**Testes de hipóteses
em eleições majoritárias**

Victor Fossaluza

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Área de Concentração: Estatística
Orientador: Prof. Dr. Luís Gustavo Esteves

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro do CNPq

São Paulo, junho de 2008

Testes de hipóteses em eleições majoritárias

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e defendida por Victor Fossaluza
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Prof. Dr. Luís Gustavo Esteves (orientador) - IME-USP.
- Prof. Dr. Sergio Wechsler - IME-USP.
- Prof. Dr. Fabio Gagliardi Cozman - POLI-USP.

Agradecimentos

Em primeiro lugar, agradeço ao meu mestre e amigo, Professor Luís Gustavo, por todos os ensinamentos, dedicação e paciência ao longo dos cinco anos em que trabalhamos juntos.

Agradeço à minha família, em especial meus pais, que incondicionalmente incentivaram meus estudos e são os grandes responsáveis pela minha formação como ser humano.

Aos amigos que caminharam comigo em direção ao meu objetivo e contribuíram, direta ou indiretamente, para a realização deste trabalho. Em especial, agradeço à Laís Tubertini pela leitura cuidadosa deste manuscrito, pelo apoio incansável, companhia e paciência, e ao meu grande amigo Professor Carlos Alberto de Bragança Pereira, o Carlinhos, pelo incentivo e sugestões valiosas. Não é possível incluir aqui todos os nomes mas gostaria de deixar registrado também meu agradecimento à Patrícia Viana da Silva e ao Jony Arrais Pinto Júnior, que estiveram ao meu lado ao longo de todo o mestrado, ao Fernando Rodrigo Rago, a quem dei muito trabalho e foi fundamental na parte computacional, ao Paulo Cilas Marques Filho, por uma idéia essencial para a conclusão desse trabalho, e ao Rafael Izbicki, pela ajuda em uma demonstração.

Resumo

O problema de Inferência sobre uma proporção, amplamente divulgado na literatura estatística, ocupa papel central no desenvolvimento das várias teorias de Inferência Estatística e, invariavelmente, é objeto de investigação e discussão em estudos comparativos entre as diferentes escolas de Inferência. Ademais, a estimação de proporções, bem como teste de hipóteses para proporções, é de grande importância para as diversas áreas do conhecimento, constituindo um método quantitativo simples e universal. Nesse trabalho, é feito um estudo comparativo entre as abordagens clássica e bayesiana do problema de testar as hipóteses de ocorrência ou não de 2º turno em um cenário típico de eleição majoritária (maioria absoluta) em dois turnos no Brasil.

Palavras-chave: testes de hipóteses, inferência bayesiana, razão de verossimilhanças, princípio da verossimilhança, “monotonicidade”, inferência sobre proporções.

Abstract

The problem of inference about a proportion, widely explored in the statistical literature, plays a key role in the development of several theories of statistical inference and, invariably, is the object of investigation and discussion in comparative studies among different schools of inference. In addition, the estimation of proportions, as well as test of hypothesis for proportions, is very important in many areas of knowledge as it constitutes a simple and universal quantitative method. In this work a comparative study between the Classical and Bayesian approaches to the problem of testing the hypothesis of occurrence of second round (or not) in a typical scenario of a majoritarian election (absolute majority) in two rounds in Brazil is developed.

Keywords: hypothesis tests, bayesian inference, likelihood ratio, likelihood principle, “monotonicity”, inference about proportions.

Sumário

Lista de Abreviaturas	ix
Lista de Símbolos	xi
Lista de Figuras	xiii
Lista de Tabelas	xv
1 Introdução	1
1.1 Considerações Preliminares	1
1.2 Objetivos	2
1.3 Organização do Trabalho	3
2 Testes de Hipóteses	5
2.1 Testes de Hipóteses Clássicos	7
2.2 Testes de Hipóteses Bayesianos	12
2.3 Monotonicidade e Princípio da Verossimilhança	14
3 Exemplos	17
3.1 Modelos Multinomial e Multinomial Negativo	18
3.2 Exemplo: Abordagem Clássica	21

3.3 Exemplo: Abordagem Bayesiana	34
4 Conclusões	39
4.1 Considerações Finais	39
4.2 Sugestões para Pesquisas Futuras	40
A Demonstrações Adicionais	41
A.1 <i>p-value</i> da Multinomial	41
A.2 RV nos Modelos Multinomial e Multinomial Negativo Truncados	46
B Rotinas Computacionais	53
B.1 <i>p-value</i> da Multinomial Negativa	53
B.2 RV na Multinomial Truncada	55
Referências Bibliográficas	57
Índice Remissivo	59

Lista de Abreviaturas

FBST	Teste de Significância Totalmente Bayesiano (<i>Full Bayesian Significance Test</i>).
RC	Região Crítica.
RV	Razão de Verossimilhanças.
UMP	Teste Uniformemente Mais Poderoso.

Lista de Símbolos

θ	Vetor de parâmetros.
Θ	Espaço paramétrico.
$\mathcal{X}$	Espaço amostral.
$\mathcal{P}$	Conjunto das distribuições indexadas por θ .
$\mathcal{D}$	Espaço de ações.
$\propto$	Proporcional.
H	Hipótese nula.
A	Hipótese alternativa.
φ	Função de teste.
α	Nível de significância.
π	Função poder.
$f(x_-)$	$\lim_{t \uparrow x} f(t)$.
$f(x_+)$	$\lim_{t \downarrow x} f(t)$.

Lista de Figuras

3.1	Região definida pela hipótese H: $\theta_1 \leq 0.5$	21
3.2	Região definida pela hipótese H': $\bigcap_{i=1}^3 \{\theta_i \leq 0.5\}$	25
3.3	Região C_y no modelo Multinomial Negativo.	29
3.4	$P(\mathbf{Y} \in C_y \boldsymbol{\theta})$ para o modelo multinomial, obtida por simulação	30

Lista de Tabelas

2.1	Tipos de erro em teste de hipóteses.	7
2.2	Função de perda para teste de hipóteses.	13
3.1	RV para o modelo multinomial truncado.	32
3.2	RV para o modelo multinomial negativo truncado.	32

Capítulo 1

Introdução

1.1 Considerações Preliminares

O problema de Inferência sobre uma proporção, amplamente divulgado na literatura estatística, ocupa papel central no desenvolvimento das várias teorias de Inferência Estatística e, invariavelmente, é objeto de investigação e discussão em estudos comparativos entre as diferentes escolas de Inferência (Barnett [1]). Ademais, a estimação de proporções, bem como teste de hipóteses para proporções, é de grande importância para as diversas áreas do conhecimento, constituindo um método quantitativo simples e universal. Nesse trabalho, pretendemos fazer um estudo comparativo entre as abordagens clássica e bayesiana do problema de testes de hipóteses para proporções considerando um cenário típico de eleição majoritária (maioria absoluta) em dois turnos (Nicolau [12]) no Brasil.

A abordagem clássica do problema de inferência sobre proporções é comumente utilizada nas principais instituições de pesquisas de intenção de voto no Brasil. Em tais pesquisas, usualmente são exibidas, para cada candidato em um pleito majoritário, estimativas pontual e intervalar para a proporção de eleitores com intenção de voto no candidato. Em geral, essas estimativas são obtidas a partir de estatísticas negativamente correlacionadas, fato algumas vezes negligenciado na derivação de estimadores por região (intervalo) para proporções (e diferenças entre proporções) (Roussas [21]). Além disso, a metodologia utilizada para avaliar hipóteses de interesse é baseada somente na sobreposição dos intervalos de confiança, de forma que não são realizados testes de hipóteses propriamente ditos. Assim, é dito, por exemplo, que ocorreu um “empate técnico” entre dois ou mais candidatos quando há superposição dos respectivos intervalos de confiança. Além disso, “para poder afirmar que uma eleição acabará no primeiro turno, a diferença entre a intenção de voto do primeiro colocado e

a soma das demais candidaturas deve ser superior à margem de erro da pesquisa” (Ibope [8]).

