição uni-variada) ou para o vetor de parâmetros (distribuição multivariada). Na CBCA, para modelar a incerteza associada aos verdadeiros valores de $\boldsymbol{\beta}$, Rossi et al. (2005) sugeriram o uso de uma distribuição normal r-variada, $\boldsymbol{\beta} \sim \text{Normal}_r(\bar{\boldsymbol{\beta}}, A)$, cuja função densidade de probabilidade é dada em (12):

$$\begin{aligned}
 p(\boldsymbol{\beta}) &= \frac{1}{\sqrt{(2\pi)^r |\Sigma|}} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' \Sigma^{-1} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})] \right\} \\
 p(\boldsymbol{\beta}) &= \frac{1}{\sqrt{(2\pi)^r} \sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' \Sigma^{-1} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})] \right\} \\
 p(\boldsymbol{\beta}) &\propto \frac{1}{\sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' \Sigma^{-1} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})] \right\} \\
 p(\boldsymbol{\beta}) &\propto \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' \Sigma^{-1} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})] \right\} \\
 p(\boldsymbol{\beta}) &\propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' \Sigma^{-1} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})] \right\} \\
 p(\boldsymbol{\beta}) &\propto (|\Sigma|^{-1})^{1/2} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' \Sigma^{-1} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})] \right\}
 \end{aligned}$$

Considerando $\Sigma^{-1} = A$.

$$p(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' A (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})] \right\} \quad (12)$$

Em que $\bar{\boldsymbol{\beta}}$ é o vetor que contém os valores médios da distribuição dos parâmetros e Σ a matriz de variâncias e covariâncias do vetor $\boldsymbol{\beta}$. Especifica-se a dispersão por meio do parâmetro A, denotado também como a precisão dos valores da distribuição. No caso uni-variado A é o inverso da variância ($A = \sigma^{-2}$) e, no multivariado, a inversa da matriz de variâncias e covariâncias ($A = \Sigma^{-1}$). Os parâmetros $\bar{\boldsymbol{\beta}}$ e A são mencionados na literatura Bayesiana como hiperparâmetros e podem também assumir distribuições de probabilidade, constituindo uma hierarquia de distribuições.

Para maiores esclarecimentos, em um estudo de CBCA no qual os tratamentos são compostos pela combinação dos níveis de r fatores, Rossi et al. (2005) sugeriram a seguinte especificação para a distribuição a priori:

$$\boldsymbol{\beta} \sim \text{Normal}_r \left(\begin{matrix} \left[\begin{matrix} \mu_{\beta_1} \\ \mu_{\beta_2} \\ \vdots \\ \mu_{\beta_r} \end{matrix} \right], \left[\begin{matrix} 1/\sigma_{\beta_1}^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_{\beta_2}^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1/\sigma_{\beta_r}^2 \end{matrix} \right] \end{matrix} \right)$$

Com os seguintes valores:

$$\bar{\boldsymbol{\beta}} = \begin{bmatrix} \mu_{\beta_1} \\ \mu_{\beta_2} \\ \vdots \\ \mu_{\beta_r} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{\beta_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\beta_2}^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_{\beta_r}^2 \end{bmatrix} = \begin{bmatrix} 100 & 0 & \dots & 0 \\ 0 & 100 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 100 \end{bmatrix}, \text{ portanto,}$$

$$\mathbf{A} = \begin{bmatrix} 1/\sigma_{\beta_1}^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_{\beta_2}^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1/\sigma_{\beta_r}^2 \end{bmatrix} = \begin{bmatrix} 0,01 & 0 & \dots & 0 \\ 0 & 0,01 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0,01 \end{bmatrix}, \text{ o que resulta em:}$$

$$\boldsymbol{\beta} \sim \text{Normal}_r \left(\begin{matrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0,01 & 0 & \dots & 0 \\ 0 & 0,01 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0,01 \end{bmatrix} \end{matrix} \right).$$

Os valores anteriores, estipulados para $\bar{\boldsymbol{\beta}}$ e \mathbf{A} na distribuição a priori normal r-variada, equivalem à especificação de uma distribuição pouco informativa para $\boldsymbol{\beta}$, devido à escolha de um elevado valor para as variâncias dos parâmetros (100), o que implica em uma baixa precisão (0,01). No entanto, esta é uma abordagem frequentemente utilizada, visto que em muitas situações não há informações a priori sobre os parâmetros do modelo em estudo.

Outras sugestões para os valores médios da distribuição a priori normal r-variada poderiam ser as estimativas de Máxima Verossimilhança ou de Máxima Verossimilhança Parcial ($\bar{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$) Frequentistas. Para o parâmetro de precisão, pode-se utilizar o negativo da matriz Hessiana [$\mathbf{A} = \mathbf{H}^* = -\mathbf{H}$]. Conforme será explicado mais adiante no texto, à medida que o tamanho da amostra (número de consumidores) aumenta os resultados Bayesianos não irão divergir em relação a diferentes especificações para os parâmetros da distribuição a priori.

Um detalhe importante refere-se à utilização de distribuições a priori conjugadas naturais, uma prática muito comum em estudos de inferência Bayesiana. Este

procedimento permite obter soluções ou sumarizar os principais momentos das distribuições a posteriori de forma analítica. A ideia central consiste em selecionar uma distribuição a priori que seja conjugada ou proporcional à função de verossimilhança dos dados. Se isso for possível, a distribuição a posteriori pertencerá à mesma classe de distribuição da priori, com modificações (atualizações) nos parâmetros. No entanto, para os estudos de CBCA, Koop e Poirier (1993) afirmaram que não há uma priori conjugada para os parâmetros do modelo Logit Multinomial (ou Conditional Logit Model). Por este motivo, Rossi et al. (2005) explicaram que o uso de uma priori normal r-variada, como elucidado anteriormente, seria razoável.

Note que o uso de uma distribuição a priori normal r-variada, quando combinada com a função de verossimilhança dos dados mencionada¹⁵ em (8), não é garantia de que a densidade a posteriori conjunta seja também uma normal ou que possua a forma de uma distribuição conhecida. No entanto, com alguns artifícios teóricos sob a função de verossimilhança o uso desta distribuição a priori favorecerá a solução dos problemas Bayesianos envolvidos na análise da CBCA.

¹⁵ $L(\boldsymbol{\beta}) = \prod_{n=1}^N \prod_{j=1}^J \left(\frac{e^{x_j \boldsymbol{\beta}}}{\sum_{k=1}^J e^{x_k \boldsymbol{\beta}}} \right)^{I_{nj}}$

3.3 Função de Verossimilhança

Devido ao fato de não existir uma conjugação natural para a análise da CBCA, Rossi et al. (2005) propuseram uma aproximação de Laplace (LAPLACE, 1986) para a função de verossimilhança do modelo Logit Multinomial ou, em outras palavras, uma aproximação normal assintótica. Note que ao se utilizar uma priori normal r-variada combinada com uma verossimilhança aproximadamente normal, a densidade a posteriori conjunta também possuirá uma densidade normal r-variada. Em outras palavras, o uso da aproximação de Laplace para a função de verossimilhança permite a ocorrência de uma conjugação natural, ou seja, em que a verossimilhança dos dados será proporcional a priori empregada.

Tal procedimento pode ser realizado pela expansão de seu logaritmo, $l(\boldsymbol{\beta}) = \ln[L(\boldsymbol{\beta})]$, em uma série de Taylor de segunda ordem, avaliada no vetor de estimadores de Máxima Verossimilhança (ou Máxima Verossimilhança Parcial), conforme a modelagem apresentada a seguir:

$$l(\boldsymbol{\beta}) \approx l(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \left[\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right] + \frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right] (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

Como $\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$ é um valor de máximo, visto que maximiza a densidade conjunta da amostra aleatória, por definição:

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0, \quad \text{portanto}$$

$$l(\boldsymbol{\beta}) \approx l(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right] (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$$l(\boldsymbol{\beta}) \approx l(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' H (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

Se o objetivo é obter uma expansão para a Função de Verossimilhança, então:

$$L(\boldsymbol{\beta}) \approx \exp\{l(\boldsymbol{\beta})\}$$

$$L(\boldsymbol{\beta}) \approx \exp \left\{ l(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' H (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}$$

$$L(\boldsymbol{\beta}) \approx \exp\{l(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} \exp \left\{ \frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' H (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}$$

$$L(\boldsymbol{\beta}) \approx L(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \exp \left\{ \frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' H (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}$$

Como $L(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$ é uma constante que representa o valor máximo da função de verossimilhança, pode ser incorporada a proporção, resultando em (13).

$$L(\boldsymbol{\beta}) \approx \exp \left\{ \frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' H (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}$$

$$L(\boldsymbol{\beta}) \approx \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' (-H) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}$$

Utilizando $H^* = -H = -E \left[\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right]$

$$L(\boldsymbol{\beta}) \approx \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' H^* (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\} \quad (13)$$

Note que o resultado acima se assemelha ao Kernel¹⁶, base de uma distribuição Normal Multivariada com vetor de médias igual aos estimadores de Máxima Verossimilhança (ou Máxima Verossimilhança Parcial) e matriz de precisão igual ao negativo do valor esperado da matriz Hessiana, ou seja, $L(\boldsymbol{\beta}) \sim N_r(\boldsymbol{\mu} = \hat{\boldsymbol{\beta}}, A = H^*)$.

O fato do logaritmo da função de verossimilhança do modelo Logit Multinomial possuir uma forma globalmente côncava, apresentando apenas um único ponto de máximo (máximo relativo = máximo global), garante que a utilização de métodos assintóticos (como o da aproximação de Laplace) funcione muito bem. Conforme mencionado anteriormente, se o logaritmo da função de verossimilhança é uma função côncava, H é uma matriz negativa definida. Portanto, $H^* = -H$ será uma matriz positiva definida e, conseqüentemente, sua inversa também. Por isso, novamente, todos os resultados aqui apresentados serão baseados no valor esperado do negativo da matriz Hessiana.

¹⁶ O Kernel é também conhecido como o núcleo de uma distribuição e no contexto Bayesiano, como função de importância.

3.4 Distribuição a Posteriori

Ao combinar a informação dos dados via função de verossimilhança com a informação que se possui a priori, tem-se a distribuição a posteriori conjunta dos parâmetros do modelo Logit Multinomial. A seguir, apresenta-se a modelagem completa para a obtenção desta densidade e dos seus respectivos parâmetros:

$$\pi(\boldsymbol{\beta}) \propto L(\boldsymbol{\beta}) p(\boldsymbol{\beta})$$

$$\pi(\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' H^* (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\right\} |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}[(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' A (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})]\right\}$$

$$\pi(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' H^* (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\right\} \exp\left\{-\frac{1}{2}[(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' A (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})]\right\}$$

$$\pi(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' H^* (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' A (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})\right\}$$

$$\pi(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}[(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' H^* (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' A (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})]\right\}$$

$$\begin{aligned} {}^{17} \pi(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[{}^{18} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' (H^* + A) (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' H^* (\widehat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \right. \right. \\ \left. \left. + (\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' A (\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\right]\right\} \end{aligned}$$

Para simplificar, tome que $SS = (\widehat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' H^* (\widehat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) + (\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' A (\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$:

$$\pi(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}[(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' (H^* + A) (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + SS]\right\}$$

$$\pi(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}SS\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' (H^* + A) (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right\}$$

Multiplicando e dividindo a equação acima por $|H^* + A|^{\frac{1}{2}}$, tem-se:

$$\pi(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \frac{|H^* + A|^{\frac{1}{2}}}{|H^* + A|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}SS\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' (H^* + A) (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right\}$$

¹⁷ $(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' H^* (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' A (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) = (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' (H^* + A) (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' H^* (\widehat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) + (\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' A (\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$

¹⁸ $\tilde{\boldsymbol{\beta}} = (H^* + A)^{-1}(H^* \widehat{\boldsymbol{\beta}} + A \bar{\boldsymbol{\beta}})$

$$\pi(\boldsymbol{\beta}) \propto \frac{|A|^{\frac{1}{2}}}{|H^* + A|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}SS\right\} |H^* + A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'(H^* + A)(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right\}$$

Incorporando os dois primeiros termos (constantes) na proporção da equação anterior, tem-se uma densidade similar à de uma normal multivariada, que possui um vetor de médias $\tilde{\boldsymbol{\beta}}$ e matriz de precisão $(H^* + A)$ como parâmetros:

$$\pi(\boldsymbol{\beta}) \propto |H^* + A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'(H^* + A)(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right\}$$

Ao avaliar a sugestão de Rossi et al. (2005), descrita na seção 3.2, para a especificação da distribuição a priori dos parâmetros, em que $\bar{\boldsymbol{\beta}}$ é um vetor nulo de dimensão $r \times 1$ e A uma matriz identidade com dimensão $r \times r$ e diagonal igual a 0,01, a seguinte conclusão pode ser obtida:

$$H^* + A \approx H^*$$

Visto que os valores presentes na diagonal de H^* serão somados a 0,01 e os demais a 0, conclui-se que não haverá alterações. Além disso:

$$\tilde{\boldsymbol{\beta}} = (H^* + A)^{-1}(H^*\hat{\boldsymbol{\beta}} + A\bar{\boldsymbol{\beta}})$$

$$\tilde{\boldsymbol{\beta}} = (H^*)^{-1}(H^*\hat{\boldsymbol{\beta}} + \mathbf{0})$$

$$\tilde{\boldsymbol{\beta}} = (H^*)^{-1}H^*(\hat{\boldsymbol{\beta}})$$

$$(H^*)^{-1}H^* = I = \textit{Identidade}$$

$$\tilde{\boldsymbol{\beta}} = I\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$$

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$$

Substituindo estes resultados, tem-se em (14) a expressão final para a distribuição a posteriori:

$$\pi(\boldsymbol{\beta}) \propto |H^* + A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'(H^* + A)(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right\}$$

$$\pi(\boldsymbol{\beta}) \propto |H^*|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'H^*(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \quad (14)$$

Logo, a densidade conjunta a posteriori é uma normal r -variada com vetor de médias igual aos estimadores de Máxima Verossimilhança (neste caso de Máxima Verossimilhança Parcial) e matriz de precisão igual ao negativo do valor esperado da matriz Hessiana, ou seja, $\pi(\boldsymbol{\beta}) \sim N_r(\boldsymbol{\mu} = \hat{\boldsymbol{\beta}}, A = H^*)$.

3.4.1 Distribuição a Posteriori Marginal

No enfoque Bayesiano todo procedimento inferencial é baseado na distribuição marginal de cada elemento β_i que compõe o vetor paramétrico $\boldsymbol{\beta}$, ou seja, em cada $\pi(\beta_i)$ e não em $\pi(\boldsymbol{\beta})$. Note que sob o enfoque inferencial Frequentista, deve-se obter a estimativa pontual de cada uma dos r parâmetros (fatores), enquanto que pela metodologia Bayesiana, é necessário encontrar a distribuição marginal de cada um deles.

Pelos conceitos básicos da estatística Clássica, dado que X e Y são duas variáveis aleatórias contínuas e que se conhece sua densidade conjunta $[f_{X,Y}(x,y)]$, é possível escrever suas respectivas densidades marginais $[f_X(x)$ e $f_Y(y)]$ como:

$$f_X(x) = \int_{\Omega_Y} f_{X,Y}(x,y) dy$$

$$f_Y(y) = \int_{\Omega_X} f_{X,Y}(x,y) dx$$

Adaptando estes conceitos à situação da CBCA, com o vetor de parâmetros $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_r)'$, tem-se que:

$$\pi(\beta_i) = \int_{\Omega_{\beta_{-i}}} \pi(\boldsymbol{\beta}) d\beta_{-i}$$

Note que a distribuição a posteriori marginal de cada β_i é obtida pela integração da distribuição conjunta a posteriori em relação a todos os demais elementos do vetor paramétrico $\boldsymbol{\beta}$, com exceção de β_i , por isso a notação $d\beta_{-i}$. Em (15) tem-se esta explicação de forma matemática:

$$\pi(\beta_i) = \int_{\Omega_{\beta_r}} \dots \int_{\Omega_{\beta_{i+1}}} \int_{\Omega_{\beta_{i-1}}} \dots \int_{\Omega_{\beta_2}} \int_{\Omega_{\beta_1}} \pi(\boldsymbol{\beta}) d\beta_1 d\beta_2, \dots, d\beta_{i-1} d\beta_{i+1}, \dots, d\beta_r \quad (15)$$

A distribuição a posteriori marginal de cada β_i contém toda a informação probabilística referente a este elemento do vetor paramétrico. Após o conhecimento destas distribuições, é possível sumarizar os seus principais momentos e funções. Sob este aspecto existe uma analogia entre a estimação pontual Frequentista e a Bayesiana, visto que na segunda, tende-se a tomar como estimativas, algumas medidas associadas

às distribuições a posteriori marginais. Os principais estimadores Bayesianos são a média (16), moda (17) e mediana (18), a posteriori. Naturalmente, medidas que descrevam a incerteza associada a estas distribuições devem ser calculadas, tais como a variância (19) ou o desvio-padrão (20), a posteriori. A escolha de qual medida tomar para representar a estimativa Bayesiana irá depender da forma da distribuição e dos objetivos ou pressuposições sobre o seu uso (EHLERS, 2007).

$$E[\beta_i] = \int_{\Omega_{\beta_i}} \beta_i \pi(\beta_i) d\beta_i \quad (16)$$

$$\text{máx } \{\pi(\beta_i)\} = \frac{d \{\log[\pi(\boldsymbol{\beta})]\}}{d\beta_i} \quad (17)$$

$$P[\beta_i \geq \text{med}(\beta_i)] \geq \frac{1}{2} \text{ e } P[\beta_i \leq \text{med}(\beta_i)] \geq \frac{1}{2} \quad (18)$$

$$\sigma^2(\beta_i) = E[\beta_i^2] - (E[\beta_i])^2 \quad (19)$$

$$\sigma(\beta_i) = \sqrt{\sigma^2(\beta_i)} \quad (20)$$

No entanto, a obtenção de tais medidas pode não ser uma tarefa simples, visto que a solução da integral em (15) por métodos analíticos pode não ser possível. Isso ocorre devido à complexidade da distribuição a posteriori conjunta e, ou, devido ao elevado número de elementos que compõe o vetor de parâmetros, o que aumenta consideravelmente a dificuldade do procedimento matemático, por causa do número de integrais múltiplas (SORENSEN, 1996).