1.2 Objetivos

Dentro do contexto de pesquisa de intenção de voto em um pleito majoritário, os objetivos desse trabalho são:

1. estudar a construção de testes para as hipóteses de ocorrência ou não de segundo turno em um pleito majoritário, considerando dois esquemas amostrais: o primeiro baseado nas contagens (amostrais) de votos dos candidatos, fixando o número de eleitores entrevistados (cujo modelo estatístico associado é Multinomial (Pereira e Stern [15]; Pereira e Viana [17])) e um segundo que consiste em registrar os votos nos candidatos até que um candidato específico, digamos candidato 1, obtenha um número de votos fixado (cujo modelo associado é Multinomial Negativo (Mosimann [10]; Olkin e Sobel [13])). Nesse exemplo, é estudado, com base no nível descritivo (chamado doravante de *p-value*), o teste da razão de verossimilhanças num cenário em que a hipótese (nula) composta não é precisa, caso pouco ilustrado na literatura de inferência estatística;
2. estudar a construção de testes para as hipóteses de um candidato fixado vencer ou não a eleição no primeiro turno utilizando o modelo Multinomial, comparar esse procedimento de teste com aquele obtido em (1) e avaliar se esses testes clássicos apresentam a propriedade de “monotonicidade”, propriedade desejável em testes de hipóteses, formulada aqui, ainda em caráter preliminar: tal propriedade possui conexão com as idéias propostas por Royall [22] e Sen [23];
3. estudar a concordância dos procedimentos de teste construídos em (1) com o Princípio da Verossimilhança (DeGroot [6]; Inoue [9], Berger e Wolpert [3]);
4. apresentar, nesse cenário de eleições, uma solução bayesiana para a questão da ocorrência (ou não) de segundo turno, adotando o modelo conjugado Dirichlet-Multinomial (DeGroot [5]; Pereira e Viana [17]), avaliar a concordância da operação bayesiana com o Princípio da Verossimilhança e verificar a “monotonicidade” dos testes sob essa abordagem.

1.3 Organização do Trabalho

No Capítulo 2, serão apresentadas as definições básicas de testes de hipóteses sob as abordagens clássica e bayesiana. Exemplos de testes de hipóteses clássicos e bayesianos em um cenário de eleições serão exibidos no Capítulo 3. Finalmente, no Capítulo 4 serão discutidas as conclusões obtidas neste trabalho. Analisaremos as vantagens e desvantagens dos testes apresentados. Em anexo constam as demonstrações adicionais dos resultados apresentados ao longo do trabalho (Apêndice A) e as rotinas computacionais desenvolvidas (Apêndice B).

Capítulo 2

Testes de Hipóteses

Um problema de inferência estatística consiste em fazer afirmações sobre quantidades desconhecidas, ou seja, não observáveis num determinado contexto. Uma quantidade desconhecida é usualmente denominada *parâmetro* e será denotada por θ . Supõe-se conhecido ao menos o conjunto em que θ está contido, o qual dá-se o nome de *espaço paramétrico* e será denotado por Θ . Para fazer afirmações sobre θ , realiza-se um experimento que permite observar uma realização x de um conjunto $\mathcal{X}$ ($x \in \mathcal{X}$) associado a um vetor aleatório X , onde $\mathcal{X}$ é o conjunto em que esse vetor aleatório assume valores, denominado *espaço amostral*. Condicionada ao conhecimento de θ , supõe-se que o vetor aleatório X possui distribuição conhecida $P_\theta \in \mathcal{P}$, onde $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ é o conjunto de distribuições indexadas por θ .

Para cada $x \in \mathcal{X}$, a função $L_x : \Theta \rightarrow \mathbb{R}_+$ que a cada $\theta \in \Theta$ associa $L_x(\theta) = P(X = x|\theta)$ (ou $f(x|\theta)$, no caso em que X é absolutamente contínuo) é denominada *função de verossimilhança* de θ , resultante da observação x . Essa função é o elemento utilizado para descrever a relação entre o parâmetro θ (desconhecido) e o resultado experimental x (conhecido) e desempenha importante papel em diversas escolas de inferência, em particular nas escolas *clássica* (ou frequentista) e *bayesiana*, que serão discutidas nesse trabalho.

A escola clássica, impulsionada pelos trabalhos de Karl Pearson, Ronald A. Fisher e Jerzy Neyman, desempenhou um papel predominante na primeira metade deste século e, ainda hoje, é a que mantém maior número de seguidores. A inferência clássica é sustentada pela interpretação frequentista de probabilidades (Venn, von Mises, Reichenbach, etc). Tal interpretação do conceito de probabilidades baseia-se na regularidade das frequências relativas observadas de um evento e sustenta que a probabilidade de um evento pode ser medida observando a frequência relativa deste numa sucessão

numerosa de experimentos idênticos e independentes. Um estatístico clássico deve ter em mente que todas as conclusões inferenciais são sustentadas pelo princípio da amostragem repetida (Paulino et al. [14]), que afirma que a particular amostra observada x é apenas um dos muitos conjuntos de dados (hipotéticos) que poderiam ter sido observados nas mesmas circunstâncias. Assim, as probabilidades são definidas em (uma σ -álgebra de) $\mathcal{X}$ e são calculadas supondo que se possa realizar o mesmo experimento um grande número de vezes, o que nem sempre é possível. Sob a perspectiva clássica, o parâmetro θ é desconhecido mas fixo (não aleatório, no sentido de desprovido de probabilidade), ou seja, é o particular valor que indexa a distribuição na família $\mathcal{P}$ que descreve “apropriadamente” o processo que gera as observações.

A abordagem bayesiana em problemas de inferência teve sua raiz em 1763 quando Richard Price publicou a obra póstuma do Reverendo Thomas Bayes, intitulada “*An Essay Towards Solving a Problem in the Doctrine of Chances*”. Contudo, foi por volta de 1937 que a escola bayesiana começou a ganhar força, com Bruno de Finetti e Harold Jeffreys. A inferência bayesiana baseia-se na interpretação subjetivista de probabilidades (Ramsey, de Finetti, Savage, etc). Nesse contexto, a probabilidade representa a medida de incerteza de um indivíduo sobre determinado objeto. Dessa forma, a probabilidade atribuída a um evento pode variar de indivíduo para indivíduo, pois depende do conhecimento que cada um tem sobre aquele evento. Nesse sentido, a incerteza sobre o parâmetro desconhecido θ deve ser quantificada em termos de probabilidade. A incerteza inicial ou *a priori* - anterior ou externa em relação ao experimento - deve traduzir-se formalmente por uma distribuição de probabilidade para θ , $f(\theta)$, denominada *distribuição a priori*. A operação bayesiana atualiza a incerteza inicial de um indivíduo sobre θ , juntando seu conhecimento a priori com a informação trazida pelos dados experimentais. O resultado dessa operação corresponde a *distribuição a posteriori*, $f(\theta|x)$. Em outras palavras, a passagem de $f(\cdot)$ para $f(\cdot|x)$ é feita por meio do teorema de Bayes

$$f(\theta|x) = \frac{f(\theta)L_x(\theta)}{\int_{\Theta} f(\theta)L_x(\theta)d\theta} \propto f(\theta)L_x(\theta).$$

A distribuição a posteriori incorpora toda a informação disponível que se tem sobre θ e, assim, todos os procedimentos de inferência bayesiana serão baseados exclusivamente em $f(\theta|x)$. Note que, diferentemente da abordagem clássica, θ é considerado uma variável aleatória pois se trata de algo desconhecido e, dentro da perspectiva subjetivista, deve ganhar probabilidade. As distribuições de probabilidade para θ estão definidas em (uma σ -álgebra de) Θ .