Para solucionar tais problemas, podem ser empregados métodos aproximados (aproximando a integral de interesse por uma integral que possua solução analítica), métodos de quadratura (aproximando a soma desconhecida representada pela integral por uma soma conhecida) ou os mais recentes métodos que utilizam algoritmos MCMC. Uma justificativa para a última é fornecida por Rossi et al. (2005), visto que como a aproximação normal assintótica para a função de verossimilhança funciona muito bem, a construção de um algoritmo MCMC que utilize as vantagens desta aproximação poderia apresentar resultados também satisfatórios. Veja o seguinte trecho citado pelos autores:

“Experience with the Multinomial Logit likelihood is that the asymptotic normal approximation is excellent. This suggests that Metropolis algorithms based on the asymptotic approximation will perform extremely well. We also note that the Multinomial Logit likelihood has exponential tails (even without the normal prior) and this should provide very favorable theoretical convergence properties”

3.5 Conceitos Iniciais sobre os Algoritmos MCMC

Os MCMCs são métodos de simulação iterativos, originados sob os princípios de uma Cadeia de Markov. Segundo Gamerman e Lopes (2006), uma Cadeia de Markov é definida como um conjunto de variáveis aleatórias (X_0, X_1, \dots, X_k) , que constituem um processo estocástico, cujo próximo estado da cadeia (X_{k+1}) depende apenas e exclusivamente do estado anterior (X_k). Em termos probabilísticos, a distribuição ou a probabilidade condicional de X_{k+1} dado todos os demais valores anteriores, depende apenas de X_k , ou seja:

$$P(X_{k+1} \in A | X_k, X_{k-1}, \dots, X_1, X_0) = P(X_{k+1} \in A | X_k)$$

Este tipo de processo estocástico é conhecido como processo sem memória (memoryless process), visto que o estado futuro depende apenas do estado presente, sendo o passado desprezado. Mais especificamente, se a probabilidade de transição independe do tempo, dado um espaço de estado (discreto) finito, definido como $A = \{x_1, x_2, \dots, x_k\}$, para qualquer X_{k+1} pode-se escrever:

$$P(X_{k+1} = x_{k+1} | X_k = x_k) = P(X_1 = x_1 | X_0 = x_0) = p_{ij}$$

Onde i e j representam quaisquer estados pertencentes ao espaço de estado de A . Note que nestas situações é preciso especificar um valor para o estado inicial da cadeia. Denota-se por $\mathbf{P}(i, j)$ a matriz de probabilidades de transição entre quaisquer estados i e j , para i e $j = 1, 2, 3, \dots, k$ pertencentes ao espaço de estado A :

$$\mathbf{P}(i, j) = \begin{bmatrix} p_{i=1,j=1} & p_{i=1,j=2} & \cdots & p_{i=1,j=k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i=k,j=1} & p_{i=k,j=2} & \cdots & p_{i=k,j=k} \end{bmatrix}$$

A matriz $\mathbf{P}(i, j)$ retrata as probabilidades de mudança entre os estados da cadeia apenas no primeiro passo (ou na primeira iteração), ou seja, de $X_k \rightarrow X_{k+1}$. No entanto, pode haver o interesse no conhecimento de tais probabilidades a R passos no tempo, ou seja, de $X_k \rightarrow X_{k+2}$, ou $X_k \rightarrow X_{k+3}, \dots$, ou $X_k \rightarrow X_{k+R}$. Pelas equações de Chapman-Kolmogorov e conforme pode ser provado por indução matemática, tem-se que:

$$\mathbf{P}(i, j)^{(R)} = \mathbf{P}(i, j)^R$$

Definindo π_0 como o vetor de probabilidades (ou a distribuição) associado ao estado inicial da cadeia e baseado no resultado anterior, chega-se à seguinte associação:

$$\pi_1 = \pi_0 \mathbf{P}(i, j)^1$$

$$\pi_2 = \pi_0 \mathbf{P}(i, j)^2$$

...

$$\pi_R = \pi_0 \mathbf{P}(i, j)^R$$

Note que a cada iteração a distribuição dos estados da cadeia passa por uma transformação. Quando o número de iterações é suficientemente grande (tendendo ao infinito) a cadeia de Markov converge para uma distribuição estacionária, invariante ou de equilíbrio. No contexto da inferência Bayesiana, a distribuição estacionária representa a distribuição de interesse. Para que isso ocorra, a cadeia deve ser ainda: homogênea, irredutível e aperiódica.

Uma cadeia é homogênea se as probabilidades de transição de um estado para qualquer outro independe do tempo (probabilidades de transição constantes). Quando em um movimento de qualquer estado para outro existir uma probabilidade de transição não nula, diz-se que a cadeia é irredutível e que, portanto, haverá comunicação entre todos os seus estados. A propriedade de irredutibilidade garante que:

$$\lim_{R \rightarrow \infty} \pi_0 \mathbf{P}(i, j)^R = \pi$$

$$\pi \mathbf{P}(i, j) = \pi$$

Por estas notações, conclui-se porque a distribuição estacionária é também conhecida como invariante, visto que com o aumento do número de iterações, a partir de qualquer distribuição inicial chega-se à distribuição estacionária. O mesmo acontece se partir da distribuição estacionária. Por fim, uma cadeia é dita aperiódica quando ela não se prende a ciclos (TIERNEY, 1994).

Se uma cadeia de Markov atende a estas três condições (sendo homogênea, irredutível e aperiódica), é também denominada como ergódica. O conceito de ergodicidade é muito importante para a solução dos problemas de integração presentes no contexto da inferência Bayesiana. Se uma cadeia é ergódica, é possível, por exemplo, utilizar sua média amostral para estimar o respectivo valor esperado da distribuição de interesse, e assim sucessivamente para as demais funções (GAMERMAN; LOPES, 2006). Ou seja:

$$\lim_{R \rightarrow \infty} \frac{1}{R} \sum_{j=1}^R \beta_{ij} \approx E_{\pi}[\beta_i], i = 1, 2, \dots, r$$

$$\lim_{R \rightarrow \infty} \frac{1}{R} \sum_{j=1}^R h(\beta_{ij}) \approx E_{\pi}[h(\beta_i)], i = 1, 2, \dots, r$$

Tais resultados, se uma cadeia é ergódica, podem ser provados de acordo com a lei forte dos grandes números, e convergem quase certamente (isto é, com probabilidade igual a 1) quando R tende ao infinito (FELLER, 1968).

Destaca-se ainda que a ergodicidade de uma cadeia de Markov pode ser uniforme ou geométrica. Segundo Tierney (1994), uma cadeia é geometricamente ergódica se existe um valor constante real $0 \leq \lambda < 1$ e também uma função $M(x)$, tal que:

$$|\mathbf{P}(i, \cdot)^R - \pi(\cdot)| \leq M(x)\lambda^R$$

Se M independe de x , a ergodicidade é uniforme. O referido autor afirma que o menor valor de λ que satisfaça à desigualdade, é conhecido como taxa de convergência. Taxas de convergências geométricas implicam em rápida convergência, sem a necessidade de muitas iterações, por isso, são desejáveis. Ainda segundo Tierney (1994), “Uniform ergodicity implies geometric ergodicity which implies *ergodicity*”.

3.5.1 Aspectos Computacionais sobre os Algoritmos MCMC

No aspecto computacional, a implementação de algoritmos MCMC em softwares estatísticos exige o estabelecimento de alguns critérios. Naturalmente, é preciso definir um número de iterações R para que as propriedades das cadeias e , conseqüentemente, sua convergência seja alcançada. No processo de simulação, as primeiras iterações sofrem a influência do estado inicial da cadeia e , portanto, devem ser desprezadas. Isto é realizado por meio da especificação de um argumento denominado Burn-in, ou número de iterações que precisam ser eliminadas para que ocorra o aquecimento da cadeia.

Devido à dependência entre os estados de uma cadeia de Markov, as amostras geradas por métodos MCMC não serão independentes e possuirão uma estrutura de autocorrelação. Para minimizar tal efeito, especifica-se o termo thinning (ou simplesmente thin), que indica um espaçamento entre as iterações armazenadas, implicando que as amostras simuladas serão selecionadas a cada thin iterações (GAMERMAN, 1997). Este argumento visa reduzir o grau de autocorrelação da amostra, no intuito de que esta seja a mais aleatória possível, pois muitas versões da lei forte dos grandes números assumem independência entre as observações. No entanto, Rossi et al. (2005) afirmam:

“As long as the dependence is not pathologically strong, then the intuition behind the law of large numbers still goes through.”

3.5.2 Independence Metropolis Hasting

Os dois principais algoritmos MCMC utilizados em estudos de inferência Bayesiana são o amostrador de Gibbs ou Gibbs sampler (GEMAN; GEMAN, 1984) e o Metropolis Hasting, desenvolvido por Metropolis et al. (1953) e aprimorado por Hastings (1970). A principal distinção entre os dois métodos é que o amostrador de Gibbs é indicado para gerar uma amostra da distribuição a posteriori de interesse, quando a distribuição condicional completa a posteriori dos parâmetros possui uma forma conhecida, isto é, semelhante a algum modelo de probabilidades documentado na literatura. Caso contrário, se a forma da condicional completa for desconhecida ou se o processo de amostragem a partir desta for muito complexo, o algoritmo mais indicado é o Metropolis Hasting, que utiliza distribuições candidatas para o processo de amostragem.

Neste trabalho será detalhado e utilizado o algoritmo Independence Metropolis Hasting (um caso especial¹⁹ do algoritmo de Metropolis), devido à estratégia desenvolvida por Zellner e Rossi (1984), que envolve o emprego de uma distribuição candidata t-Student multivariada.

Este algoritmo MCMC está disponível no pacote bayesm do software livre R, via função `rmnlIndepMetrop`. A distribuição candidata utilizada é uma distribuição t-Student k-variada (simétrica e pertencente à classe de distribuições elípticas), com v graus de liberdade e parâmetros de locação e escala, respectivamente, iguais a $\boldsymbol{\mu}$ e A . Sua função densidade pode ser escrita conforme (21):

$$p(\mathbf{y}|v, \boldsymbol{\mu}, \Sigma) = \frac{\Gamma\left(\frac{v+k}{2}\right)}{\Gamma\left(\frac{v}{2}\right) v^{\frac{k}{2}} \pi^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}} \left[1 + \frac{1}{v}(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})\right]^{\frac{k+v}{2}}}$$
$$p(\mathbf{y}|v, \boldsymbol{\mu}, \Sigma) \propto \frac{1}{|\Sigma|^{\frac{1}{2}} \left[1 + \frac{1}{v}(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})\right]^{\frac{k+v}{2}}}$$

¹⁹ O amostrador de Gibbs e o Independence Metropolis Hasting são casos particulares do algoritmo original de Metropolis Hasting. Além disso, um terceiro caso particular é o algoritmo conhecido como Random-Walk Metropolis Hasting (ou passeio aleatório).

$$\begin{aligned}
p(\mathbf{y}|v, \boldsymbol{\mu}, \Sigma) &\propto \frac{1}{|\Sigma|^{\frac{1}{2}} \left[\frac{v + (\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})}{v} \right]^{\frac{k+v}{2}}} \\
p(\mathbf{y}|v, \boldsymbol{\mu}, \Sigma) &\propto \frac{1}{|\Sigma|^{\frac{1}{2}} \left(\frac{1}{v} \right)^{\frac{k+v}{2}} [v + (\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})]^{\frac{k+v}{2}}} \\
p(\mathbf{y}|v, \boldsymbol{\mu}, \Sigma) &\propto \frac{1}{|\Sigma|^{\frac{1}{2}} [v + (\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})]^{\frac{k+v}{2}}} \\
p(\mathbf{y}|v, \boldsymbol{\mu}, \Sigma) &\propto |\Sigma|^{-\frac{1}{2}} [v + (\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})]^{-\frac{k+v}{2}} \\
p(\mathbf{y}|v, \boldsymbol{\mu}, \Sigma) &\propto |\Sigma^{-1}|^{\frac{1}{2}} [v + (\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})]^{-\frac{k+v}{2}} \\
p(\mathbf{y}|v, \boldsymbol{\mu}, A) &\propto |A|^{\frac{1}{2}} [v + (\mathbf{y} - \boldsymbol{\mu})^T A (\mathbf{y} - \boldsymbol{\mu})]^{-\frac{k+v}{2}} \tag{21}
\end{aligned}$$

Analogamente, para os estudos de CBCA, Rossi et al. (2005) sugerem que o vetor de parâmetros do modelo Logit Multinomial seja gerado de acordo com a seguinte especificação de $\boldsymbol{\beta} \sim MSt_r(v = 6, \boldsymbol{\mu} = \widehat{\boldsymbol{\beta}}, A = H^*)$. Reescrevendo (21) e substituindo os parâmetros, tem-se:

$$p(\boldsymbol{\beta}|v, \widehat{\boldsymbol{\beta}}, H^*) \propto |H^*|^{\frac{1}{2}} \left[6 + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^T H^* (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \right]^{-\frac{r+6}{2}}$$

Repare que o Independence Metropolis Hasting utiliza (implicitamente) a densidade conjunta a posteriori em (14) como distribuição candidata ou como uma função de importância. No entanto, são procedidas algumas modificações para transformá-la em uma t-Student r-variada, o que contempla a inclusão do parâmetro v , necessário para tornar a distribuição candidata mais densa ou com caudas mais pesadas do que a distribuição de interesse. Esta é a estratégia proposta por Zellner e Rossi (1984). Dado que X e Y são duas variáveis aleatórias, define-se uma terceira variável Z conforme o esquema a seguir:

$$X \sim N(0, A = H^*)$$

$$Y \sim \chi_v^2$$

$$Z = \frac{X}{\left(\frac{Y}{v}\right)^{\frac{1}{2}}} + \boldsymbol{\mu} \quad \Rightarrow \quad Z = \frac{X}{\left(\frac{Y}{6}\right)^{\frac{1}{2}}} + \widehat{\boldsymbol{\beta}}$$

$$Z \sim MSt(v = 6, \boldsymbol{\mu} = \widehat{\boldsymbol{\beta}}, A = H^*)$$

Rossi et al. (2005) sugerem que a escolha de ν varie entre 5 e 15. Valores muito pequenos, como 4 ou menos, implicariam em uma distribuição com um pico ou com “ombros” extremamente estreitos, o que resultaria em uma função de importância ruim. O valor $\nu = 6$, é o sugerido e implementado como default no pacote bayesm pelos autores.

Como mencionado anteriormente, prova-se pela lei forte dos grandes números que a média da distribuição estacionária converge quase certamente para o valor esperado da distribuição de interesse, se a cadeia é ergódica (o mesmo vale para as suas respectivas funções). A confiabilidade destes resultados também está associada à escolha da distribuição candidata para o algoritmo MCMC. Em geral, estes resultados são válidos apenas se a distribuição candidata possuir o mesmo suporte da distribuição de interesse. Um detalhe importante é que se a distribuição candidata deve possuir caudas mais densas que a distribuição de interesse, caso contrário, os valores da cadeia serão rejeitados com maior frequência, diminuindo a eficiência do algoritmo e aumentando a autocorrelação entre as amostras. Além disso, as estimativas obtidas pela média amostral da distribuição estacionária tenderão a possuir variância infinita (CHIB; GREENBERG, 1995 e GEWEKE, 1989). Portanto, deve-se atentar à escolha da distribuição candidata, visando sempre que esta possua caudas mais densas do que a distribuição alvo.

Apresentaremos a seguir algumas observações a respeito da utilização do enfoque Bayesiano (priori, aproximação de Laplace, posteriori, algoritmo Independence Metropolis Hasting) para a análise da CBCA, baseado nas condições mencionadas anteriormente.

1 - Devido ao aumento do tamanho da amostra os valores simulados da distribuição a posteriori dos parâmetros tendem a se concentrar em torno de seus verdadeiros valores paramétricos, ou seja, assintoticamente a distribuição a priori não possui influência no processo inferencial. Veja no trecho a seguir a explicação de Rossi et al. (2005):

“As n increases, the posterior concentrates more and more mass in the vicinity of the ‘true’ value of β . The likelihood term dominates the prior and the prior becomes more and more uniform in appearance in the region in which the likelihood is concentrating. Thus, the prior has no asymptotic influence and the posterior starts to look MORE AND MORE NORMAL $\pi(\beta) \sim N_r(\hat{\beta}, \Sigma = [H^]^{-1})$ ”*

2 - Segundo Gamerman e Lopes (2006), ao tomar uma aproximação de Laplace para a função de verossimilhança e uma priori pouco informativa (aproximadamente

constante), a moda da densidade a posteriori, simulada via MCMC, irá convergir para os estimadores de Máxima Verossimilhança (neste caso de Máxima Verossimilhança Parcial). Naturalmente, com o aumento do tamanho da amostra e a distribuição a posteriori se aproximando mais e mais de uma Normal, a média, mediana e moda serão estimadores Bayesianos equivalentes.

3 - Rossi et al. (2005) informam que realizaram diversas simulações com o Independence Metropolis Hasting baseado na aproximação normal assintótica, inclusive com pequenos tamanhos de amostra, e concluíram que os resultados ainda foram satisfatórios. Veja o trecho citado por estes autores:

“Even in this extreme case, the actual posterior is only slightly skewed to the left from the asymptotic approximation.”

4 – No que se refere às propriedades de convergência do MCMC Independence Metropolis Hasting, o fato de utilizarmos uma priori normal facilita a obtenção de algumas propriedades desejáveis. Conforme Rossi et al. (2005):

“The theoretical properties of MCMC methods are quite appealing. For many algorithms and problems, especially with normal priors, it is easy to establish ergodicity and even stronger results such as geometric rates of convergence and uniform ergodicity”

5 - O fato da distribuição a posteriori conjunta ser aproximadamente uma normal r-variada (devido à aproximação de Laplace para a função de verossimilhança e uma priori normal), garante que sua densidade possua caudas exponenciais. Distribuições com caudas exponenciais favorecem à obtenção de excelentes propriedades de convergência para a cadeia de Markov (ergodicidade geométrica), mesmo que distribuições a priori normais não sejam utilizadas. Jarner e Tweedie (2003) provaram que as caudas exponenciais para a distribuição a posteriori são restrições necessárias para a ergodicidade geométrica em alguns MCMCs. Flajolet e Soria (1993) explicam que esta característica da distribuição, além de influenciar para uma rápida convergência (com taxas geométricas) do MCMC garante que valores muito diferentes dos que seriam esperados ocorram com probabilidade extremamente baixa.

6 - Os resultados obtidos serão válidos apenas se a distribuição a posteriori dos parâmetros for uni-modal (TIERNEY; KADAME, 1986).

7 - A abordagem descrita funciona muito bem se a função de verossimilhança apresentar uma forma globalmente côncava, como é o caso do modelo Logit

Multinomial. Além disso, Rossi et al. (2005) aconselham o seu uso apenas para problemas em que a dimensão do vetor de parâmetros varia entre 2 e 20.

3.5.2.1 Funcionamento do Independence Metropolis Hasting

Para explicar o funcionamento do algoritmo Independence Metropolis Hasting e consequentemente o procedimento de geração das amostras da distribuição a posteriori de interesse, denotaremos a distribuição candidata t-Student por $q(\beta^*|\beta^t)$. É importante relatar que a distribuição candidata deve ser uma função densidade de probabilidade, ou seja, $\int q(\beta^*|\beta^t) d\beta^* = 1$.

Suponha que a cadeia de Markov esteja no estado inicial $\beta^t = \beta_0$, cujos valores para β_0 podem ser definidos²⁰ arbitrariamente. Gera-se uma amostra β^* por meio da distribuição candidata $q(\beta^*|\beta^t)$, que será aceita com uma determinada probabilidade $\alpha(\beta^t, \beta^*)$, também conhecida como a probabilidade da cadeia se mover (probability of move).

$$\alpha(\beta^t, \beta^*) = \min\left(1, \frac{\pi(\beta^*) q(\beta^t|\beta^*)}{\pi(\beta^t) q(\beta^*|\beta^t)}\right)$$

Acima temos o cálculo de $\alpha(\beta^t, \beta^*)$ para o algoritmo original de Metropolis Hasting. No entanto, a principal diferença deste para o Independence Metropolis Hasting é que se considera que os valores gerados por $q(\beta^*|\beta^t)$ são independentes do valor atual da cadeia, o que garante que a estrutura de autocorrelação entre as amostras seja menor do que a exibida por outros algoritmos. Aplicando conceitos básicos de probabilidade condicional, é possível então escrever $\alpha(\beta^t, \beta^*)$ da seguinte forma:

$$\alpha(\beta^t, \beta^*) = \min\left(1, \frac{\pi(\beta^*) \left[\frac{q(\beta^t, \beta^*)}{q(\beta^*)}\right]}{\pi(\beta^t) \left[\frac{q(\beta^*, \beta^t)}{q(\beta^t)}\right]}\right)$$

$$\alpha(\beta^t, \beta^*) = \min\left(1, \frac{\pi(\beta^*) \left[\frac{q(\beta^t) q(\beta^*)}{q(\beta^*)}\right]}{\pi(\beta^t) \left[\frac{q(\beta^*) q(\beta^t)}{q(\beta^t)}\right]}\right)$$

$$\alpha(\beta^t, \beta^*) = \min\left(1, \frac{\pi(\beta^*) q(\beta^t)}{\pi(\beta^t) q(\beta^*)}\right)$$

Após calcular $\alpha(\beta^t, \beta^*)$, um valor aleatório proveniente de uma distribuição uniforme com suporte entre 0 e 1 é gerado, ou seja, $U \sim Unif(0,1)$. Se constatado que

²⁰ Geralmente os softwares estatísticos utilizam as estimativas de Máxima Verossimilhança dos parâmetros como os valores iniciais.

$u \leq \alpha(\beta, \beta^*)$ então $\beta^{t+1} = \beta^*$. Caso contrário, ocorre à rejeição de β^* e tem-se que $\beta^{t+1} = \beta_0$. Posteriormente, incrementa-se o tempo $t = t + 1$ e repete-se o referido procedimento R vezes.

3.6 Critérios de Convergência

A primeira forma de se avaliar a convergência de uma cadeia de Markov, embora não seja a mais recomendada, é por meio de métodos informais, como sugerem os autores Gelfand e Smith (1990), via análise gráfica da trajetória da cadeia durante o processo iterativo. Neste sentido, a partir de um determinado número de iterações, se for possível observar que o comportamento da cadeia de Markov é similar ao comportamento de um processo estacionário, então pode-se concluir sobre sua convergência.

A segunda forma e a mais indicada na literatura envolve a utilização de critérios estatísticos para diagnosticar a convergência das cadeias. Dentre estes critérios destacam-se os de Heidelberg e Welch (1983), Geweke (1992), Gelman e Rubin (1992) e Raftery e Lewis (1992). Além das referências citadas, em Cowles e Carlin (1996), e Nogueira et al. (2004), é possível encontrar uma revisão e discussões sobre estes critérios. Descreve-se a seguir, sucintamente, os dois critérios empregados neste trabalho, o critério de Raftery-Lewis e de Geweke.

O critério de Raftery-Lewis oferece cinco resultados interessantes: i) a estimativa do número total de iterações para que ocorra a convergência da cadeia na distribuição estacionária, ii) o número de Burn-in, iii) o número de thin, iv) o número mínimo ou o tamanho da amostra piloto (Lower-Bound) e v) o fator de dependência, calculado como o quociente entre a amostra de tamanho mínimo e o número total de iterações estimadas. Se o fator de dependência for superior a 5, tem-se um indicativo de que a cadeia não alcançou a convergência. O valor de Lower-Bound ou da amostra piloto pode ser calculado conforme em (22) (RAFTERY; LEWIS, 1992):

$$n_{min} = \left[\phi^{-1} \left(\frac{s+1}{2} \right) \frac{\sqrt{q(1-q)}}{r} \right]^2 \quad (22)$$

Em que usualmente $q = 0,025$, $r = 0,005$, $s = 0,95$ e ϕ^{-1} é o inverso da função de distribuição acumulada da normal padrão.

O critério de Geweke, com uma determinada probabilidade α , testa a hipótese de que a média da primeira parte da cadeia de Markov é estatisticamente igual à última parte. Normalmente, tomam-se as primeiras 10% e as últimas 50% observações simuladas. O procedimento consiste em um teste de médias baseado na diferença entre

as médias amostrais da primeira e da última parte da cadeia, ou seja, entre \bar{x}_a e \bar{x}_b . Por métodos padrões de análise espectral estimam-se variâncias assintóticas \hat{S}_a e \hat{S}_b , em que $\hat{S}_{(\cdot)}$ é a densidade espectral estimada da série de cada parte da cadeia de Markov. Logo, tem-se a seguinte estatística (GEWEKE, 1992):

$$\frac{\bar{x}_a - \bar{x}_b}{\sqrt{\left(\frac{\hat{S}_a}{n_a}\right) - \left(\frac{\hat{S}_b}{n_b}\right)}} \sim N(0,1)$$

Tais procedimentos podem ser realizados em vários softwares estatísticos. Em alguns destes é possível obter estas conclusões em outputs (saída) de resultados, como ocorre nos PROCs (procedimentos) do software SAS (Statistical Analysis System, SAS Institute Inc.). O software livre R possui pacotes específicos para a obtenção destes diagnósticos, dentre os quais destacamos o BOA (Bayesian Output Analysis) e o CODA (Convergence Output Diagnostic Analysis) (BEST et al., 1996). Em Albert (2009), pode-se encontrar os scripts e as explicações para a programação em R.

3.7 Medidas Adicionais

O Erro de Monte Carlo, segundo Sorensen et al. (1994), representa a variância de Monte Carlo para os estimadores Bayesianos ou, em outras palavras, a acurácia ou a qualidade dos resultados aproximados (visto que os métodos MCMC oferecem resultados aproximados para algumas integrais), informando o quanto as estimativas Bayesianas podem variar de acordo com a dimensão da cadeia de Markov (número de iterações). Conforme Van Tassel e Van Vleck (1996), o Erro de Monte Carlo é considerado pequeno se o seu valor somado à estimativa Bayesiana dos coeficientes não alterar o valor desta até a segunda casa decimal. Ainda segundo os referidos autores, baixos valores para o Erro de Monte Carlo são bons indicadores de que houve a convergência da cadeia de Markov. Este indicador pode ser calculado conforme apresentado em (23):

$$E.M.C. = Var[\widehat{g}(\beta)] = \sqrt{\left[\frac{\sum_{j=1}^R g(\beta_j) - \overline{g(\beta)}}{R} \right]^2} = \sqrt{\frac{1}{R} \left\{ \frac{[\sum_{j=1}^R g(\beta_j) - \overline{g(\beta)}]^2}{R} \right\}}$$

$$E.M.C. = Var[\widehat{g}(\beta)] = \sqrt{\frac{1}{R} \sigma^2(\beta)} = \frac{\sigma(\beta)}{\sqrt{R}} \quad (23)$$

Em que $\sigma(\beta)$ representa o desvio-padrão a posteriori e R o número de iterações.

A segunda medida, denominada como Fator de Ineficiência ou Tempo de Autocorrelação Integrado (Integrated Autocorrelation Time), segundo Green e Han (1992), indica o quanto à eficiência da cadeia de Markov é afetada pela autocorrelação presente entre as observações simuladas. Além disso, outra interpretação para essa medida é a quantificação de quanto a variância de uma estimativa (geralmente da média a posteriori) aumenta devido ao fato da amostra não ser aleatória. Logo, quanto menor o valor de \hat{f}_R mais precisa será a estimativa obtida. O cálculo para o Fator de ineficiência é apresentado em (24):

$$\hat{f}_R = 1 + \sum_{j=1}^m \left(\frac{m+1-j}{m+1} \right) \hat{\rho}_j \quad (24)$$

Em que m é o número de defasagens (lags) utilizadas para o cálculo e $\hat{\rho}_j$ é o coeficiente de autocorrelação estimado na defasagem j .

Normalmente o cálculo de \hat{f}_R é baseado no truncamento da soma em um determinado valor de m , que deve ser no mínimo 100. Rossi et al. (2005) explicam que a estrutura de autocorrelação de uma cadeia pode apresentar lags significantes após a defasagem de ordem 100 e que, em contra partida, o uso de todas as R iterações pode conduzir a estimativas tendenciosas e não consistentes de f_R . A estimativa de f_R pode ser obtida pela função `numEff` do software R, disponível no pacote `bayesm`. Na programação default, calcula-se m como:

$$m = \text{mín} \left[\dim(\beta), \frac{100}{\sqrt{5000}} \sqrt{\dim(\beta)} \right]$$

A escolha do número de graus de liberdade (ν) para a distribuição candidata t-Student multivariada é baseada na minimização do Fator de Ineficiência. Por isso, como mencionado anteriormente, Rossi et al. (2005) afirmam que quando ν varia entre 5 e 15, tem-se os melhores resultados para o Independence Metropolis Hasting. Um fato curioso é que escolher parâmetros que minimizem o fator de ineficiência é o mesmo que maximizar a eficiência numérica relativa, um terceiro critério, porém similar, que também pode ser encontrado na literatura de inferência Bayesiana.

3.8 Principais Vantagens do Enfoque Bayesiano

A metodologia Bayesiana oferece algumas vantagens em relação à Frequentista, sendo a mais notória associada à possibilidade de se incluir na análise estatística, a informação adquirida a priori a respeito dos parâmetros em estudo. Tal abordagem apresenta ainda maior flexibilidade em relação ao número de observações (tamanho da amostra), uma vez que diferentemente de como ocorre em análises estatísticas clássicas, na inferência Bayesiana não há nenhuma restrição ou conceito envolvido quanto aos graus de liberdade, ou seja, o número de parâmetros a serem estimados não precisa ser necessariamente menor que o número de observações. Além disso, o fato de considerarmos os parâmetros envolvidos como variáveis aleatórias minimiza o viés associado às estimativas (BARRETO; ANDRADE, 2004 e SILVA et al., 2006).

Outra vantagem desta metodologia é a possibilidade de se obter os chamados intervalos de credibilidade. Eles são semelhantes aos intervalos de confiança (da análise convencional), porém, são representados pelos quantis da distribuição a posteriori marginal do parâmetro de interesse. Logo, não é necessário o uso de aproximações e, ou, variâncias assintóticas para a sua construção, o que é comprovadamente uma vantagem em relação aos intervalos clássicos de confiança (O'HAGAN, 1994, BARRETO; ANDRADE, 2004 e SILVA et al., 2006). Em outras palavras, os intervalos de credibilidade são construídos especificamente para os parâmetros (que são tratados como variáveis aleatórias).

Um intervalo $C = (L, U)$, com limite inferior L e limite superior U será um intervalo de credibilidade $100(1 - \alpha)\%$ para β_i , com $i = 1, 2, \dots, r$, se $P[\beta_i \in C] \geq 1 - \alpha$. Matematicamente:

$$\int_L^U \pi(\beta_i) d\beta_i = 1 - \alpha$$

Em que $\pi(\beta_i)$ é a função densidade de probabilidade a posteriori para β_i .

Logo, com um mesmo grau de credibilidade, é possível estabelecer infinitas regiões para o parâmetro de interesse. É comum em trabalhos de inferência Bayesiana apresentar os quantis de 2,5%, 50% e 97,5%. Uma alternativa também empregada são os intervalos HPD (Highest Probability Density), definidos por Bernardo e Smith (1994) como os intervalos de maior densidade e mais curtos. Portanto, quanto mais

estrito for o intervalo HPD, maior será a credibilidade ou a certeza a respeito de β_i .