Geralmente, a inferência estatística concentra-se em dois tipos de problemas: estimação e teste de hipóteses, sendo o último o foco deste trabalho. Uma *hipótese estatística* é uma afirmação sobre um parâmetro θ . Um *teste de hipóteses* consiste em um problema de decisão, em que deve-se escolher uma entre duas (ou mais) hipóteses. No caso mais comum da prática estatística, onde tem-se duas hipóteses, o espaço paramétrico é dividido em subconjuntos complementares, Θ_0 e $\Theta_1 = \Theta_0^c$. Às afirmações $H : \theta \in \Theta_0$ e $A : \theta \in \Theta_1$ dá-se o nome de *hipótese nula* e *hipótese alternativa*, respectivamente.

2.1 Testes de Hipóteses Clássicos

Definição 2.1. Um *teste de hipóteses* $\varphi : \mathcal{X} \rightarrow \{0, 1\}$ é uma regra de decisão que especifica para quais pontos $x \in \mathcal{X}$ a hipótese $H : \theta \in \Theta_0$ deve ser rejeitada ($\varphi(x) = 1$) ou aceita ($\varphi(x) = 0$). O subconjunto do espaço amostral $\mathcal{X}$ que contém os pontos que levam a rejeição de H ($\varphi^{-1}(\{1\})$) é chamado *região crítica* (RC) ou *região de rejeição* do teste φ .

Ao se realizar um teste de hipóteses pode-se cometer dois tipos de erros. Se $\theta \in \Theta_0$ mas, erroneamente, decide-se por rejeitar H , diz-se que foi cometido o *erro tipo I*. Por outro lado, se $\theta \in \Theta_1$ e decide-se pela não rejeição de H , o *erro tipo II* foi cometido. Essas duas situações são apresentadas na Tabela (Tab. 2.1).

		$\theta \in \Theta_0$	$\theta \in \Theta_1$
Decisão	H	Decisão Correta	Erro Tipo II
	A	Erro Tipo I	Decisão Correta

Tabela 2.1: Tipos de erro em teste de hipóteses.

Um teste ideal não apresentaria erros de nenhuma espécie, ou seja, é um teste que possuiria probabilidades de erros tipo I e tipo II iguais a zero. Contudo, encontrar um teste com essa propriedade é inviável e uma possível solução seria encontrar testes que apresentem pequenas probabilidades de erros. De acordo com a abordagem proposta por Neyman-Pearson, uma maneira de encontrar um teste razoável é fixar a probabilidade máxima de erro tipo I que o decisor está disposto a cometer e tentar encontrar um teste com a menor probabilidade de erro tipo II. Assim, vamos restringir nossa atenção aos testes com probabilidade de rejeitar H menor ou igual a um valor α fixado, $0 \leq \alpha \leq 1$, para todo $\theta \in \Theta_0$. Diz-se que tais testes têm *nível de significância* α . Além disso, diz-se que um teste

tem *tamanho* α se $\sup_{\theta \in \Theta_0} P(\varphi^{-1}(\{1\})|\theta) = \alpha$. Note que quando $\Theta_0 = \{\theta_0\}$, $\alpha = P(\text{Rejeitar } H \mid \theta_0)$ é a probabilidade de cometer o erro do tipo I.

Usualmente a região crítica é também expressa por meio de uma *estatística de teste*, $W(X)$, que, em geral, é uma função da estatística suficiente. A estatística de teste é utilizada para ordenar o espaço paramétrico de modo que grandes valores de W indicam pontos mais desfavoráveis a $H : \theta \in \Theta_0$. Uma vez fixado o nível de significância α , pode-se determinar a região crítica de um teste, encontrando o *valor crítico* c tal que

$$\alpha = \sup_{\theta \in \Theta_0} P(\varphi^{-1}(\{1\}) \mid \theta) = \sup_{\theta \in \Theta_0} P(W(X) \geq c \mid \theta) = \alpha(c) \quad (2.1)$$

Para um teste de hipóteses φ , também é possível obter as probabilidades de não rejeitar H para cada $\theta \in \Theta_1$. Desse modo, normalmente é considerada a *função poder*, definida a seguir.

Definição 2.2. A *função poder* de um teste de hipóteses φ é a função

$$\begin{aligned} \pi_\varphi : \Theta &\rightarrow [0, 1] \\ \theta &\mapsto \pi_\varphi(\theta) = P(\{x \in \mathcal{X} : \varphi(x) = 1\} \mid \theta) = P(\varphi^{-1}(\{1\}) \mid \theta). \end{aligned}$$

Note que, para um dado $\theta \in \Theta_1$, $P(\text{Aceitar } H \mid \theta) = 1 - \pi_\varphi(\theta)$. Note que se $\Theta_1 = \{\theta_1\}$, $\beta = P(\text{Aceitar } H \mid \theta_1) = 1 - \pi_\varphi(\theta_1)$ é a probabilidade do erro tipo II do teste φ .

Um possível critério para avaliar um teste de hipóteses é estudar sua função poder. Uma função poder ideal valeria 0 para todo $\theta \in \Theta_0$ e 1 caso contrário. Um bom teste possui função poder próxima de 0 quando $\theta \in \Theta_0$ e próxima de 1 quando $\theta \in \Theta_1$. Uma maneira de encontrar um bom teste segundo esse critério é procurar na classe dos testes de tamanho menor ou igual a α aquele que apresenta maior função poder e, conseqüentemente, menor probabilidade de cometer um erro tipo II, para cada $\theta \in \Theta_1$.

Definição 2.3. Seja Λ_α a classe de testes nível de significância igual a α para testar $H : \theta \in \Theta_0$ contra $A : \theta \in \Theta_1$. Um teste φ na classe Λ_α , com função poder π_φ , é um **teste uniformemente mais poderoso** (UMP) se $\pi_\varphi(\theta) \geq \pi_{\varphi'}(\theta)$, $\forall \theta \in \Theta_1$ e para todo teste φ' na classe Λ_α .

Os testes uniformemente mais poderosos são considerados procedimentos ótimos sob a perspectiva clássica e, para o caso em que ambas as hipóteses são simples, podem ser obtidos por meio de

um resultado muito conhecido na literatura estatística, chamado lema de Neyman-Pearson. Além disso, existem algumas extensões que permitem encontrar testes UMP para casos específicos como, por exemplo, quando a distribuição de X pertence à uma família de distribuições com razão de verossimilhanças monótona (Casella e Berger [4]). Entretanto, para a maioria dos problemas, encontrar tais testes não é uma tarefa trivial. Dessa forma, faz-se necessário construir procedimentos de testes que sejam mais gerais e que forneçam resultados razoáveis na maioria dos casos. Sendo assim, consideremos na sequência o teste da razão de verossimilhanças (*generalizada*).

Definição 2.4. Considere o problema de testar as hipóteses $H : \theta \in \Theta_0$ contra $A : \theta \in \Theta_1$. A estatística **razão de verossimilhanças** $RV : \mathcal{X} \rightarrow [0, 1]$ é definida, para cada $x \in \mathcal{X}$, por

$$RV(x) = \frac{\sup_{\theta \in \Theta_0} L_x(\theta)}{\sup_{\theta \in \Theta} L_x(\theta)}$$

Um **teste de razão de verossimilhanças** é qualquer teste que possui uma região de rejeição da forma $\{x \in \mathcal{X} : RV(x) \leq c\}$, onde $c \in [0, 1]$ (Casella e Berger [4]). Note que $0 \leq RV(x) \leq 1$ pois $\Theta_0 \subseteq \Theta$ e $f(x|\theta) \geq 0$.

Em geral, os testes de razão de verossimilhanças não são testes UMP. Contudo, estes testes são bastante intuitivos e, sob algumas condições de regularidade, apresentam boas propriedades assintóticas (Roussas [21]).