Capítulo 4

Resultados

4.1 Descrição do Conjunto de Dados

Neste trabalho, optou-se por estabelecer tratamentos hipotéticos. Os 12 tratamentos foram constituídos pela combinação em um fatorial completo de três fatores, dois com dois níveis cada e o terceiro com três níveis. Para os fatores com dois níveis, a codificação ocorreu por meio de variáveis Dummy (0 e 1) e, para o fator com três níveis como 0, 1 e 2, conforme apresentado na Tabela 3. Para a variável dependente (escolhas), estabeleceu-se o número de vezes que cada tratamento foi escolhido, conforme apresentado na Tabela 4. Especificamente para esta situação, considerou-se a abordagem básica da CBCA, em que cada avaliador escolhe apenas um único tratamento (não mais de um ou nenhum destes). Definiu-se o tamanho amostral (número de avaliadores hipotéticos) como $n = 96$, que de acordo com MacFie et al. (1989), é suficiente para estimar de modo não-viesado os efeitos principais dos três fatores em estudo (Ver no apêndice A que o referido delineamento para 12 tratamentos indica $n = 48$). Esta etapa foi desenvolvida com o auxílio do software livre R, utilizando o pacote `Conjoint` e a função `caFactorialDesign`.

Tabela 3 - Fatores, níveis e codificação

Fatores	A			B		C	
Níveis	a_1	a_2	a_3	b_1	b_2	c_1	c_2
Codificação	0	1	2	0	1	0	1

Tabela 4 - Número de vezes em que cada tratamento foi escolhido.

Tratamento	1	2	3	4	5	6	7	8	9	10	11	12	Total
Número de Escolhas	10	5	2	16	7	13	9	3	4	14	7	6	96

Por ser um estudo hipotético não foi utilizado nenhum tipo de delineamento fatorial para reduzir o número de tratamentos. No entanto, vale ressaltar que em

situações reais tal abordagem poderia ser utilizada no intuito de evitar a fadiga dos avaliadores. Na Tabela 5 são apresentadas todas as possíveis combinações entre os níveis de cada fator e, conseqüentemente, as características específicas que compõem cada um dos 12 tratamentos avaliados.

Tabela 5 - Tratamentos avaliados em um Esquema Fatorial Completo

Tratamento	A	B	C
1	a ₁	b ₁	c ₁
2	a ₂	b ₁	c ₁
3	a ₃	b ₁	c ₁
4	a ₁	b ₂	c ₁
5	a ₂	b ₂	c ₁
6	a ₃	b ₂	c ₁
7	a ₁	b ₁	c ₂
8	a ₂	b ₁	c ₂
9	a ₃	b ₁	c ₂
10	a ₁	b ₂	c ₂
11	a ₂	b ₂	c ₂
12	a ₃	b ₂	c ₂

Na Tabela 6, é apresentada a estrutura geral do conjunto de dados utilizado para o ajuste do modelo Logit Multinomial, pelas abordagens Frequentista e Bayesiana da CBCA. Apenas para recordar, o valor 1 na variável Escolha implica que aquele tratamento foi escolhido pelo avaliador. Os demais não escolhidos recebem o valor 0.

Tabela 6 - Organização do conjunto de dados para os estudos de CBCA

Avaliador	Tratamento	A	B	C	Escolha
1	1	0	0	0	0
1	2	1	0	0	0
1	3	2	0	0	1
1	4	0	1	0	0
1	5	1	1	0	0
1	6	2	1	0	0
1	7	0	0	1	0
1	8	1	0	1	0
1	9	2	0	1	0
1	10	0	1	1	0
1	11	1	1	1	0
1	12	2	1	1	0
2	1	0	0	0	0
2	2	1	0	0	0
2	3	2	0	0	0
2	4	0	1	0	0
2	5	1	1	0	0
2	6	2	1	0	0
2	7	0	0	1	0
2	8	1	0	1	0
2	9	2	0	1	1
2	10	0	1	1	0
2	11	1	1	1	0
2	12	2	1	1	0
...
...
...
...
...
...
96	1	0	0	0	0
96	2	1	0	0	0
96	3	2	0	0	0
96	4	0	1	0	0
96	5	1	1	0	0
96	6	2	1	0	1
96	7	0	0	1	0
96	8	1	0	1	0
96	9	2	0	1	0
96	10	0	1	1	0
96	11	1	1	1	0
96	12	2	1	1	0

4.2 Resultados Frequentistas para Choice-Based Conjoint Analysis

Esta etapa foi realizada com o auxílio do software livre R por meio do pacote *survival*, via função *clogit*. Antes de analisar os resultados associados à estimação dos parâmetros do modelo Logit Multinomial, avaliou-se a adequação do modelo aos dados. Foram utilizadas algumas medidas de comparação e escolha de modelos, ou seja, comparou-se por meio destas o modelo com a presença das covariáveis (completo) com um segundo modelo (reduzido), que apresentava apenas o intercepto (β_0) como variável explicativa. As medidas utilizadas foram a estimativa do logaritmo da verossimilhança de cada modelo, avaliada em seu ponto de máximo, vezes -2 (-2log L), o Critério de Informação de Akaike (AIC) (AKAIKE, 1974 e AKAIKE, 1987) e o Critério Bayesiano de Schwars (SBC)²¹ (SCHWARS, 1978 e SCLOVE, 1987), conforme apresentados na Tabela 7.

Tabela 7 - Comparação de modelos sem e com a presença de covariáveis

Critério	Modelo Sem Covariáveis	Modelo Com Covariáveis
-2Log L	1353,457	1333,711
AIC	1355,457	1339,770
SBC	1360,506	1354,919

Notou-se que o modelo parametrizado (completo ou com a presença das três covariáveis) apresentou menores valores de -2Log L, AIC e SBC. Isso implica que mesmo com um maior número de parâmetros, o que o torna menos parcimonioso, o modelo completo está mais próximo de um modelo teórico ideal, ou em outras palavras, que representa satisfatoriamente o mecanismo da geração dos dados. Logo, concluiu-se que a influência das covariáveis foi significativa.

Um segundo resultado oferecido foram os testes assintóticos da Razão de Verossimilhanças, Wald e Score. Nestes procedimentos, a estatística do teste tem distribuição qui-quadrado em grandes amostras e a hipótese nula é de que os coeficientes paramétricos são iguais a zero ($H_0: \beta_1 = \beta_2 = \beta_3 = 0$). Portanto, a sua

²¹ Também conhecido como Critério de Informação Bayesiano (BIC de Bayesian Information Criterion)

rejeição implica que pelo menos um destes difere de zero ou é significativo. O nível de significância (α) adotado foi de 5%. Na Tabela 8 estão apresentados os resultados para estes testes (graus de liberdade, valor da estatística qui-quadrado e valor-p).

Tabela 8 - Testes de hipóteses globais para os três coeficientes do modelo

Testes	G.L	$\chi^2_{v=3}$	valor-p
Razão de Verossimilhanças	3	19,69	0,000197
Wald	3	18,88	0,000290
Score	3	19,42	0,000224

Como o valor-p associado aos três testes foi inferior a 0,05, rejeitou-se a hipótese nula a 5% de probabilidade. As estimativas pontuais Frequentistas para os parâmetros do modelo Logit Multinomial e o teste de Wald para a significância dos coeficientes estimados estão apresentados na Tabela 9.

Tabela 9 - Estimação dos parâmetros do modelo pela abordagem Frequentista e o teste de Wald.

Fator	Estimativas dos β's	Erro Padrão	Z	valor-p
A	-0,3842	0,1296	-2,964	0,00304*
B	0,6466	0,2149	3,009	0,00262*
C	-0,2091	0,2052	-1,019	0,30832

* Significativo a 5% de probabilidade

Verificou-se pela Tabela 9 que os fatores “A” e “B” influenciam significativamente a escolha do consumidor no momento da escolha do tratamento (valor-p < 0,05) e que o fator “C” não (valor-p > 0,05).

Após obter as estimativas pontuais dos parâmetros do modelo Logit Multinomial, foram estimadas as probabilidades de escolha associadas a cada um dos $\mathbf{J} = 12$ tratamentos avaliados. Novamente, ressalta-se que este é um estudo hipotético, cujo delineamento fatorial utilizado foi o do tipo completo, no qual todos os tratamentos são avaliados. No entanto, se este fosse um estudo real, e um delineamento fatorial fracionado tivesse sido empregado utilizando-se, por exemplo, apenas 6 tratamentos (1/2 dos tratamentos), após o conhecimento das estimativas pontuais dos parâmetros seria possível estimar as probabilidades dos mesmos $\mathbf{J} = 12$ tratamentos (inclusive dos 6

tratamentos que não tivessem sido avaliados), visto que as utilidades e consequentemente, as probabilidades de escolha são calculadas baseadas no modelo que define uma combinação linear dos parâmetros. Esta é uma das principais vantagens da metodologia CBCA em relação a uma análise, por exemplo, de frequências relativas. Na Tabela 10 estão apresentadas as probabilidades estimadas pelo modelo Logit Multinomial (Prob. Est.) e também as probabilidades observadas (Prob. Obs.), obtidas por meio das frequências relativas de escolhas.

Tabela 10 - Probabilidades observadas e estimadas pelo modelo Logit Multinomial

Tratamento	A	B	C	Prob. Obs.	Prob. Est.
1	a ₁	b ₁	c ₁	0,1041	0,0884
2	a ₂	b ₁	c ₁	0,0520	0,0602
3	a ₃	b ₁	c ₁	0,0208	0,0410
4	a ₁	b ₂	c ₁	0,1666	0,1689
5	a ₂	b ₂	c ₁	0,0729	0,1150
6	a ₃	b ₂	c ₁	0,1354	0,0783
7	a ₁	b ₁	c ₂	0,0937	0,0717
8	a ₂	b ₁	c ₂	0,0312	0,0488
9	a ₃	b ₁	c ₂	0,0416	0,0332
10	a ₁	b ₂	c ₂	0,1458	0,1370
11	a ₂	b ₂	c ₂	0,0729	0,0933
12	a ₃	b ₂	c ₂	0,0625	0,0635

Verificou-se que para os tratamentos avaliados, as probabilidades estimadas e observadas foram muito próximas, apresentando diferenças numéricas de ordem decimal. Apenas para os tratamentos 5 e 6 constataram-se diferenças um pouco maiores, explicadas devido ao fato das probabilidades observadas estarem sujeitas a fatores aleatórios não controláveis no estudo.

Por fim, foram avaliadas as Razões de Escolhas (R.E.), no intuito de verificar quais os níveis de cada fator atraíram mais a atenção ou foram probabilisticamente preferidos pelo consumidor na escolha do tratamento.

Razão de Escolha para os níveis do fator A:

$$RE(1,2) = \frac{P(Y_n = 1)}{P(Y_n = 2)} = \frac{0,08848}{0,06026} = 1,46$$

Mantendo-se fixos os níveis dos fatores B (b_1) e C (c_1) e modificando apenas os níveis do fator A (a_1 para a_2), pode-se concluir que o tratamento 1, com nível a_1 , é 1,4683 vezes mais provável de ser escolhido do que o tratamento 2, cujo nível é a_2 . Repare que o mesmo resultado poderia ser obtido se fossem comparados os tratamentos 4 e 5, 7 e 8 ou 10 e 11.

$$RE(1,3) = \frac{P(Y_n = 1)}{P(Y_n = 3)} = \frac{0,08848}{0,04104} = 2,15$$

Mantendo-se fixos os níveis dos fatores B (b_1) e C (c_1) e modificando apenas os níveis do fator A (a_1 para a_3), pode-se concluir que o tratamento 1, com nível a_1 , é 2,1559 vezes mais provável de ser escolhido do que o tratamento 3, cujo nível é a_3 . Repare que o mesmo resultado poderia ser obtido se fossem comparados os tratamentos 4 e 6, 7 e 9 ou 10 e 12.

Por meio destes dois resultados já foi possível concluir que o tratamento com nível a_1 (tratamento 1) foi probabilisticamente preferido pelo consumidor em relação aos demais (tratamento 2 e 3), com níveis, respectivamente, a_2 e a_3 .

$$RE(2,3) = \frac{P(Y_n = 2)}{P(Y_n = 3)} = \frac{0,06026}{0,04104} = 1,46$$

Mantendo-se fixos os níveis dos fatores B (b_1) e C (c_1) e modificando apenas os níveis do fator A (a_2 para a_3), pode-se concluir que o tratamento 2, com nível a_2 , é 1,4683 vezes mais provável de ser escolhido do que o tratamento 3, cujo nível é a_3 . Repare que o mesmo resultado poderia ser obtido se fossem comparados os tratamentos 5 e 6, 8 e 9 ou 11 e 12.

Razão de Escolha para os níveis do fator B:

$$RE(1,4) = \frac{P(Y_n = 1)}{P(Y_n = 4)} = \frac{0,08848}{0,16892} = 0,52$$

Como $RE(1,4)$ foi menor do que 1, isso implica que a Probabilidade de se escolher o tratamento 1 é menor do que a probabilidade de se escolher o tratamento 4, neste sentido, torna-se mais interessante avaliar essa razão conforme apresentado abaixo:

$$RE(4,1) = \frac{1}{RE(1,4)} = \frac{P(Y_n = 4)}{P(Y_n = 1)} = \frac{0,16892}{0,08848} = 1,90$$

Logo, mantendo-se fixos os níveis dos fatores A (a_1) e C (c_1) e modificando apenas os níveis do fator B (b_2 para b_1), pode-se concluir que o tratamento 4, com nível b_2 , é 1,9091 vezes mais provável de ser escolhido do que o tratamento 1, com nível b_1 . Repare que o mesmo resultado poderia ser obtido se fossem comparados os tratamentos 5 e 2, 6 e 3, 10 e 7, 11 e 8 ou 12 e 9.

Razão de Escolha para os níveis do fator C:

Embora o fator C não tenha sido significativo (o que torna este resultado desprezível, visto que a mudança de níveis não irá influenciar para maior aceitação do tratamento) também apresentaremos e interpretaremos a Razão de Escolhas para este fator.

$$RE(1,7) = \frac{P(Y_n = 1)}{P(Y_n = 7)} = \frac{0,08848}{0,07179} = 1,23$$

Mantendo-se fixos os níveis dos fatores A (a_1) e B (b_1) e modificando apenas os níveis do fator C (c_1 para c_2), pode-se concluir que o tratamento 1, com nível c_1 , é 1,2324 vezes mais provável de ser escolhido do que o tratamento 7, com nível c_2 . Repare que o

mesmo resultado poderia ser obtido se fossem comparados os tratamentos 2 e 8, 3 e 9, 4 e 10, 5 e 11 ou 6 e 12.

4.3 Resultados Bayesianos para Choice-Based Conjoint Analysis

Nesta seção, os resultados da abordagem Bayesiana para a CBCA, foram obtidos com o pacote bayesm do software livre R e sua função `rmnlIndepMetrop`. Utilizou-se o pacote BOA para estudar as cadeias de Markov e obter as inferências a respeito dos parâmetros e das Probabilidades e Razões de Escolhas dos tratamentos. Além disso, tal pacote foi empregado para avaliar a convergência.

Para implementação do enfoque Bayesiano utilizou-se uma distribuição a priori normal tri-variada pouco informativa, conforme mencionado no capítulo anterior. Baseado em uma amostra piloto, calculada conforme em (22), simulou-se os resultados iniciais com $R = 3746$ iterações, $\text{thin} = 1$ e $\text{burn-in} = 0$. Posteriormente, verificou-se o critério de Raftery-Lewis. Os resultados deste critério estatístico estão apresentados na Tabela 11.

Tabela 11 - Resultados Iniciais do critério de Raftery-Lewis

Fator	thin	Burn-in	R	Lower Bound	Fator de Dependência
A	2	4	7.468	3746	1,9936
B	1	3	4.165	3746	1,1119
C	2	4	7.150	3746	1,9087

Nota-se que com poucas iterações (menos de 10.000) as cadeias parecem convergir para suas respectivas distribuições estacionárias, visto que o Fator de Dependência mostra-se menor do que 5. Isso pode ser explicado devido à ergodicidade geométrica (que implicam em taxas geométricas de convergência) das cadeias geradas pelo MCMC empregado, principalmente pelo uso de uma priori normal e das caudas exponenciais presentes na densidade a posteriori, que surgem devido à aproximação de Laplace para a função de verossimilhança.

Neste sentido, optou-se por trabalhar com os seguintes valores: $R = 24.000$ iterações, $\text{thin} = 2$ (amostras selecionadas 2 a 2) e um $\text{Burn-in} = 2.000$. Isso implica que serão selecionadas para compor a distribuição a posteriori marginal dos parâmetros apenas as últimas 10.000 iterações. A superestimação do número de iterações e de Burn-in tem por finalidade eliminar quaisquer amostras indevidas no início da cadeia e trabalhar com um conjunto de no mínimo 10.000 observações simuladas.