Em algumas situações onde a obtenção de testes por meio das metodologias descritas anteriormente é complicada e a hipótese nula pode ser expressa convenientemente como uma intersecção de conjuntos, pode-se utilizar um teste do tipo *união-intersecção*, descrito em Casella e Berger [4]. Nesses casos, temos que a hipótese nula $H : \theta \in \Theta_0$ pode ser escrita como

$$H : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma$$

onde Γ é um conjunto arbitrário de índices. Suponha que, para cada $\gamma \in \Gamma$, existe uma regra de decisão para testar $H_\gamma : \theta \in \Theta_\gamma$ contra $A_\gamma : \theta \in \Theta_\gamma^c$, de modo que a região crítica correspondente é

dada por $\{x \in \mathcal{X} : T_\gamma(x) \in R_\gamma\}$. Então, a região crítica de um teste união-intersecção é dada por

$$\bigcup_{\gamma \in \Gamma} \{x \in \mathcal{X} : T_\gamma(x) \in R_\gamma\}.$$

Note que, em geral, este teste não coincide com o teste de razão de verossimilhanças e tampouco constitui um teste UMP.

Quando é realizado um teste de hipóteses, é possível apresentar as conclusões de diversas formas. Uma possível maneira de apresentar o resultado é dizendo o nível de significância α adotado e a decisão tomada para esse nível de significância. É claro que quando decide-se por rejeitar H e o nível de significância é pequeno, temos um resultado convincente pois a chance de cometer um erro do tipo I é pequena. Por outro lado, se o nível de significância adotado é grande e decide-se por rejeitar H , há uma grande probabilidade de tomar a decisão errada. Assim, outra maneira de reportar o resultado de um teste de hipóteses é apresentando o valor de uma estatística denominada “*p-value*”, sugerida por Fisher.

Definição 2.5. Um ***p-value*** $p(X)$ é uma estatística de teste satisfazendo $0 \leq p(x) \leq 1$ para todo ponto amostral $x \in \mathcal{X}$. Um *p-value* é **válido** se, para todo $\theta \in \Theta_0$ e todo $\alpha \in [0, 1]$,

$$P(p(X) \leq \alpha \mid \theta) \leq \alpha.$$

Valores “pequenos” de $p(X)$ fornecem evidência contra a hipótese H . Segue da definição de *p-value* válido que rejeitar H se, e somente se, $p(X) \leq \alpha$ é um teste de nível α . O *p-value* $p(x)$ pode ser interpretado como o menor nível de significância para o qual o teste conduziria a rejeição de H , com base na observação x . Muitos autores contemporâneos não deixam claro a distinção entre os testes sugeridos por Neyman-Pearson e os testes baseados no *p-value*, sugeridos por Fisher e comumente chamados de testes de significância. Diferentemente da abordagem de Neyman-Pearson, os testes de significância propõem o uso de uma medida de “força de evidência contra a hipótese” (nula).

Como nos testes descritos anteriormente, o *p-value* é calculado, em geral, com auxílio de uma estatística de teste $W(X)$. Dessa forma, a especificação de *p-value* válido baseado em uma estatística W é apresentada na sequência.

Teorema 2.1. *Seja $W(X)$ uma estatística de teste tal que grandes valores de W fornecem evidência*

contra a hipótese $H : \theta \in \Theta_0$. Para cada $x \in \mathcal{X}$, definimos

$$p(x) = \sup_{\theta \in \Theta_0} P(W(X) \geq W(x)|\theta).$$

Então, $p(X)$ é um p -value válido.

Demonstração. Considere $\theta \in \Theta_0$ fixado. Seja $F_\theta(w)$ a função de distribuição de $-W(X)$. Defina

$$p_\theta(x) = P(W(X) \geq W(x)|\theta) = P(-W(X) \leq -W(x)|\theta) = F_\theta(-W(x))$$

A variável aleatória $p_\theta(X) = F_\theta(-W(X))$ é estocasticamente maior ou igual (Ross [20]) a uma variável com distribuição *uniforme*(0, 1), ou seja, $P(p_\theta(X) \leq \alpha|\theta) \leq \alpha$, $\alpha \in [0, 1]$. Mas

$$p(x) = \sup_{\theta' \in \Theta_0} p_{\theta'}(x) \geq p_\theta(x), \forall x \in \mathcal{X} \Rightarrow P(p(X) \leq \alpha|\theta) \leq P(p_\theta(X) \leq \alpha|\theta) \leq \alpha.$$

Como isso é verdade para todo $\theta \in \Theta_0$ e para todo $\alpha \in [0, 1]$, $p(X)$ é um p -value válido.

□

Como interpretação do Teorema 2.1, considere X um vetor aleatório, x uma realização de X e W uma estatística de teste. Então, se um teste tem valor crítico c , ao observar $x \in \mathcal{X}$ decide-se por rejeitar H se, e somente se, $W(x) \geq c$. Dessa forma, o valor $c = W(x)$ é o maior valor de c que conduziria a rejeição de H com base na observação x . Mas, por (2.1), o tamanho de um teste com valor crítico c é $\alpha(c)$ e a função $\alpha(c)$ é decrescente em c . Sendo assim, o menor α que conduziria a rejeição de H é $\alpha(W(x))$.

Nesse contexto, é possível também especificar o p -value para o teste da razão de verossimilhanças. Assim, considere o Corolário 2.1 apresentado na sequência.

Corolário 2.1. *Considere o problema de testar as hipóteses $H : \theta \in \Theta_0$ contra $A : \theta \in \Theta_1$. Para cada $x \in \mathcal{X}$, seja $C_x = \{w \in \mathcal{X} : RV(w) \leq RV(x)\}$ o conjunto dos pontos “mais desfavoráveis” à H que o ponto x sob a ordenação do espaço amostral por meio da razão de verossimilhanças. O p -value $p(X) : \mathcal{X} \rightarrow [0, 1]$ definido, para cada $x \in \mathcal{X}$, por*

$$p(x) = \sup_{\theta \in \Theta_0} P(X \in C_x|\theta)$$

é um p -value válido.

Demonstração. Basta considerar o Teorema 2.1 e tomar, para cada $x \in \mathcal{X}$, $W(x) = 1/RV(x)$.

□

2.2 Testes de Hipóteses Bayesianos

Como já foi dito anteriormente, a inferência bayesiana baseia-se na distribuição a posteriori $f(\theta|x)$. Essa distribuição é obtida por meio do teorema de Bayes, juntando a informação trazida pela observação x , expressa na função de verossimilhança $L_x(\theta)$, com a distribuição a priori $f(\theta)$.

Os procedimentos bayesianos de estimação e teste de hipóteses normalmente são baseados na *teoria da decisão* cujos elementos são apresentados na sequência. Seja Θ o espaço paramétrico (ou *espaço de estados*) tal que cada $\theta \in \Theta$ representa uma possível realização do “estado da natureza”, o elemento desconhecido de interesse, e considere o *espaço de ações* $\mathcal{D}$ onde cada elemento $d \in \mathcal{D}$ representa uma possível ação (ou afirmação sobre θ) entre as quais o decisor deve escolher.

A informação a priori e a informação experimental são complementadas com um novo tipo de “informação” (Paulino et al. [14]) que diz respeito às consequências das decisões na sua interação com o estado da natureza envolvente: a *função de perda* $l : \Theta \times \mathcal{D} \rightarrow \mathbb{R}_+$ que, para cada $\theta \in \Theta$ e cada $d \in \mathcal{D}$, associa a perda $l(\theta, d)$ ocasionado pela escolha de d quando θ é a realização do estado da natureza.

Assim, considere uma *função de decisão* $\delta : \mathcal{X} \rightarrow \mathcal{D}$, que para cada ponto amostral $x \in \mathcal{X}$, associa-se uma decisão $\delta(x) \in \mathcal{D}$. A classe de todas as possíveis funções de decisão será denotada por Δ . Os valores de X (e de θ) são desconhecidos (aleatórios) antes da realização do experimento, de forma que a perda associada à decisão δ é também aleatória. A função de decisão ótima nesse caso é a função $\delta^* \in \Delta$ que minimiza a função de *perda esperada* (ou *risco*), definida por

$$r(\delta) = E[l(\theta, \delta(X))] = \int_{\Theta} \int_{\mathcal{X}} l(\theta, \delta(x)) f(x|\theta) f(\theta) dx d\theta.$$

A abordagem descrita anteriormente é denominada por Raiffa e Schlaifer [19] *forma normal* do problema de decisão. Essa abordagem consiste em escolher a regra de decisão δ^* que minimiza a perda esperada, considerando todas as possíveis regras de decisão $\delta \in \Delta$. Esse procedimento pode ser

bastante trabalhoso (e muitas vezes inviável), de forma que na abordagem bayesiana normalmente trata-se esse problema utilizando a *forma extensiva*, que conduz a resultados idênticos (Raiffa e Schlaifer [19]). Sob essa abordagem, supõe-se que o experimento já foi realizado e observou-se um particular resultado x , obtendo-se a distribuição posterior para θ , $f(\theta|x)$. Nesse caso, a decisão ótima $\delta^*(x)$, também chamada de *decisão de Bayes*, é a ação $d^* \in \mathcal{D}$ que minimiza a função de *risco posterior*, definida por

$$r_x(d) = E[l(\theta, d)|x] = \int_{\Theta} l(\theta, d) f(\theta|x) d\theta.$$

ou seja, $r_x(d^*) = \inf_{d \in \mathcal{D}} r_x(d)$.

O problema de testes de hipóteses pode ser visto como um caso particular da teoria das decisões, onde $\mathcal{D} = \{d_0, d_1\}$ tal que d_0 e d_1 são as decisões correspondentes a aceitar as hipóteses $H : \theta \in \Theta_0$ e $A : \theta \in \Theta_1$, respectivamente. Uma função de perda usual para esse problema é apresentada a seguir.

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
d_0	b_0	a_1
d_1	a_0	b_1

Tabela 2.2: Função de perda para teste de hipóteses.

Note que a função de perda definida na Tabela 2.2 é bem geral pois considera perdas também para os casos em que se toma a decisão correta (b_0 e b_1). Dessa forma, é razoável sempre tomar $a_0 > b_0$ e $a_1 > b_1$. Comumente, toma-se $b_0 = b_1 = 0$. Deste modo, a solução bayesiana para um problema de teste de hipóteses desse tipo (sob a abordagem normal) consiste em escolher a função $\delta \in \Delta$ que minimiza a perda esperada, de forma que a cada $x \in \mathcal{X}$, $\delta(x)$ corresponde à decisão de rejeitar (d_1) ou não (d_0) a hipótese H . Como já foi dito, dado um valor $x \in \mathcal{X}$, esse procedimento é equivalente a escolher a decisão $d \in \{d_0, d_1\}$ que minimiza a função de risco posterior.

Proposição 2.1. *Um teste de Bayes com relação a perda definida na Tabela 2.2 consiste em rejeitar H a partir de $x \in \mathcal{X}$, se, e somente se,*

$$P(\theta \in \Theta_0|x) < \frac{a_1 - b_1}{(a_0 - b_0) + (a_1 - b_1)} \quad , \quad \text{com } a_0 > b_0 \text{ e } a_1 > b_1.$$

Demonstração. O risco posterior da decisão d_0 é

$$r(d_0|x) = E[l(\theta, d_0)|x] = \int_{\Theta} l(\theta, d_0) f(\theta|x) d\theta = \int_{\Theta_0} b_0 f(\theta|x) d\theta + \int_{\Theta_1} a_1 f(\theta|x) d\theta$$

$$= b_0 P(\theta \in \Theta_0|x) + a_1 P(\theta \in \Theta_1|x) = a_1 - (a_1 - b_0)P(\theta \in \Theta_0|x).$$

Por outro lado, o risco posterior da decisão d_1 é

$$\begin{aligned} r(d_1|x) &= E[l(\theta, d_1)|x] = \int_{\Theta} l(\theta, d_1) f(\theta|x) d\theta = \int_{\Theta_0} a_0 f(\theta|x) d\theta + \int_{\Theta_1} b_1 f(\theta|x) d\theta \\ &= a_0 P(\theta \in \Theta_0|x) + b_1 P(\theta \in \Theta_1|x) = b_1 + (a_0 - b_1)P(\theta \in \Theta_0|x). \end{aligned}$$

Assim, um procedimento de teste consiste em rejeitar H (decidir por d_1) se, e somente se, $r(d_1|x) < r(d_0|x)$, ou seja,

$$\begin{aligned} r(d_1|x) < r(d_0|x) &\iff b_1 + (a_0 - b_1)P(\theta \in \Theta_0|x) < a_1 - (a_1 - b_0)P(\theta \in \Theta_0|x) \\ &\iff P(\theta \in \Theta_0|x) < \frac{a_1 - b_1}{(a_0 - b_0) + (a_1 - b_1)} \end{aligned}$$

□

2.3 Monotonicidade e Princípio da Verossimilhança

No próximo capítulo, serão apresentados exemplos de testes de hipóteses sob as abordagens clássica e bayesiana. Mais precisamente, o *p-value* clássico, determinado a partir do teste da razão de verossimilhanças, e a probabilidade a posteriori da hipótese de interesse (bayesiana) serão estudados, pois são os procedimentos mais usuais para testar hipóteses sob essas duas abordagens. Em particular, estamos interessados em estudar esses procedimentos de teste em relação a duas características: a propriedade da “monotonicidade” e o princípio da verossimilhança.

A propriedade da “*monotonicidade*” em teste de hipóteses, segundo a qual a rejeição de uma hipótese “ $\theta \in \Theta_0$ ” deve implicar a rejeição da hipótese “ $\theta \in \Theta'_0$ ”, se $\Theta'_0 \subseteq \Theta_0$, parece ser uma característica desejada. Assim, para qualquer medida de evidência μ espera-se que, para dois conjuntos A e B tais que $A \subseteq B$, $\mu(A) \leq \mu(B)$. Como o *p-value* e a probabilidade posterior são propostas como medidas de evidência contra a hipótese nula, é desejável que estes possuam tal propriedade.

O *princípio da verossimilhança*, bastante conhecido na literatura, diz, resumidamente, que se desejamos fazer afirmações sobre uma quantidade desconhecida θ e se para isso são realizados dois experimentos E e E' que geram as observações x e y , respectivamente, de modo que $L_x(\theta) \propto L'_y(\theta)$,

$\forall \theta \in \Theta$, então a inferência a respeito de θ deve ser idêntica a partir desses dois experimentos (para uma discussão detalhada veja, por exemplo, Barnett [1], Berger e Wolpert [3] e Inoue [9]).

Capítulo 3

Exemplos

Nesse capítulo serão apresentados exemplos de testes de hipóteses sob as perspectivas clássica e bayesiana, bem como uma discussão sobre as correspondentes soluções. A motivação desses exemplos é o problema de pesquisa de intenção de voto em eleições majoritárias em dois turnos, descrito a seguir.

Eleição é o processo mediante o qual um grupo social escolhe seu governante ou seu representante político por meio do voto. Nos países democráticos, o exercício do voto é um dos direitos fundamentais dos cidadãos. É por meio dele que o indivíduo participa do poder público e manifesta sua vontade. No Brasil, o voto é obrigatório para os maiores de 18 anos e facultativo para os analfabetos, maiores de 70 anos e para os jovens que tenham entre 16 e 18 anos.