Na Tabela 12 estão apresentados os resultados iniciais da CBCA, obtidos pela implementação do MCMC conforme a definição dos parâmetros mencionados acima. São apresentados a média, a moda e o desvio padrão a posteriori, assim como o Erro de Monte Carlo e intervalos com 95% de credibilidade para os betas (quantis 2,5% e 97,5%) e o valor do quantil 50% (mediana a posteriori).

Tabela 12 - Resultados da distribuição a posteriori dos betas para a CBCA via método de estimação Bayesiano

Fatores	Média	Moda	Mediana	Desvio Padrão	Erro de M.C.	$Q_{2,5\%}$	$Q_{97,5\%}$
A	-0,3875	-0,3518	-0,3857	0,1307	0,0014	-0,6460	-0,1347
B	0,6552	0,5889	0,6529	0,2165	0,0019	0,2390	1,0889
C	-0,2116	-0,1654	-0,2110	0,2052	0,0024	-0,6156	0,1933

Conclui-se que as estimativas Bayesianas de média, mediana e moda a posteriori de cada fator foram muito semelhantes às Frequentistas. Tal fato já era esperado, devido à aproximação de Laplace para a função de verossimilhança e o uso de uma priori normal pouco informativa, o que implica que a moda a posteriori converge para o estimador de Máxima Verossimilhança Parcial. Mais especificamente, nesta situação, quaisquer destas três medidas podem ser tratadas como o estimador Bayesiano, pois o tamanho da amostra utilizada foi suficientemente grande para tornar a distribuição a posteriori marginal de cada parâmetro aproximadamente uma normal (simétrica e com Média = Mediana = Moda). Logo, as marginais de cada fator serão normais uni-variadas e especificadas como:

$$\pi(\beta_i) \sim N\left(\mu_i = \hat{\beta}_i, \sigma_i^2 = H_{i,i}^* \Big|_{\beta = \hat{\beta}}\right).$$

Por meio da função `mnlHess` do pacote `bayesm`, foi computado o negativo do valor esperado da Matriz Hessiana²². Posteriormente, foram plotadas as amostras simuladas de cada parâmetro em um histograma de probabilidades e, construídos sobrepostos a estes, as curvas das distribuições normais conforme a especificação a seguir. A Figura 1 ilustra os resultados.

$$\hat{\beta} = (-0,3841; 0,6466; -0,2090)'$$

²² Destacamos ao leitor que para construir as curvas no software R, com a função `curve`, a especificação da dispersão da distribuição ocorre por meio do desvio-padrão e não da precisão, portanto, é preciso encontrar a inversa da matriz H^* e tomar a raiz quadrada dos valores de sua diagonal (variâncias), conforme apresentado nos códigos disponíveis no Apêndice D.

$$H^* = \begin{bmatrix} 64 & 0 & 0 \\ 0 & 24 & 0 \\ 0 & 0 & 24 \end{bmatrix}$$

Portanto,

$$\pi(A) = \pi(\beta_1) \sim N(-0,3841; 64)$$

$$\pi(B) = \pi(\beta_2) \sim N(0,6466; 24)$$

$$\pi(C) = \pi(\beta_3) \sim N(-0,2090; 24)$$

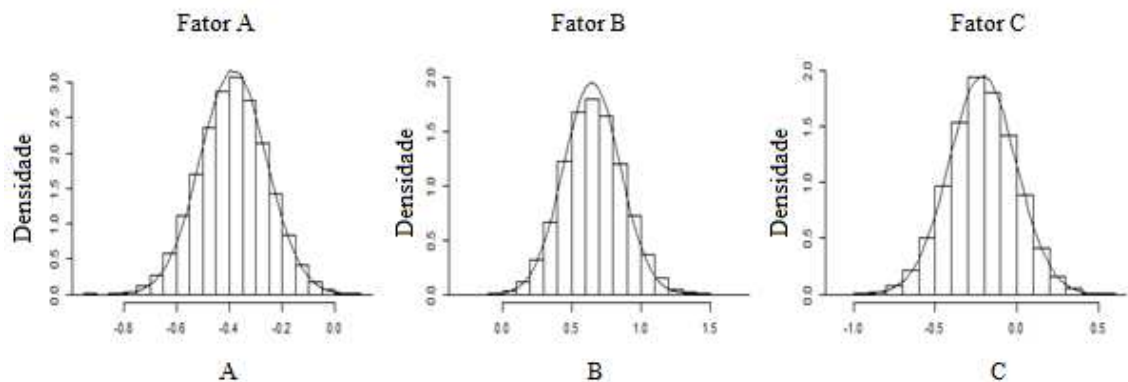


Figura 1 – Histograma de Probabilidades e densidades esperadas

Observou-se ainda que o desvio-padrão a posteriori foi similar à estimativa do erro padrão de cada coeficiente, quando confrontado com a abordagem Frequentista. Através do Erro de Monte Carlo, foi possível concluir que os resultados Bayesianos foram considerados satisfatórios e confiáveis, visto que os valores obtidos foram pequenos quando somados à estimativa Bayesiana (média, mediana ou moda a posteriori), não alterando o seu resultado. Este fato garantiu que o número de iterações foi apropriado, sendo um bom indicativo da convergência das cadeias.

Na Figura 2 estão apresentados os gráficos que descrevem o comportamento das cadeias de Markov simuladas para cada um dos fatores (após a eliminação das 2.000 amostras iniciais), que auxiliam a verificar se houve a convergência. Além disso, foram também apresentados os gráficos das Funções de Autocorrelação (ACF), que são úteis para verificar o grau de autocorrelação presente na amostra simulada.

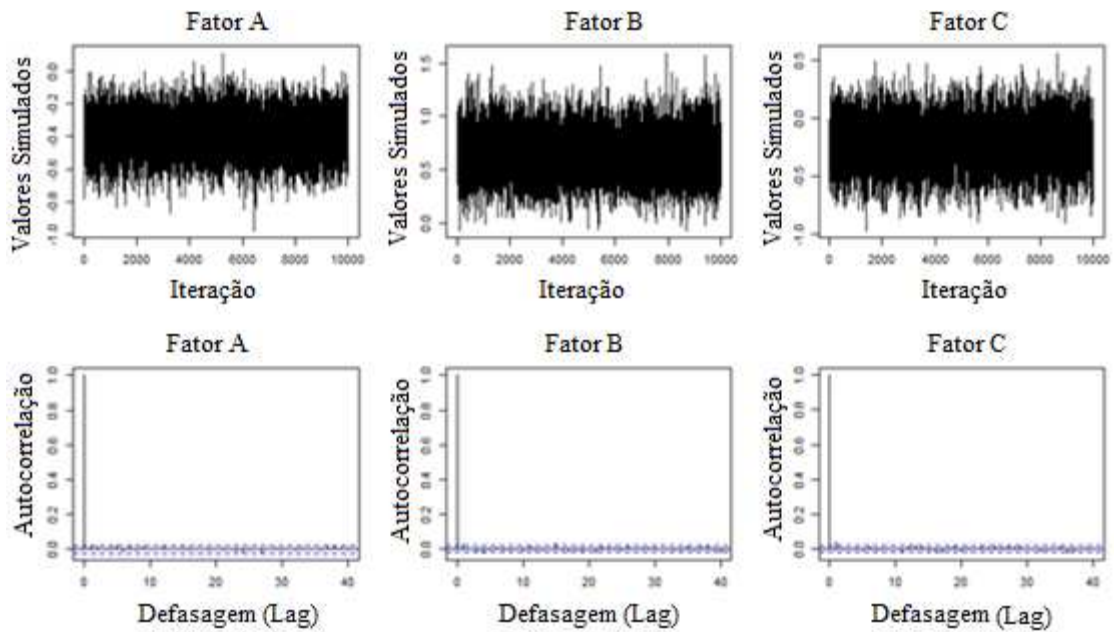


Figura 2 - Gráficos para as distribuições a posteriori marginais dos parâmetros

Conforme observado na Figura 2, existiu forte evidência de que ambas as cadeias alcançaram a convergência, visto que o conjunto de amostras pareceu constituir distribuições estacionárias perante as iterações realizadas. Graficamente, entende-se que uma distribuição é estacionária se seus valores oscilam em torno de um valor médio constante, refletindo alguma forma de equilíbrio. Tal conclusão foi corroborada pelos resultados da Tabela 13, por meio da análise do valor-p para o critério de Geweke, ambos superiores a 0,05, o que comprovou, estatisticamente, que houve a convergência.

Tabela 13 - Verificação da convergência das cadeias pelo critério de Geweke.

Fator	A	B	C
Z-Score	-0,7047	0,8895	-0,9675
valor-p	0,4809	0,3737	0,3332

Ainda conforme a Figura 2 notou-se que as amostras simuladas apresentaram baixo grau de autocorrelação, visto que em cada ACF não houve a presença de muitos lags ou defasagens que ultrapassassem os limites de significância propostos por Bartlett (1946). Este foi um ótimo resultado, pois indicou que a amostra simulada se aproximou de uma amostra aleatória. No entanto, tal conclusão já era esperada, visto que as cadeias de Markov foram geradas pelo algoritmo Independence Metropolis Hasting, que segundo Rossi et al. (2005), para este tipo de análise, é mais eficiente em relação a

outros algoritmos, como por exemplo, o Random-Walk Metropolis Hasting. Estes autores apresentam a seguinte justificativa baseados em alguns de seus resultados experimentais:

“In comparison to the Random-Walk, the independence Metropolis functions perform much more efficiently with an acceptance rate of 0.70 and a numerical efficiency of 1.44 or barely less than iid sampling. Differences between the ACFs of the RW and independence chains result in a numerical efficiency ratio of 3:1.”

Nesta análise, a taxa de aceitação para o algoritmo foi de 0,8548 e o fator de ineficiência f_R para os fatores A, B e C, foram, respectivamente, de 1,2920, 1,0106 e 1,0611. Tais resultados foram calculados com 141 defasagens e demonstraram que a variância da média a posteriori dos fatores não foi demasiadamente superior ao seu valor pontual. Como a amostra simulada para cada fator pode ser considerada aleatória (devido ao baixo grau de autocorrelação), as estimativas de eficiência relativa foram adequadas, visto que seriam esperados valores menores que 1,44.

Na Figura 3 foi demonstrado que após poucas iterações, o comportamento das médias ergódicas tornou-se assintótico e convergiu para os respectivos valores esperados, fato explicado, principalmente, devido à ergodicidade geométrica da cadeia. Notou-se ainda que o Burn-in utilizado (de 2.000) eliminou qualquer possibilidade de que a amostra tomada como base para a realização de inferências contenha valores inadequados ou obtidos antes do aquecimento da cadeia.

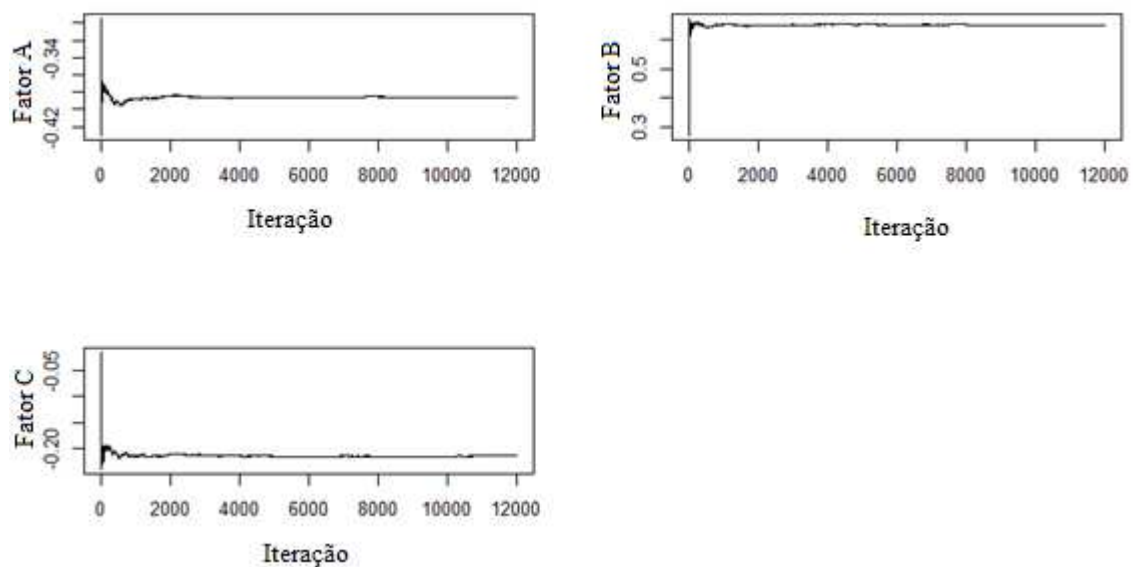


Figura 3 - Comportamento de Convergência das cadeias de Markov

No contexto da inferência Bayesiana, para verificar se uma determinada estimativa foi significativa ou não, pode-se recorrer aos intervalos de credibilidade e realizar tal análise de forma similar à abordagem Frequentista, por meio de intervalos de confiança. Em CBCA, o interesse consiste em testar a hipótese de que cada estimativa Bayesiana, associada a um determinado fator, é igual a zero. Neste sentido, basta avaliar se o valor de interesse (zero) encontra-se dentro dos respectivos intervalos com 95% de credibilidade (determinado pelos quantis 2,5% e 97,5%).

De acordo com os resultados obtidos na Tabela 12, repare que o valor zero não se encontrou presente dentro dos intervalos de credibilidade para os fatores A e B, o que implica que suas estimativas Bayesianas (média, mediana ou moda a posteriori) divergem estatisticamente de zero. Para o fator C, notou-se que o valor 0 encontrou-se dentro dos intervalos de credibilidade. Logo, concluiu-se que a estimativa Bayesiana associada a este coeficiente foi igual a 0. Tais conclusões foram similares às obtidas por meio do teste de hipótese de Wald na abordagem Frequentista. A Figura 4 ilustra este resultado. Nesta, tomou-se a média a posteriori (apenas para caráter ilustrativo) como valor central e os quantis de 2,5% e 97,5% como, respectivamente, os limites inferiores e superiores dos intervalos de credibilidade. A linha contínua indica o valor a ser testado, equivalente à zero.

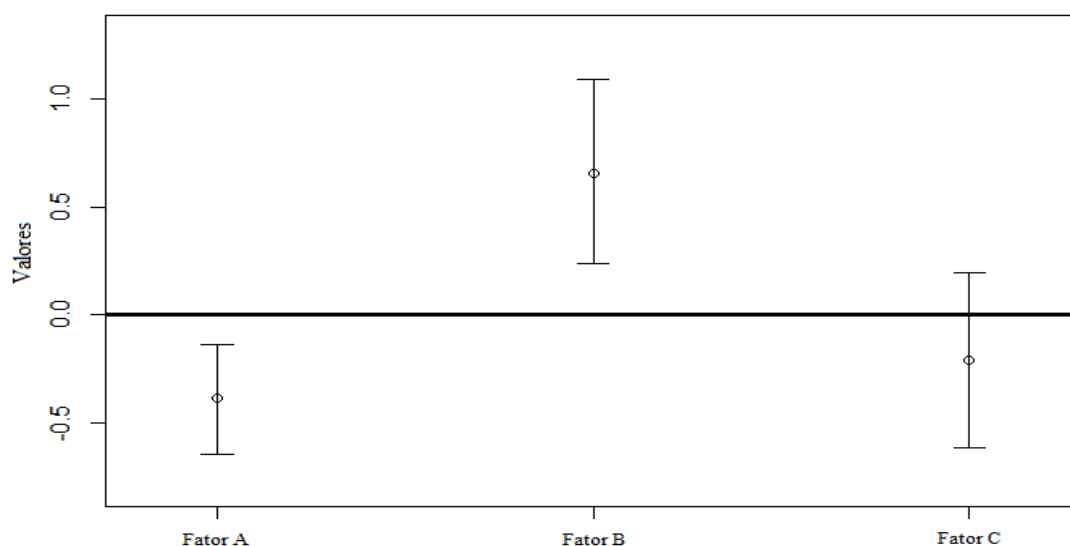


Figura 4 – Teste de Hipótese Bayesiano para os fatores (Quantil de 2,5%, Média a posteriori e Quantil de 97,5%)

Para avaliar as probabilidades de escolha, associadas a cada um dos $J = 12$ tratamentos, foram estabelecidas doze novas variáveis aleatórias, uma para cada probabilidade. Se há uma variável aleatória referente a cada parâmetro e composta por

10.000 observações (valores simulados das distribuições a posteriori marginais dos fatores), dado que as utilidades e probabilidades de escolha são funções diretas ou combinações lineares destes parâmetros, foi possível construir distribuições para estas quantidades procedendo a operações matemáticas básicas sobre estas variáveis aleatórias. A Tabela 14 apresenta os resultados Bayesianos para as probabilidades de escolha.