Existem dois sistemas eleitorais no Brasil, o majoritário e o proporcional. O sistema de eleições majoritárias em dois turnos é usado para a escolha do presidente da república, governadores e prefeitos nas cidades com mais de 200 mil eleitores. No caso dos senadores e dos prefeitos nas cidades com menos de 200 mil eleitores, a escolha é feita por meio de eleições majoritárias em turno único. Para os cargos de vereadores e deputados estaduais e federais, a escolha se dá por eleição proporcional. O foco desse trabalho são as eleições majoritárias em dois turnos onde o candidato é eleito no primeiro turno caso obtenha a maioria absoluta dos votos válidos, ou seja, 50% dos votos válidos mais um. Caso esse percentual não seja atingido, os dois candidatos mais votados disputam o segundo turno, em que se considera eleito quem obtiver a maioria simples dos votos válidos. Entenda por votos válidos todos aqueles diferentes de brancos ou nulos.

Usualmente, a estatística clássica é utilizada para as pesquisas de intenção de voto no Brasil. As principais instituições que realizam tais pesquisas, apresentam os resultados sob a forma de esti-

mativas por intervalo (intervalos de confiança). Em geral, esses intervalos tem o mesmo “tamanho” pois são construídos com base na chamada “margem de erro”, que é o maior erro estimado para as proporções de votos dos candidatos. Aparentemente, metodologia similar é utilizada em instituições estrangeiras (ver, por exemplo, Newport et al. [11]). Apesar dessas instituições utilizarem procedimentos de amostragem bastante complexos, as estimativas são calculadas supondo um esquema de amostragem aleatória simples (Ibope [8]). Deste modo, a metodologia utilizada para avaliar hipóteses de interesse é baseada somente na sobreposição dos intervalos de confiança, de forma que não são realizados testes de hipóteses propriamente ditos. Assim, é dito, por exemplo, que ocorreu um “empate técnico” entre dois ou mais candidatos quando há superposição dos respectivos intervalos de confiança. Além disso, “para poder afirmar que uma eleição acabará no primeiro turno, a diferença entre a intenção de voto do primeiro colocado e a soma das demais candidaturas deve ser superior à margem de erro da pesquisa” (Ibope [8]). Desta forma, um estudo mais detalhado de testes de hipóteses se faz necessário e será apresentado adiante.

Suponha que existam $k + 1$ candidatos e considere θ_i , $i \in \{1, 2, \dots, k + 1\}$, a proporção de votos válidos no i -ésimo candidato. Deste modo, estudaremos testes de hipóteses considerando inicialmente $H : \theta_j \leq 1/2$ contra $A : \theta_j > 1/2$, ou seja, testaremos se o j -ésimo candidato vencerá a eleição no primeiro turno. Em seguida, será testada a hipótese $H : \bigcap_{i=1}^{k+1} \{\theta_i \leq 1/2\}$ (nenhum candidato conseguiu maioria absoluta e haverá segundo turno) contra a hipótese $A : \bigcup_{i=1}^{k+1} \{\theta_i > 1/2\}$ (algum dos candidatos vencerá a eleição no primeiro turno). Como é necessário que um candidato obtenha a maioria absoluta dos votos válidos para a não ocorrência de um segundo turno, serão modelados, por simplicidade, apenas os votos válidos ao longo do capítulo. Além disso, para os resultados desenvolvidos ao longo do capítulo é feita a suposição usual de que a população é infinita.

3.1 Modelos Multinomial e Multinomial Negativo

Considere os seguintes esquemas amostrais:

- observar o candidato preferido de n eleitores, com $n \in \mathbb{N}$ fixado;
- observar o candidato preferido dos eleitores até que o candidato $k + 1$ obtenha y_{k+1} votos, com $y_{k+1} \in \mathbb{N}$ fixado.

Admitindo que as respostas dos indivíduos com relação ao candidato de preferência são independentes (permutáveis) sob a perspectiva clássica (bayesiana), o vetor aleatório $\mathbf{X}$ de intenções de

votos nos candidatos, condicionado ao conhecimento de θ , tem distribuição multinomial no primeiro esquema amostral e o espaço amostral correspondente é $\mathcal{X} = \{(x_1, \dots, x_k) \in \mathbb{N}^k : \sum_{i=1}^k x_i \leq n\}$. Já no segundo esquema amostral, o vetor aleatório $\mathbf{Y}$ de intenções de votos, condicionado ao conhecimento de θ , tem distribuição multinomial negativa e o espaço amostral, nesse caso, é $\mathcal{Y} = \mathbb{N}^k$. Essas distribuições, bem como algumas de suas propriedades (Pereira e Stern [15]; Pereira e Viana [17]; Mosimann [10]), são apresentadas na sequência. É importante salientar que o parâmetro de interesse θ é o mesmo, independente do esquema amostral adotado.

Definição 3.1. *Seja $\mathbf{X} = (X_1, X_2, \dots, X_k)$ um vetor aleatório cujas componentes assumem valores inteiros não negativos tal que $\sum_{i=1}^k x_i \leq n$, com $n \in \mathbb{N} \setminus \{0\}$ fixado. Diz-se que $\mathbf{X}$ tem **distribuição multinomial** com parâmetros n e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ se a distribuição de probabilidade de $\mathbf{X}$ é dada por:*

$$f(\mathbf{x}|\boldsymbol{\theta}) = n! \prod_{i=1}^{k+1} \frac{\theta_i^{x_i}}{x_i!},$$

tal que $\theta_i \geq 0$, $\forall i \in \{1, \dots, k\}$, $\sum_{i=1}^k \theta_i \leq 1$, $\theta_{k+1} = 1 - \sum_{i=1}^k \theta_i$ e $x_{k+1} = n - \sum_{i=1}^k x_i$. No particular caso em que $k = 1$ temos a distribuição **binomial**(n, θ_1).

Nesse caso, vamos denotar $\mathbf{X}|\boldsymbol{\theta} \sim \text{Mult}(n, \boldsymbol{\theta})$. Mais formalmente, a notação poderia ser $\mathbf{X}|n, \boldsymbol{\theta}$. Entretanto, como n é conhecido (fixado de antemão), optamos pela primeira notação.

Resultado 3.1. *Considere o vetor aleatório $\mathbf{X}$ que, dado $\boldsymbol{\theta}$, tem distribuição multinomial, como na Definição 3.1. Então $X_i \mid \boldsymbol{\theta} \sim \text{Bin}(n, \theta_i)$, $i \in \{1, 2, \dots, k+1\}$.*

Resultado 3.2. *Considere o vetor aleatório $\mathbf{X}$ que, dado $\boldsymbol{\theta}$, tem distribuição multinomial, como na Definição 3.1. Seja $X_{k+1} = n - (X_1 + \dots + X_k)$, de modo que $(X_1, \dots, X_{k+1})$ é também vetor aleatório discreto, e considere $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{k+1})$ uma permutação dos índices $1, \dots, k+1$. Então, para $q \in \{1, \dots, k-1\}$,*

$$(X_{\pi_1}, \dots, X_{\pi_q}) \mid X_{\pi_{q+2}} = x_{\pi_{q+2}}, \dots, X_{\pi_{k+1}} = x_{\pi_{k+1}}, \boldsymbol{\theta} \sim \text{Multinomial} \left(n - \sum_{i=q+2}^{k+1} x_{\pi_i}, \boldsymbol{\mu} \right),$$

onde $\boldsymbol{\mu} = (\mu_1, \dots, \mu_q)$ com $\mu_j = \frac{\theta_{\pi_j}}{\sum_{i=1}^q \theta_{\pi_i}}$, $j \in \{1, \dots, q\}$ e $\mu_{q+1} = 1 - \sum_{i=1}^q \mu_i$.