Tabela 14 - Resultados Bayesianos para as Probabilidades de Escolha.

Tratamento	Média	Moda	Mediana	Desvio Padrão	Erro de M.C.	Q_{2,5%}	Q_{97,5%}
1	0,0885	0,0905	0,0873	0,0172	0,0001	0,0585	0,1257
2	0,0599	0,0588	0,0593	0,0103	0,0001	0,0414	0,0813
3	0,0412	0,0392	0,0403	0,0097	0,0001	0,0248	0,0626
4	0,1692	0,1680	0,168	0,0258	0,0002	0,1221	0,2231
5	0,1145	0,1122	0,1141	0,014	0,0001	0,0880	0,1429
6	0,0788	0,0711	0,0777	0,0158	0,0001	0,0511	0,1125
7	0,0718	0,0681	0,0708	0,0148	0,0001	0,0459	0,1036
8	0,0486	0,0474	0,0481	0,0089	0,0001	0,0325	0,0674
9	0,0334	0,0326	0,0327	0,0082	0,0001	0,0197	0,0516
10	0,1373	0,1298	0,1361	0,0229	0,0001	0,0963	0,1858
11	0,0929	0,0904	0,0925	0,0129	0,0001	0,0688	0,1190
12	0,0639	0,0621	0,063	0,0134	0,0001	0,0404	0,0932
Total	1,0000	0,9707	0,9899				

De acordo com a Tabela 14, a soma dos valores da mediana e da moda a posteriori não resultou na unidade, fato que violou um dos principais conceitos axiomáticos de probabilidade. Neste sentido, optou-se por utilizar a média a posteriori como estimador Bayesiano para as probabilidades de Escolha. Verificou-se que as médias a posteriori assumem valores próximos aos obtidos no contexto Frequentista para as probabilidades de escolha. Esta similaridade entre os resultados Frequentista e Bayesiano permitiu concluir a respeito da eficiência do enfoque proposto para CBCA, visto que um dos principais resultados associados a tal metodologia não divergiu ou apresentou valores discrepantes, portanto, não comprometeram as interpretações da análise. Notou-se ainda que a média, mediana e moda a posteriori são muito próximas, o

que indicou que a distribuição das probabilidades de escolha parece ser simétrica, conforme ilustrado na Figura 5.

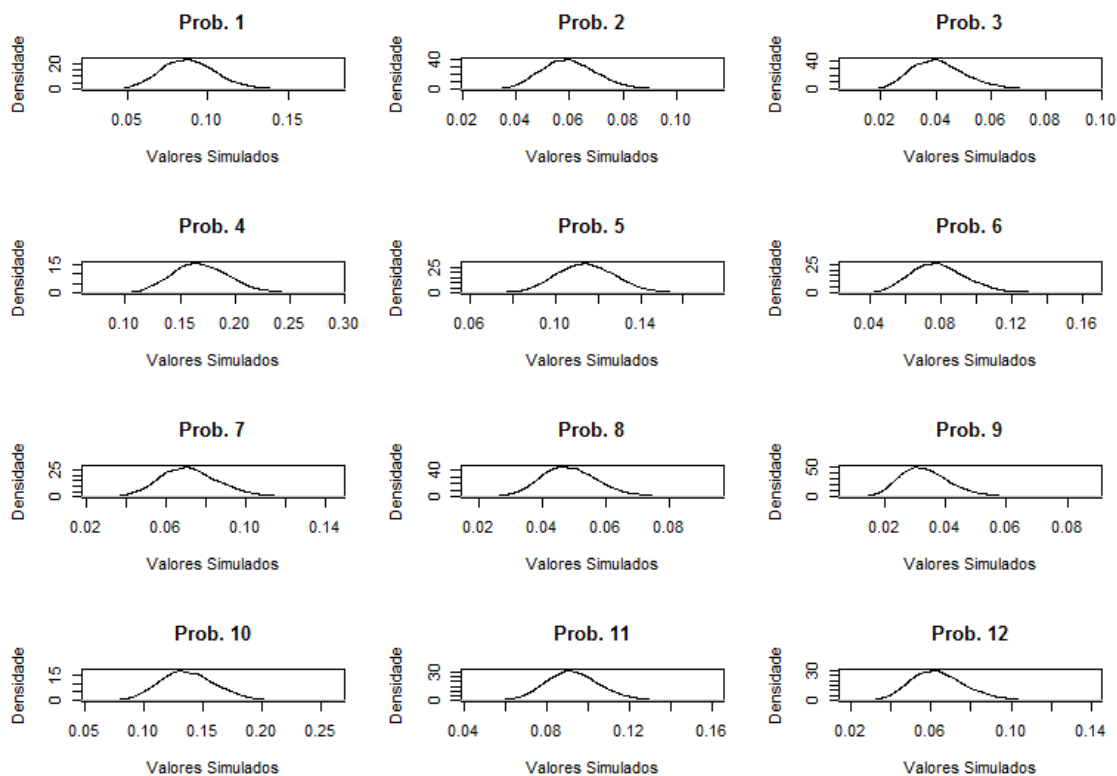


Figura 5 - Densidades a posteriori estimadas para cada uma das probabilidades de escolha

Um aspecto interessante associado ao enfoque Bayesiano para CBCA, é que ele disponibiliza sem grandes dificuldades o desvio-padrão a posteriori e os intervalos de credibilidade para cada probabilidade de escolha (Tabela 14). Pela abordagem Frequentista não é uma tarefa simples obter estimativas de erros-padrão e intervalos de confiança para as probabilidades de escolha, visto que a distribuição destas quantidades não é conhecida. Uma alternativa seria por meio de métodos computacionais de reamostragem, como jackknife e bootstrap.

Neste sentido, de um ponto de vista prático, tais resultados poderiam ser utilizados para testar a hipótese de que duas ou mais probabilidades de escolha são estatisticamente iguais. Esta análise pode ser realizada graficamente, plotando as estimativas Bayesianas de probabilidade e seus respectivos intervalos de credibilidade. Se houver sobreposição dos limites de credibilidade é possível concluir que não há divergência estatística entre as probabilidades. Caso contrário, que há.

Baseado nos resultados deste trabalho imagine a seguinte situação hipotética. Um empresário encontra-se em dúvida sobre qual tratamento investir dentre os que obtiveram as maiores probabilidades de escolha, ou seja, os tratamentos 4, 5 e 10 (respectivamente: 0,1692; 0,1145 e 0,1373). Suponha ainda que cada um destes tratamentos gere custos unitários, respectivamente, iguais a R\$7,00 (4), R\$5,00 (5) e R\$3,00 (10). Repare que se for possível concluir sobre a igualdade (estatística) das probabilidades de escolha, este empresário poderia investir na produção do tratamento com preço unitário de produção mais baixo, visto que ambos seriam igualmente prováveis de serem escolhidos pelos consumidores. Em uma alta escala de produção, esta decisão poderia lhe render uma grande economia financeira. A Figura 6 apresenta os resultados gráficos que permitem realizar esta comparação.

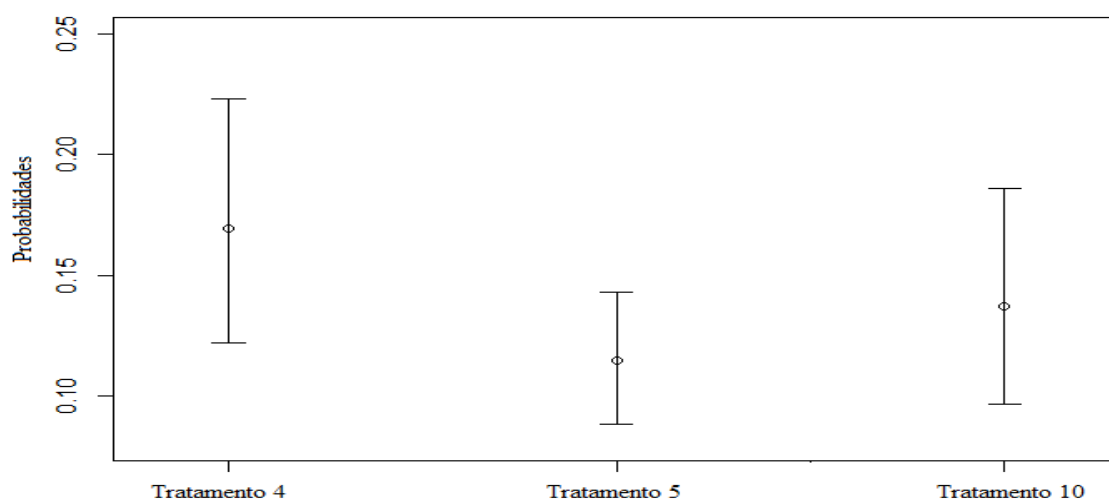


Figura 6 - Comparação entre as probabilidades de escolha para os tratamentos 4, 5 e 10

Logo, é possível concluir que, estatisticamente, as probabilidades de escolha associadas aos tratamentos 4, 5 e 10 são iguais, visto que existe sobreposição dos limites de credibilidade, ou seja, $P(Y_n = 4|\mathbf{X}) = P(Y_n = 5|\mathbf{X}) = P(Y_n = 10|\mathbf{X})$. Isso implica que o empresário poderia investir no tratamento 10, que possui o menor preço unitário de produção e, também, em termos estatísticos, a mesma probabilidade de escolha do tratamento 4, que está associado ao maior valor de probabilidade.

Outro detalhe é que as Probabilidades e Razões de escolha poderiam ser calculadas utilizando apenas as estimativas Bayesianas dos parâmetros (média, mediana ou moda a posteriori), ou seja, pelo produto matricial destas estimativas pela matriz de delineamento do experimento \mathbf{X} , posteriormente, tomando-se os exponenciais, divisões etc., como ocorre na abordagem Frequentista.

Para finalizar esta seção, são apresentados na Tabela 15 os resultados do enfoque Bayesiano para as Razões de Escolha. Ressalta-se que também foram construídas distribuições para tais quantidades, procedendo a operações matemáticas sobre as variáveis aleatórias referentes às probabilidades de escolha.

Tabela 15 - Estimativas Bayesianas para as Razões de Escolhas.

Razão de Escolhas	Média	Moda	Mediana	Desvio Padrão	Erro de M.C.	$Q_{2,5\%}$	$Q_{97,5\%}$
<i>RE(1,2)</i>	1,4847	1,4842	1,4710	0,1949	0,0010	1,1436	1,9069
<i>RE(1,3)</i>	2,2424	2,0212	2,1638	0,5983	0,0029	1,3078	3,6361
<i>RE(2,3)</i>	1,4847	1,4842	1,4710	0,1949	0,0010	1,1436	1,9069
<i>RE(4,1)</i>	1,9723	1,7756	1,9191	0,4359	0,0026	1,2642	2,9684
<i>RE(1,7)</i>	1,2617	1,1614	1,2368	0,2627	0,0015	0,8251	1,8515

Avaliando os resultados da Tabela 15, notou-se que ao representar as estimativas Bayesianas pela mediana a posteriori (quantil 0,5), ocorreram resultados mais semelhantes com os obtidos no contexto Frequentista. A média e a moda a posteriori não apresentaram valores tão diferentes dos valores medianos, ao ponto de descaracterizar os resultados. No entanto, conforme visualizado na Figura 6, a distribuição da Razão de Escolha tende a apresentar um comportamento assimétrico à direita, ou seja, com o valor da Moda < Mediana < Média. Este comportamento fica claramente evidenciado em *RE(1,3)*, *RE(4,1)* e *RE(1,7)*, conforme os valores da Tabela 15 e as densidades estimadas na Figura 7.

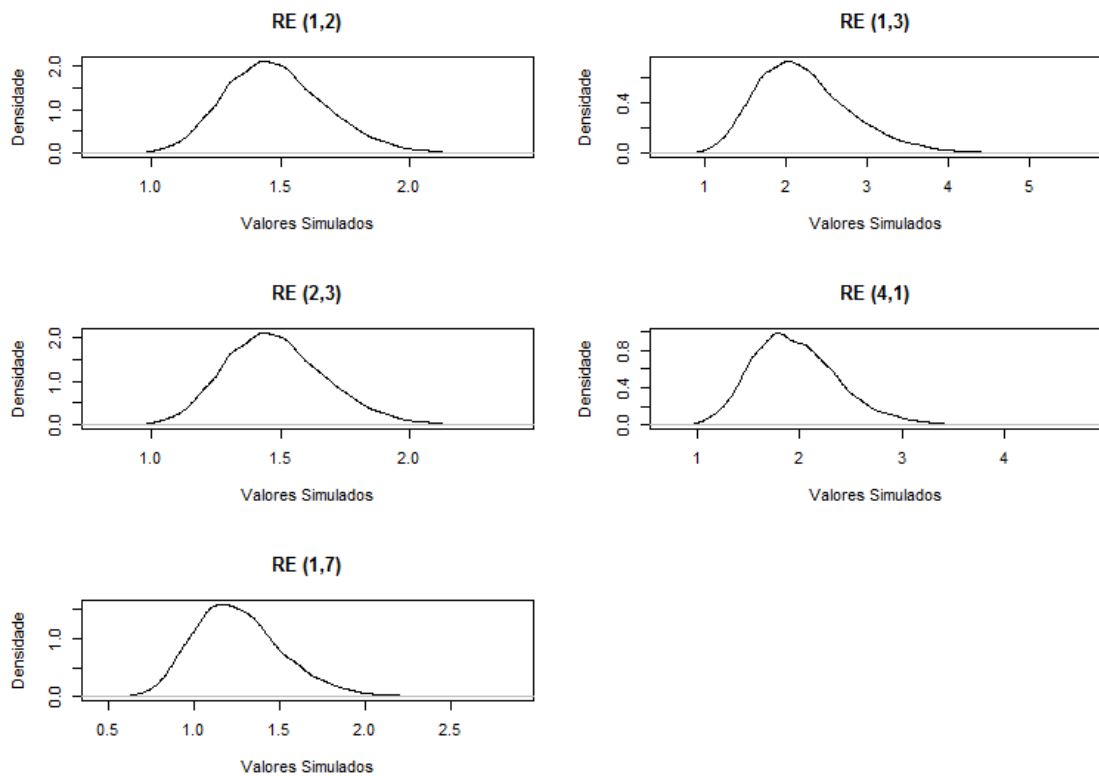


Figura 7 - Densidades a posteriori estimadas para as razões de escolha.

Pelo enfoque Bayesiano, foi possível obter o desvio-padrão a posteriori da distribuição de cada razão de escolha e também os respectivos intervalos de credibilidade. Note ainda que os valores da média, da mediana e da moda a posteriori para $RE(1,7)$ localizam-se entre os quantis de 2,5% e 97,5%. Pode-se concluir, portanto, que estatisticamente, $RE(1,7) = 1$. Como $RE(1,7)$ retrata a mudança de níveis do fator C, que não foi significativo, este resultado já era esperado.

Considerações Finais

Foi deduzida e demonstrada a implementação do enfoque Bayesiano para a Choice-Based Conjoint Analysis (CBCA). Os resultados indicaram que a metodologia proposta, embora recente, apresentou resultados satisfatórios e muito próximos aos da abordagem Clássica (Frequentista), amplamente mencionada e empregada na literatura da CBCA. Além disso, pelo enfoque Bayesiano foi possível obter intervalos de credibilidades para os parâmetros (fatores) de interesse, probabilidades e razões de escolha, permitindo ao pesquisador testar hipóteses sobre tais quantidades, o que é uma vantagem em relação à abordagem Frequentista, principalmente no que se refere às probabilidades de escolha.

Para futuros trabalhos, sugere-se que tal enfoque seja implementado de diferentes formas, como por exemplo, verificando os resultados que seriam obtidos se outras distribuições de probabilidades fossem utilizadas como a priori, além da função de verossimilhança original do modelo Logit Multinomial (sem a aproximação de Laplace). Tal procedimento poderia ser realizado em pacotes Bayesianos específicos, como o WinBUGS ou OpenBUGS, disponíveis no software R.

Um segundo ponto interessante seria o desenvolvimento de estudos por simulação de dados que avaliassem a eficiência do enfoque Bayesiano em relação ao Frequentista em diferentes situações práticas, visto que conforme explicado neste trabalho, pela metodologia aqui proposta, os resultados Bayesianos tendem a convergir para os Frequentistas, principalmente para grandes amostras. Neste sentido, simulações poderiam ser conduzidas em situações hipotéticas, em que o tamanho da amostra é pequeno, para diferentes números de fatores e níveis, presença de interações, experimentos com mais de um choice set (ou conjunto de escolhas, que utilizam delineamentos não ortogonais), entre outras situações.

Destaca-se que em estudos de CBCA em que a metodologia Hierárquica Bayesiana (CBC-HB) é empregada, estima-se um vetor de parâmetros para cada indivíduo da amostra. Logo, trabalhos que visem à comparação dos resultados em nível individual (por exemplo, a média das estimativas individuais) com os resultados agregados poderiam auxiliar a verificar qual abordagem é mais interessante e representativa no contexto de CBCA.

Referências Bibliográficas

AAKER, D. A.; KUMAR, V.; DAY, G. S. **Marketing research**. 7th edition. New York: John Wiley & Sons, 2001. 751 p.

AIZAKI, H., NISHIMURA, K. Design and analysis of choice experiments using R: a brief introduction. **Agricultural Information Research**, v. 17, n. 2, p. 86-94, 2008.

AKAIKE, H. **A new look at the statistical model identification**, IEEE Transactions on Automatic Control, 1974.

AKAIKE, H. Factor Analysis and AIC. **Psychometrika**, v. 52, n. 3, p. 317-332, 1987.

ALBERT, J. **Bayesian Computation with R**. Ed. Springer. 2009. 312 p.

ARTES, R. **Análise de Preferência “Conjoint Analysis”**. 1991. 189 p. Tese (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1991.

BARRETO, G.; ANDRADE, M. G. Robust Bayesian Approach for AR(p) Models Applied to Streamflow Forecasting. **Journal Applied Statistical Science**, New York, v. 12, n. 3, p. 269-292, 2004.

BARTLETT, M.S. On the theoretical specification and sampling properties of auto correlated time-series. **Supplement to the Journal of the Royal Statistical Society**, v. 8, n. 1, p. 27-41, 1946.

BASTOS, F. S. **Análise conjunta de Fatores Baseada em Escolhas: Estimção e Inferências**. 2010. 82 p. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, 2010.

BAYES, T. An essay towards solving in the doctrine of chances. **Philosophical Transactions of the Royal Society London**, v. 53, p. 370–418, 1763.

BEM-AKIVA, M.; S. R. Lerman. **Discrete Choice Analysis**. MIT Press, Cambridge, MA, 1985.

BERNARDO, J.; SMITH, A. **Bayesian theory**. New York: John Wiley and Sons, 586 p., 1994.

BEST, N.; COWLES, M. E V.; K. **CODA – Convergence diagnostics and output analysis software for Gibbs sampling output: Version 0.4**. Cambridge: MRC Biostatistics Units, 1997.

BOX, G. E. P.; TIAO, G. C. **Bayesian Inference in Statistical Analysis**. Wiley Classics Library ed. Wiley-Interscience, 1992, 608 p.

BRESLOW, N. E. Covariance Analysis of Censored Survival Data, **Biometrics**, v. 30, p. 89–99, 1974.

CARDELLO, A. V. **Consumer expectations and their role in food acceptance**. In: MacFIE, H. J. H.; THOMSON, D. M. H. Measurement of food preferences. Glasgow: Blackie Academic & Professional, 1994. cap. 10, p. 253-297.

CARNEIRO, J. D. S.; SILVA, C. H. O. S.; MINIM, V. P. R.; REGAZZI, A. J.; SUDA, I. R. Princípios básicos da conjoint analysis em estudos do consumidor. **Ciência e Tecnologia de Alimentos**, Campinas, v. 37, n. Supl., p. 107-114, 2004.

CASELLA, G.; BERGER, R. L. **Inferência Estatística**. 2 edição, São Paulo. 2010. 586 p.

CHANDUKALA, S.R.; KIM, J.; OTTER, T.; ROSSI, P.; ALLENBY, G. M. Choice Models in Marketing: Economic Assumptions, Challenges and Trends. **Foundations and Trends in Marketing**, v. 2, n. 2, p. 97–184, 2007.

CHESHER, A., SANTOS SILVA, J. C. Taste variation in discrete choice models. **Review of Economic Studies**, v. 69, n. 1, p. 147–168, 2002.

CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings algorithm. **The American Statistician**, Alexandria, v. 49, n. 4, p. 327-335, 1995.

CHURCHILL Jr., G. A.; NIELSEN Jr., A. C. **Marketing Research: Methodological Foundations**. 6th edition. Wisconsin: The Dryden Press, 1996. 602 p.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de Sobrevivência Aplicada**. São Paulo, Editora Edgard Blucher, 2006.

COOK, R. D.; CHRISTOPHER, J. N. A Comparison of Algorithms for Constructing Exact D-optimal Designs, **Technometrics**, v. 22, p. 315-24, 1980.

COWLES, K., CARLIN, B. Markov chain Monte Carlo convergence diagnostics: a comparative review, **Journal of the American Statistical Association**, v. 91, p. 883–904, 1996.

COX, D. R. Partial Likelihood, **Biometrika**, v. 62, p. 269-276, 1975.

DAZIANO, R. A. Conditional-logit Bayes estimators for consumer valuation of electric vehicle driving range. **Journal Resource and Energy Economics**, v. 35, p. 429-450, 2013.

DELIZA, R. **The effects of expectation on sensory perception and acceptance**. 1996. 198p. (PhD thesis) - University of Reading, Inglaterra, 1996.

DELIZA, R.; ROSENTHAL, A.; HEDDERLEY, D.; JAEGER, S. R. Consumer Perception of Irradiated Fruit: A case study using Choice-Based Conjoint Analysis. **Journal of Sensory Studies**, v. 25, p. 184-200, 2010.

DELLA LUCIA, S. M. **Métodos Estatísticos para a Avaliação da Influência de Características não Sensoriais no Comportamento do Consumidor**. 135 p. Tese (Doutorado em Ciência e Tecnologia de Alimentos) – Universidade Federal de Viçosa, Viçosa, 2008.

DELLA LUCIA, S. M., MININ, V. P. R., SILVA, C. H. O., MININ, L. A., SILVA, R. C. S. N. Análise conjunta de fatores baseada em escolhas no estudo da embalagem de iogurte light sabor morango. **Brazilian Journal of Food technology (Online)**, v. 14, p. 11-18, 2010.

DELLA LUCIA, S. M.; MINIM, V. P. R.; SILVA, C. H. O.; MINIM, L. A. Fatores da embalagem de café orgânico torrado e moído na intenção de compra do consumidor. **Ciência e Tecnologia de Alimentos**, v. 27, p. 485-491, 2007.

DIAS, A. **Comparação entre os Métodos Soma Constante e Análise Conjunta de Fatores para Estimar a Importância Relativa**. 65p. Dissertação (Mestrado em Estatística Aplicada e Biometria), Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, 2010.

DYKSTRA, O. The Augmentation of Experimental Data to Maximize $|(X'X)^{-1}|$, **Technometrics**, v. 13, p. 682-88, 1971.

EFRON, B. The Efficiency of Cox's Likelihood Function for Censored Data. **Journal of the American Statistical Association**, v. 72, p. 557-565, 1977.

EHLERS, R. S. **Introdução a Inferência Bayesiana**. 64 p., 2003.

EHLERS, R. S. **Introdução a Inferência Bayesiana**. 68 p., 2007.

ENNEKING, U.; NEUMANN, C.; HENNEBERG, S. How important intrinsic and extrinsic product attributes affect purchase decision. **Food Quality and Preference**, Oxford, v. 18, n. 1, p. 133-138, 2007.

FEDEROV, V. V. **Theory of Optimal Experiments**, translated and edited by W.J. Studden and E.M. Klimko, New York: Academic Press, 1972.

FELLER, W. **An Introduction to Probability Theory and Its Applications**, Volume I, 3 edition, Wiley, New York, 1968.

GALLANT, A. R. **Nonlinear Statistical Models**. New York, John Wiley and Sons, Inc, forthcoming, 1984.

GAMERMAN, D. **Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. Texts in Statistical Sciences. Chapman and Hall, London, 1997.

GAMERMAN, D.; LOPES, H. F. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. London: Chapman & Hall, 2nd edition, 2006.

GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches to calculating marginal densities. **Journal of the American Statistical Association, Alexandria**, v. 85, n. 4, p. 398-409, 1990.

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

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

GEWEKE, J. Bayesian inference in econometric models using Monte Carlo integration. **Econometrica**, v. 57, p. 1317-1339, 1989.

GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. **Journal Animal Science**, v. 63, p. 217-244, 1986.

GOLDSTEIN, M. **Modelos integrativos de comportamento do consumidor: análise crítica e perspectivas**. São Paulo, 151 p. Dissertação (Mestrado em Engenharia de Produção) – Escola Politécnica da Universidade de São Paulo, (1997).

GREEN, P. J.; HAN, X-L. **Metropolis methods, Gaussian proposals and antithetic variables**. In Lecture Notes in Statistics 74: Stochastic Models (eds A. Barone, A. Frigessi and M. Piccioni), Springer Verlag, 1992.

GREEN, P.E.; KRIEGER, A.M. **Conjoint Analysis with Product Positioning Applications**, Vol. 5, 1993.

GREENE, W. H. **Econometric Analysis**. 5 edition. New Jersey: Prentice Hall, 2003. 802 p.

GUIMARÃES JÚNIOR, F. R. F.; FREITAS, A. A. F.; GUIMARÃES, L. G. A.; PRIMO, M. A. M.; MACHADO, M. A. V. Análise de preferência declarada para estudar a utilidade da produção de oleaginosas. **Produção** (Impresso), v. 23, p. 846-857, 2013.

GUSTAFSSON, A.; EKDAHL, F.; BERGMAN, B. Conjoint analysis: A usefull tool in the design process. **Total Quality Management**, v. 10, n. 3, p. 327-343, 1999.

HAIR JR, J. F., ANDERSON, R. E., & TATHAN, R. L. **Multivariate data analysis with readings**. 4^a edition: Prentice Hall, 1995. 757 p.

HAIR JR., J. F. **Análise multivariada de dados**. 5^a edição. Porto Alegre: Bookman, 2005. 688 p.

HASTING, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, p. 97-109, 1970.

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

HOFFMAN, S. D.; DUNCAN, G. J. Multinomial and Conditional Logit Discrete Choice Models in Demography. **Demography**, v. 25, n. 3, p. 415-442, 1988.

JARNER, S. F.; TWEEDIE, R. L. Necessary conditions for geometric and polynomial ergodicity of random walk-type Markov chains. **Bernoulli**, v. 9, n. 4, p. 559–578, 2003.

JOHNSON, N. L. e KOTZ, S. **Distributions in Statistics: Continuous Univariate Distributions**, Vol. 1, Houghton Mifflin, Boston, 1970. 300 p.

KOOP, G., POIRIER, D.J. Bayesian analysis of logit models using natural conjugate priors. **Journal of Econometrics**, v. 56, n. 3, p. 323–340, 1993.

KOTLER, P.; ARMSTRONG, G. **Princípios de marketing**. 12ª edição. Rio de Janeiro: Prentice Hall do Brasil LTDA, 2008. 624 p.

KUHFELD, W. F. Efficient Experimental Design Using Computerized Searches, **Sawtooth Software: Research Paper Series**, 1997.

KUHFELD, W. F.; TOBIAS, R. D.; GARRATT, M. Efficient Experimental Design with Marketing Research Applications, **Journal of Marketing Research**, v. 31, p. 545-557, 1994.

LAPLACE, P. S. Memoir on the probability of causes of events. **Statistical Science**, v. 1, p. 364-78, 1986.

LOCKSHIN, L.; JARVIS, W.; d'HAUTEVILLE, F.; PERROUTY, J. P. Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price and awards in wine choice. **Food Quality and Preference**, Oxford, v. 17, n. 3-4, p. 166-178, 2006.

LOUVIERE, J. J. Hierarchical information integration: A new method for the design and analysis of complex multiattribute judgment problems. **Association of Consumer Research**, v.11, p. 148-155, 1984.

LOUVIERE, J. J.; G. WOODWORTH. Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data, **Journal of Marketing Research**, v. 20, n. 4, p. 350-367, 1983.

MACFIE, H. J.; BRATCHELL, N.; GREENHOFF, K.; VALLIS, L. V. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. **Journal of Sensory Studies**, v. 4, p. 129-148, 1989.

MAHAJAN, V.; GREEN, P. E.; GOLDBERG, S. M. A Conjoint Model for Measuring Self and Cross-Price/Demand Relationships. **Journal of Marketing Research**, v. 19, n. 3, p. 334-342, 1982.

MALHOTRA, N. K. **Pesquisa de marketing: uma orientação aplicada**. 6ª edição. Porto Alegre: Bookman, 2012. 768 p.

MATOS, T. O. **Conjoint Analysis: Uma aplicação ao Marketing**. 2011. 126 p. Juiz de Fora Monografia (Graduação em Estatística) - Departamento de Estatística, Universidade Federal de Juiz de Fora, 2011.

MCFADDEN, D. Conditional Logit Analysis of Qualitative Choice Behavior, **Frontiers in Econometrics**, Academic Press, P. Zarembka Eds., New York, p. 105-142, 1974.

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

MINGON, H. S.; SOUZA, A. D. P.; SCHMIDT, A. M. **Modelos Hierárquicos e Aplicações**. ABE – Associação Brasileira de Estatística. 18ª Sinape, 2008. 279 p.

MITCHELL, T. J. An Algorithm for the Construction of D-optimal Experimental Designs, **Technometrics**, v. 16, p. 203-210, 1974.

MITCHELL, T. J.; MILLER, F. L. **Use of Design Repair to Construct Designs for Special Linear Models**, Math. Div. Ann. Progr. Rept. (ORNL-4661), 130-31, Oak Ridge, TN: Oak Ridge National Laboratory, 1970.

MOORE, H. L. A cross-validity comparison of rating-based and choice-based conjoint analysis models. **International Journal of Research in Marketing**. v. 21, n. 3, p. 299-312, 2004.

MOSKOWITZ, H. R.; ITTY, B.; KATZ, R.; MAIER, A.; BECKLEY, J.; FLORES, L. Hispanic and non-hispanic responses to concepts for four foods. **Journal of Sensory Studies**, Hoboken, v. 19, n. 6, p. 459-485, 2004.

NOGUEIRA, D. A.; SÁFADI, T.; FERREIRA, D. F. Avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov. **Revista Brasileira de Estatística**. Rio de Janeiro, v. 65, n. 224, p. 59-88, 2004.

O'HAGAN, A. **Kendall's advanced theory of statistics**. Volume 2B. Bayesian Inference. New York, USA: Edward Arnold Press, 1994, 402 p.

PHILIPPE, F.; MICHKLE, S. General combinatorial schemas: Gaussian limit distributions and exponential tails. **Discrete Mathematics**, North-Holland, v. 114, p. 159-180, 1993.

RAFTERY, A. E.; LEWIS, S. How many iterations in the Gibbs sampler? **Bayesian Statistics** (Eds.: J. M. Bernardo et al.), Oxford, USA: University Press, p. 763-773, 1992.

ROSSI, E., ALLENBY, G., MCCULLOCH, R. **Bayesian statistics and marketing**, John Wiley & Sons, 2005. 348 p.

SAWTOOTH SOFTWARE. **CBC/HB Technical Paper**, 2009.

SCHWARZ, G. Estimating the Dimension of a Model, **The Annals of Statistics**, v. 6, n. 2, p. 461-464, 1978.

SCLOVE, S. L. Application of model-selection criteria to some problems in multivariate analysis. **Psychometrika**, v. 52, n. 3, p. 333-343, 1987.

SILVA, C. H. O.; BATOS, F. S. **Introdução à Conjoint analysis**. IX Encontro Mineiro de Estatística (MGEST), Viçosa, 2010.

SILVA, F. A. B. **Modelos Paramétricos de Escolha Discreta Aplicados à Receita Médica e Automedicação no Continente Português**. 2004. 92 p. Dissertação

(Mestrado em Econometria Aplicada e Previsão) - Instituto Superior de Economia e Gestão, Universidade federal de Lisboa, Lisboa, 2004.

SILVA, F. F. **Análise bayesiana do modelo auto-regressivo para dados em painel: aplicação na avaliação genética de touros da raça Nelore.** 2006. 100 p. Tese (Doutorado em Estatística e Experimentação Agorpecuária) – Universidade Federal de Lavras, Lavras, 2006.

SIQUEIRA, J. de O. **Mensuração da Estrutura de Preferência do Consumidor: Uma Aplicação de Conjoint Analysis em Marketing.** 2000. 230 p. Dissertação (Mestrado em Administração, concentração em Métodos Quantitativos) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, (2000).

SO, Y.; KUHFIELD, W. F. **Multinomial Logit Models.** SAS Institute Inc., Cary, NC, 2010.