Definição 3.2. Seja $\mathbf{Y} = (Y_1, Y_2, \dots, Y_k)$ um vetor aleatório cujas componentes assumem valores inteiros não negativos e $y_{k+1} \in \mathbb{N} \setminus \{0\}$ fixado. Diz-se que $\mathbf{Y}$ tem distribuição **multinomial negativa** com parâmetros y_{k+1} e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ se sua distribuição de probabilidade é dada por:

$$f(\mathbf{y}|\boldsymbol{\theta}) = \left(y_{k+1} - 1 + \sum_{i=1}^k y_i \right)! \frac{\theta_{k+1}^{y_{k+1}}}{(y_{k+1} - 1)!} \prod_{i=1}^k \frac{\theta_i^{y_i}}{y_i!},$$

tal que $\theta_i \geq 0$, $\forall i \in \{1, \dots, k\}$, $\sum_{i=1}^k \theta_i \leq 1$ e $\theta_{k+1} = 1 - \sum_{i=1}^k \theta_i$. Em particular, se $k = 1$ temos a distribuição **binomial negativa**(y_2, θ_1) e, nesse caso, se $y_2 = 1$ temos a distribuição **geométrica**(θ_1).

Nesse caso, vamos denotar $\mathbf{Y}|\boldsymbol{\theta} \sim \text{MultNeg}(y_{k+1}, \boldsymbol{\theta})$. Analogamente à notação da Multinomial, y_{k+1} será omitido da notação por se tratar de um número fixado previamente.

Resultado 3.3. Considere o vetor aleatório $\mathbf{Y}$ que, dado $\boldsymbol{\theta}$, tem distribuição multinomial negativa, como na Definição 3.2. Então $\sum_{i=1}^k Y_i \mid \boldsymbol{\theta} \sim \text{BinNeg}(y_{k+1}, \theta_{k+1})$.

Resultado 3.4. Considere o vetor aleatório $\mathbf{Y}$ que, dado $\boldsymbol{\theta}$, tem distribuição multinomial negativa, como na Definição 3.2. Então

$$Y_1, \dots, Y_{k-1} \mid \sum_{i=1}^k Y_i = n, \boldsymbol{\theta} \sim \text{Multinomial}(n, \boldsymbol{\theta}_{-(k+1)}),$$

$$\text{onde } \boldsymbol{\theta}_{-(k+1)} = \left(\frac{\theta_1}{\theta_1 + \dots + \theta_k}, \frac{\theta_2}{\theta_1 + \dots + \theta_k}, \dots, \frac{\theta_{k-1}}{\theta_1 + \dots + \theta_k} \right).$$

Nas próximas seções serão desenvolvidos testes clássicos e bayesianos visando responder duas questões. A primeira é verificar se o j -ésimo candidato vencerá a eleição no primeiro turno, ou seja, será testada a hipótese $H : \theta_j \leq 1/2$ contra $A : \theta_j > 1/2$, supondo o modelo multinomial. A segunda questão é verificar a ocorrência ou não de segundo turno, ou seja, será testada a hipótese $H : \bigcap_{i=1}^{k+1} \{\theta_i \leq 1/2\}$ contra $A : \bigcup_{i=1}^{k+1} \{\theta_i > 1/2\}$. Neste último caso, também serão comparados os testes obtidos utilizando os dois esquemas amostrais, descritos pelos modelos multinomial e multinomial negativo.

Por simplicidade, nos exemplos que seguem vamos supor $k = 2$, isto é, um cenário com apenas três candidatos. Na sequência, apresentaremos a descrição dos testes de razão de verossimilhanças para os casos acima.

3.2 Exemplo: Abordagem Clássica

Primeiramente, será apresentado o p -value para testar se o j -ésimo candidato vencerá ou não a eleição no primeiro turno, utilizando o modelo multinomial.

Resultado 3.5. *Seja $\mathbf{X}|\boldsymbol{\theta} \sim \text{Mult}(n, \boldsymbol{\theta})$. O p -value $p_1 : \mathcal{X} \rightarrow [0, 1]$ para testar $H : \theta_j \leq 1/2$ contra $A : \theta_j > 1/2$ é dado, para cada $\mathbf{x} = (x_1, x_2, x_3) \in \mathcal{X}$, por*

$$p_1(\mathbf{x}) = \begin{cases} 1 & \text{se } x_j \leq n/2 \\ P(\text{Bin}(n, 1/2) \geq x_j) & \text{se } x_j > n/2 \end{cases}$$

Figura 3.1.

Figura 3.1: Região definida pela hipótese $H: \theta_1 \leq 0.5$.

Demonstração. Seja $\mathbf{x} = (x_1, x_2, x_3) \in \mathcal{X}$. Vamos supor, sem perda de generalidade, que $j = 1$. É

bem conhecido da literatura de inferência estatística que $\sup_{\theta \in \Theta} L_{\mathbf{x}}(\theta) = \sup_{\theta \in \Theta} \theta_1^{x_1} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{x_3}$ se dá em $\hat{\theta} = \bar{\mathbf{x}} = \mathbf{x}/n$.

Se $x_1 \leq n/2$, temos que $\hat{\theta} \in \Theta_0$ e, assim,

$$\sup_{\theta \in \Theta_0} L_{\mathbf{x}}(\theta) = \sup_{\theta \in \Theta} L_{\mathbf{x}}(\theta) = L_{\mathbf{x}}(\hat{\theta}) \Rightarrow RV(\mathbf{x}) = 1 \Rightarrow C_x = \mathcal{X} \Rightarrow p_1(\mathbf{x}) = 1$$

Considere agora $x_1 > n/2$.

Seja $\theta_2 \in [0, 1/2]$ fixado e $g_{\theta_2}(\theta_1) = \theta_1^{x_1} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{n-x_1-x_2}$, com $x_1 > n/2$. Vamos mostrar que $g_{\theta_2}(\theta_1)$ é crescente no intervalo $\left(0, \min \left\{ \frac{1}{2}; \frac{x_1}{n-x_2}(1-\theta_2) \right\} \right)$ ou, equivalentemente, que $g'_{\theta_2}(\theta_1) \geq 0$, $\forall \theta_1 \in \left(0, \min \left\{ \frac{1}{2}; \frac{x_1}{n-x_2}(1-\theta_2) \right\} \right)$.

$$g'_{\theta_2}(\theta_1) = \theta_2^{x_2} \left[x_1 \theta_1^{x_1-1} (1 - \theta_1 - \theta_2)^{n-x_1-x_2} - (n - x_1 - x_2) \theta_1^{x_1} (1 - \theta_1 - \theta_2)^{n-x_1-x_2-1} \right]$$

$$g'_{\theta_2}(\theta_1) > 0 \iff x_1(1 - \theta_1 - \theta_2) > (n - x_1 - x_2)\theta_1 \iff \theta_1 < \frac{x_1(1 - \theta_2)}{n - x_2}$$

Deste modo, $g_{\theta_2}(\theta_1)$ é crescente em $\left(0, \min \left\{ \frac{1}{2}; \frac{x_1}{n-x_2}(1-\theta_2) \right\} \right)$.