SOLGAARD, H. S.; HANSEN, T. A hierarchical Bayes model of choice between supermarket formats. **Journal of Retailing and Consumer Services**, v. 10, n. 3, p. 69-180, 2003.

SORENSEN, D. A., WANG, C. S., JENSEN, J. Bayesian analysis of genetic change due to selection using Gibbs sampling. **Genetics Selection Evolution**, v. 26, n.4, 333-360, 1994.

SORENSEN, D. **Gibbs sampling in quantitative genetics.** Intern Report, n.82, Danish Institute of Animal, Department of Breeding and Genetics, Denmark, 1996.

SOUZA, E. A. M.; MINIM, V. P. R.; CARNEIRO, J. D. S.; DANTAS, M. I. S.; RIBEIRO, M. M. **O efeito da embalagem na intenção de compra do consumidor.** In: ENCONTRO NACIONAL SOBRE PROCESSAMENTO MÍNIMO DE FRUTAS E HORTALIÇAS, 3º, 2004, Universidade Federal de Viçosa, Viçosa, MG, **Palestras, Resumos e Oficinas**, Viçosa, III EMP, 2004.

TEMPESTA, T.; GIANCRISTOFARO, R. A.; CORAIN, L.; SALMASO, L.; TOMASI, D.; BOATTO, V. The importance of landscape in wine quality perception: an integrated approach using choice-based conjoint analysis and combination-based permutation tests. **Food Quality and Preference**, Oxford, v. 21, n. 7, p. 827-836, 2010.

THEIL, H. A Multinomial Extension of the Linear Logit Model, **International Economic Review**, v. 10, p. 251-259, 1969.

Tierney, L. Markov chains for exploring posterior distributions (with discussion). **Annals of Statistics**, v. 22, p. 1701-1762, 1994.

TIERNEY, L.; KADANE, J. B. Accurate approximations for posterior moments and marginal densities. **Journal of the American Statistical Association**, v. 81, n. 393, p. 82-86, 1986.

VAN TASSELL, C. P.; VAN VLECK, L. D. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. **Journal of Animal Science**, v. 74, p. 2586-2597, 1996.

VARIAN, H. R; REPCHECK, J. **Intermediate Microeconomics: A Modern Approach**, Vol. 7, New York, Norton, 2010.

WELLMAN, R.; VIDICAN, C. Pilot study of a hierarchical Bayes method for utility estimation in a choice-based conjoint analysis of prescription benefit plans including medication therapy management services. **ScienceDirect Social and Administrative Pharmacy**, v. 4, n. 3, p. 218-230, 2008.

ZELLNER, A.; ROSSI, P. E. Bayesian analysis of dichotomous quantal response models. **Journal of Econometrics**, v. 25, n. 3, p. 365-393, 1984.

Apêndice A

A coluna da Esquerda (**Consumidor**) representa a indicação de cada consumidor (hipotético) que participou da pesquisa. As colunas de 1 a 12 (**Tratamentos**) indicam como seria a ordem de apresentação dos tratamentos para cada um destes consumidores.

Consumidor	Tratamentos											
	1	2	3	4	5	6	7	8	9	10	11	12
1	6	7	5	3	2	8	10	4	1	9	12	11
2	1	10	12	2	11	5	9	6	4	7	8	3
3	3	8	7	4	6	9	5	11	2	12	10	1
4	4	9	8	11	3	12	7	1	6	10	5	2
5	8	4	3	9	7	11	6	12	5	1	2	10
6	9	11	4	12	8	1	3	10	7	2	6	5
7	2	5	10	6	1	7	12	3	11	8	9	4
8	7	3	6	8	5	4	2	9	10	11	1	12
9	5	6	2	7	10	3	1	8	12	4	11	9
10	11	12	9	1	4	10	8	2	3	5	7	6
11	10	2	1	5	12	6	11	7	9	3	4	8
12	12	1	11	10	9	2	4	5	8	6	3	7
13	9	5	11	7	3	10	2	8	6	12	4	1
14	12	1	8	4	10	6	7	2	5	3	9	11
15	3	11	2	9	6	5	4	7	1	10	12	8
16	11	9	3	5	2	7	6	10	4	8	1	12
17	8	12	10	1	7	4	5	6	9	2	11	3
18	1	4	12	6	8	2	10	3	7	11	5	9
19	2	3	6	11	4	9	1	5	12	7	8	10
20	6	2	4	3	1	11	12	9	8	5	10	7
21	4	6	1	2	12	3	8	11	10	9	7	5
22	5	7	9	10	11	8	3	12	2	1	6	4
23	7	10	5	8	9	12	11	1	3	4	2	6
24	10	8	7	12	5	1	9	4	11	6	3	2
25	6	1	11	4	7	9	12	8	3	2	10	5
26	1	4	6	9	11	8	7	2	12	5	3	10
27	7	11	12	6	3	1	10	4	5	9	2	8
28	3	12	10	7	5	11	2	6	8	1	9	4
29	8	2	9	5	4	10	1	3	5	12	11	7
30	9	8	4	2	1	5	6	10	11	3	7	12
31	11	6	7	1	12	4	3	9	10	8	5	2
32	2	5	8	10	9	3	4	12	1	7	6	11
33	10	3	5	12	2	7	8	11	9	6	4	1
34	5	10	2	3	8	12	9	7	4	11	1	6
35	4	9	1	8	6	2	11	5	7	10	12	3
36	12	7	3	11	10	6	5	1	2	4	8	9
37	9	5	7	1	8	11	4	12	3	6	2	10

38	4	8	3	7	2	9	10	5	6	1	12	11
39	8	7	4	9	3	5	2	1	10	11	6	12
40	10	2	6	3	12	4	11	8	1	7	5	9
41	2	3	10	4	6	8	12	7	11	9	1	5
42	11	12	1	6	5	10	9	2	7	3	8	4
43	6	10	12	2	11	3	1	4	5	8	9	7
44	1	11	5	12	9	6	7	10	8	2	4	3
45	3	4	2	8	10	7	6	9	12	5	11	1
46	7	9	8	5	4	1	3	11	2	12	10	6
47	12	6	11	10	1	2	5	3	9	4	7	8
48	5	1	9	11	7	12	8	6	4	10	3	2

Apêndice B

Planejando o Experimento

```
install.packages("conjoint")
library("conjoint")

# Definindo os tratamentos

experiment<-expand.grid(
  A<-c("a1","a2","a3"),
  B<-c("b1","b2"),
  C<-c("c1","c2"))
design<-caFactorialDesign(data=experiment,type="full")
print(design)
code<-caEncodedDesign(design)
code

code[code==1]<-0
code[code==2]<-1
code[code==3]<-2
code

# Definindo o número de tratamentos a serem avaliados
p=dim(code)[1]
p

# Definindo a variável indicadora de indivíduos
n=96
ind=rep(1:n, each=p)
ind

# Definindo a variável indicadora de tratamento
trat=rep(1:p, n)
trat

# Replicando a matriz de covariáveis (X) para cada indivíduo (n)
X=as.matrix(do.call("rbind", replicate(n,code,simplify=F)))
X

# Definindo a variável indicadora de escolha (1 se o p-ésimo produto foi escolhido e 0
caso contrário)
m=diag(p)
m

n1=10; c_1=as.matrix(do.call("rbind", replicate(n1,cbind(m[,1]),simplify=F)))
n2=5; c_2=as.matrix(do.call("rbind", replicate(n2,cbind(m[,2]),simplify=F)))
n3=2; c_3=as.matrix(do.call("rbind", replicate(n3,cbind(m[,3]),simplify=F)))
n4=16; c_4=as.matrix(do.call("rbind", replicate(n4,cbind(m[,4]),simplify=F)))
```

```

n5=7; c_5=as.matrix(do.call("rbind", replicate(n5,cbind(m[,5]),simplify=F)))
n6=13; c_6=as.matrix(do.call("rbind", replicate(n6,cbind(m[,6]),simplify=F)))
n7=9; c_7=as.matrix(do.call("rbind", replicate(n7,cbind(m[,7]),simplify=F)))
n8=3; c_8=as.matrix(do.call("rbind", replicate(n8,cbind(m[,8]),simplify=F)))
n9=4; c_9=as.matrix(do.call("rbind", replicate(n9,cbind(m[,9]),simplify=F)))
n10=14; c_10=as.matrix(do.call("rbind", replicate(n10,cbind(m[,10]),simplify=F)))
n11=7; c_11=as.matrix(do.call("rbind", replicate(n11,cbind(m[,11]),simplify=F)))
n12=6; c_12=as.matrix(do.call("rbind", replicate(n12,cbind(m[,12]),simplify=F)))

Choice=rbind(c_1,c_2,c_3,c_4,c_5,c_6,c_7,c_8,c_9,c_10,c_11,c_12)
Choice

# Definindo o conjunto de dados final e o transformando em um data.frame
dados=as.data.frame(cbind(ind, trat, X, Choice))
colnames(dados)<-c("ind", "trat", "A", "B", "C", "Choice")
dados
attach(dados)

```

Apêndice C

Ajustando o Modelo Logit Multinomial pela abordagem Frequentista

```
install.packages("survival")  
library("survival")
```

```
Logit<-clogit(Choice ~ A + B + C, data=dados, method=c("breslow"))  
summary(Logit)
```

```
# Calculando as Utilidades
```

```
betas_freq=Logit$coef  
betas_freq
```

```
X_calc=as.matrix(dados[1:p,c("A", "B", "C")])  
X_calc
```

```
uti_freq=exp(X_calc%*%betas_freq)  
uti_freq
```

```
prob_freq=uti_freq/sum(uti_freq)  
prob_freq
```

Apêndice D

Abordagem Bayesiana para o Modelo Logit Multinomial

```
install.packages("bayesm")
library("bayesm")

# Vetor com o número de vezes que cada tratamento foi escolhido

n_escolhas=c(n1,n2,n3,n4,n5,n6,n7,n8,n9,n10,n11,n12)
n_escolhas

# No bayesm, o vetor y deve ser definido como uma sequencia de valores representando
o número de vezes
# que cada tratamento foi escolhido. Por exemplo, se o tratamento 1 foi escolhido 10
vezes
# temos que repetir o 1 10 vezes.

y=c(rep(seq(1:p),n_escolhas))
y

# Especificando a distribuição a Priori para os parâmetros

r=dim(X_calc)[2]
betabar=c(rep(0,r))
betabar

A=0.01*diag(r)
A

data1=list(y=y,X=X,p=p)
mcmc1=list(R=24000,keep=2)
prior1=list(A=A, betabar=betabar) # Neste caso são apenas 2 fatores e utilizei as
estimativas dos betas (frequentista) para valores de betabar da priori
prior1

out=rmnlIndepMetrop(Data=data1, Prior=prior1, Mcmc=mcmc1)

# Construindo uma matriz para cadeia de cada parâmetro

b1=cbind(out$betadraw[2001:12000,1])
b2=cbind(out$betadraw[2001:12000,2])
b3=cbind(out$betadraw[2001:12000,3])

# Definindo as utilidades como função dos betas

u1=rep(exp(0),length(b1))
u2=exp(b1)
u3=exp(2*b1)
u4=exp(b2)
```

```
u5=exp(b1+b2)
u6=exp(2*b1+b2)
u7=exp(b3)
u8=exp(b1+b3)
u9=exp(2*b1+b3)
u10=exp(b2+b3)
u11=exp(b1+b2+b3)
u12=exp(2*b1+b2+b3)
ut=u1+u2+u3+u4+u5+u6+u7+u8+u9+u10+u11+u12
```

```
# Definindo as Probabilidades de Escolha
```

```
p1=u1/ut
p2=u2/ut
p3=u3/ut
p4=u4/ut
p5=u5/ut
p6=u6/ut
p7=u7/ut
p8=u8/ut
p9=u9/ut
p10=u10/ut
p11=u11/ut
p12=u12/ut
```

```
# Definindo as Razões de Escolhas
```

```
# RE para os níveis do Fator A
```

```
RE_A_1=p1/p2
RE_A_2=p1/p3
RE_A_3=p2/p3
```

```
# RE para os níveis do Fator B
```

```
RE_B=p4/p1
```

```
# RE para os níveis do Fator C
```

```
RE_C=p1/p7
```


Resultados e Avaliação da Convergência das cadeias

```
# Criando uma matriz com as distribuições criadas dos betas
```

```
b_bayes=cbind(b1,b2,b3)
data1=b_bayes
colnames(data1)=c("A","B","C")
data1
```

```
# Criando uma matriz com as distribuições criadas para as probabilidades
```

```
prob_bayes=(cbind(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10,p11,p12))
data2=prob_bayes
colnames(data2)=c("p1","p2","p3","p4","p5","p6","p7","p8","p9","p10","p11","p12")
data2
```

```
# Criando uma matriz com as distribuições para as RE
```

```
RE_bayes=cbind(RE_A_1,RE_A_2,RE_A_3,RE_B,RE_C)
data3=RE_bayes
colnames(data3)=c("RE_A_1","RE_A_2","RE_A_3","RE_B","RE_C")
data3
```

```
install.packages("boa")
library("boa")
```

```
boa.menu()
```

```
## Moda a Posteriori
```

```
install.package("MCMCglmm")
library("MCMCglmm")
```

```
posterior.mode(data1)
posterior.mode(data2)
posterior.mode(data3)
```

```
numEff(b1, m = as.integer(min(length(b1), (100/sqrt(5000)) * sqrt(length(b1))))))
numEff(b2, m = as.integer(min(length(b2), (100/sqrt(5000)) * sqrt(length(b2))))))
numEff(b3, m = as.integer(min(length(b3), (100/sqrt(5000)) * sqrt(length(b3))))))
```

```
H=round(mnlHess(betabar,y, X),4)
H
```

```
inv_H=round(solve(H),4)
inv_H
```

```
par(mfrow=c(1,3))
hist(b1, prob=TRUE, main="Fator A", xlab="A", ylab="Densidade")
curve(dnorm(x,Logit$coef[1],sqrt(inv_H[1,1])),-2,2,add=T)
hist(b2, prob=TRUE, main="Fator B", ylim=c(0,2), xlab="B", ylab="Densidade")
curve(dnorm(x,Logit$coef[2],sqrt(inv_H[2,2])),-2,2,add=T)
```

```
hist(b3, prob=TRUE, main="Fator C", xlab="C", ylab="Densidade")
curve(dnorm(x,Logit$coef[3],sqrt(inv_H[3,3])),-2,2,add=T)
```

```
par(mfrow=c(2,3))
plot.ts(b1, main="A", xlab="Iteração", ylab="Valores Simulados")
plot.ts(b2, main="B", xlab="Iteração", ylab="Valores Simulados")
plot.ts(b3, main="C", xlab="Iteração", ylab="Valores Simulados")
acf(b1, main="A", xlab="Defasagem (Lag)", ylab="Autocorrelação")
acf(b2, main="B", xlab="Defasagem (Lag)", ylab="Autocorrelação")
acf(b3, main="C", xlab="Defasagem (Lag)", ylab="Autocorrelação")
```

```
par(mfrow=c(4,3))
plot(density(p1), main="Prob. 1", xlab="Valores Simulados", ylab="Densidade")
plot(density(p2), main="Prob. 2", xlab="Valores Simulados", ylab="Densidade")
plot(density(p3), main="Prob. 3", xlab="Valores Simulados", ylab="Densidade")
plot(density(p4), main="Prob. 4", xlab="Valores Simulados", ylab="Densidade")
plot(density(p5), main="Prob. 5", xlab="Valores Simulados", ylab="Densidade")
plot(density(p6), main="Prob. 6", xlab="Valores Simulados", ylab="Densidade")
plot(density(p7), main="Prob. 7", xlab="Valores Simulados", ylab="Densidade")
plot(density(p8), main="Prob. 8", xlab="Valores Simulados", ylab="Densidade")
plot(density(p9), main="Prob. 9", xlab="Valores Simulados", ylab="Densidade")
plot(density(p10), main="Prob. 10", xlab="Valores Simulados", ylab="Densidade")
plot(density(p11), main="Prob. 11", xlab="Valores Simulados", ylab="Densidade")
plot(density(p12), main="Prob. 12", xlab="Valores Simulados", ylab="Densidade")
```

```
par(mfrow=c(3,2))
plot(density(RE_A_1), main="RE (1,2)", xlab="Valores Simulados",
ylab="Densidade")
plot(density(RE_A_2), main="RE (1,3)", xlab="Valores Simulados",
ylab="Densidade")
plot(density(RE_A_3), main="RE (2,3)", xlab="Valores Simulados",
ylab="Densidade")
plot(density(RE_B), main="RE (4,1)", xlab="Valores Simulados", ylab="Densidade")
```