Se $\theta_2 < 1 - \frac{x_1 + x_3}{2x_1}$ (nesse caso, $\frac{x_1}{n-x_2}(1-\theta_2) > \frac{1}{2}$ e devemos tomar $\theta_1 = 1/2$), vamos encontrar o ponto $\theta_2 \in [0; 1/2]$ que maximiza a função $h(\theta_2) = \theta_2^{x_2} (1/2 - \theta_2)^{x_3}$. Assim, para $\theta_2 \in (0, 1/2)$,

$$h'(\theta_2) = x_2 \theta_2^{x_2-1} (1/2 - \theta_2)^{x_3} - x_3 \theta_2^{x_2} (1/2 - \theta_2)^{x_3-1} \quad \text{e}$$

$$h'(\theta_2) = 0 \iff x_2 \theta_2^{x_2-1} (1/2 - \theta_2)^{x_3} = x_3 \theta_2^{x_2} (1/2 - \theta_2)^{x_3-1} \iff \theta_2 = \frac{x_2}{2(x_2 + x_3)}$$

Mas, $h''(\theta_2) = \theta_2^{x_2-2} (1/2 - \theta_2)^{x_3-2} \left[\left(x_3 \theta_2 - x_2 (1/2 - \theta_2) \right)^2 - x_2 (1/2 - \theta_2)^2 - x_3 \theta_2^2 \right]$, donde

$$\begin{aligned}
& h''\left(\frac{x_2}{2(x_2+x_3)}\right) \leq 0 \\
& \iff \left[x_3 \left(\frac{x_2}{2(x_2+x_3)} \right) - x_2 \left(1/2 - \left(\frac{x_2}{2(x_2+x_3)} \right) \right) \right]^2 \\
& \quad - x_2 \left[1/2 - \left(\frac{x_2}{2(x_2+x_3)} \right) \right]^2 - x_3 \left(\frac{x_2}{2(x_2+x_3)} \right)^2 \leq 0 \\
& \iff -\frac{x_3 x_2}{4(x_2+x_3)} \leq 0
\end{aligned}$$

Nesse caso, $L_{\mathbf{x}}$ atinge valor máximo em $\left(\frac{1}{2}, \frac{x_2}{2(x_2+x_3)}, \frac{x_3}{2(x_2+x_3)}\right)$. (3.1)

Analogamente, se $\theta_2 \geq 1 - \frac{x_1+x_3}{2x_1}$ (nesse caso, $\frac{x_1}{n-x_2}(1-\theta_2) \leq \frac{1}{2}$) então a função

$$h(\theta_2) = \left(\frac{x_1}{n-x_2}(1-\theta_2) \right)^{x_1} \theta_2^{x_2} \left(1 - \theta_2 - \frac{x_1}{n-x_2}(1-\theta_2) \right)^{x_3}$$

é decrescente em $\left(1 - \frac{x_1+x_3}{2x_1}, \frac{1}{2}\right)$.

Portanto, nesse caso, $L_{\mathbf{x}}$ atinge valor máximo em $\left(\frac{1}{2}, \frac{1}{2} - \frac{x_3}{2x_1}, \frac{x_3}{2x_1}\right)$. (3.2)

Comparando os valores de $L_{\mathbf{x}}$ nos pontos obtidos em (3.1) e (3.2), concluímos que $\sup_{\theta \in \Theta_0} L_{\mathbf{x}}(\theta)$ se dá no ponto $\theta^* = \left(\frac{1}{2}, \frac{x_2}{2(x_2+x_3)}, \frac{x_3}{2(x_2+x_3)}\right)$ e, assim, a razão de verossimilhanças é dada por

$$RV(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L_{\mathbf{x}}(\theta)}{\sup_{\theta \in \Theta} L_{\mathbf{x}}(\theta)}$$

$$\begin{aligned}
&= \frac{\binom{n}{x_1, x_2, x_3} \left(\frac{1}{2}\right)^{x_1} \left(\frac{x_2}{2(x_2 + x_3)}\right)^{x_2} \left(\frac{x_3}{2(x_2 + x_3)}\right)^{n-x_1-x_2}}{\binom{n}{x_1, x_2, x_3} \left(\frac{x_1}{n}\right)^{x_1} \left(\frac{x_2}{n}\right)^{x_2} \left(\frac{x_3}{n}\right)^{n-x_1-x_2}} \\
&= \left(\frac{n}{2}\right)^n \left(\frac{1}{x_1}\right)^{x_1} \left(\frac{1}{n-x_1}\right)^{n-x_1}
\end{aligned}$$

A região C_x , neste caso, é dada por

$$\begin{aligned}
C_x &= \{\mathbf{z} \in \mathcal{X} : RV(\mathbf{z}) \leq RV(\mathbf{x})\} \\
&= \left\{ \mathbf{z} \in \mathcal{X} : \left(\frac{n}{2}\right)^n \left(\frac{1}{z_1}\right)^{z_1} \left(\frac{1}{n-z_1}\right)^{n-z_1} \leq \left(\frac{n}{2}\right)^n \left(\frac{1}{x_1}\right)^{x_1} \left(\frac{1}{n-x_1}\right)^{n-x_1} \right\} \\
&= \left\{ \mathbf{z} \in \mathcal{X} : (z_1)^{z_1} (n-z_1)^{n-z_1} \geq (x_1)^{x_1} (n-x_1)^{n-x_1} \right\} \\
&= \{\mathbf{z} \in \mathcal{X} : z_1 \geq x_1\}
\end{aligned}$$

A última passagem se deve ao fato de que a função $x^x(n-x)^{n-x}$ é crescente para $\forall x \in (n/2; n)$ pois sua derivada é $x^x(n-x)^{n-x} \ln \frac{x}{n-x} \geq 0$, $\forall x \in (n/2; n)$. Em particular, a função $x^x(n-x)^{n-x}$ é crescente para os valores inteiros nesse intervalo.

Como $X_k | \boldsymbol{\theta} \sim \text{Bin}(n, \theta_k)$

$$\begin{aligned}
p_1(\mathbf{x}) &= \sup_{\boldsymbol{\theta} \in \Theta_0} P(\mathbf{X} \in C_x | \boldsymbol{\theta}) = \sup_{\boldsymbol{\theta} \in \Theta_0} P(X_1 \geq x_1 | \boldsymbol{\theta}) = \sup_{\boldsymbol{\theta} \in \Theta_0} P(\text{Bin}(n, \theta_1) \geq x_1 | \boldsymbol{\theta}) \\
&= P(\text{Bin}(n, 1/2) \geq x_1)
\end{aligned}$$

□

Não é difícil mostrar que, nesse caso, esse teste também é o teste UMP para as referidas hipóteses. Em seguida, será especificado o *p-value* para o teste de ocorrência ou não de segundo turno.

Resultado 3.6. Seja $\mathbf{X}|\boldsymbol{\theta} \sim \text{Mult}(n, \boldsymbol{\theta})$. O p-value $p_2 : \mathcal{X} \rightarrow [0, 1]$ para testar

$H : \bigcap_{k=1}^3 \{\theta_k \leq 1/2\}$ contra $A : \bigcup_{k=1}^3 \{\theta_k > 1/2\}$ é dado, para cada $\mathbf{x} = (x_1, x_2, x_3) \in \mathcal{X}$, por

$$p_2(\mathbf{x}) = \begin{cases} 1 & \text{se } m(\mathbf{x}) \leq n/2 \\ 2P(\text{Bin}(n, 1/2) \geq m) & \text{se } m(\mathbf{x}) > n/2 \end{cases}$$

onde $m(\mathbf{x}) = \max(x_1, x_2, x_3)$.

Figura 3.2.

Figura 3.2: Região definida pela hipótese $H: \bigcap_{i=1}^3 \{\theta_i \leq 0.5\}$.

Demonstração. Se $m(\mathbf{x}) \leq n/2$ então, de forma análoga ao Resultado 3.5, temos

$$\sup_{\boldsymbol{\theta} \in \Theta_0} L(\mathbf{x}|\boldsymbol{\theta}) = \sup_{\boldsymbol{\theta} \in \Theta} L(\mathbf{x}|\boldsymbol{\theta}) = L(\mathbf{x}|\hat{\boldsymbol{\theta}}) \Rightarrow RV(\mathbf{x}) = 1 \Rightarrow C_x = \mathcal{X} \Rightarrow p_2(\mathbf{x}) = 1$$

Suponha agora, sem perda de generalidade, que $m(\mathbf{x}) = x_1 > n/2$. Temos (como no desenvolvimento do Resultado 3.5) que

- $\hat{\boldsymbol{\theta}} = \arg \sup_{\boldsymbol{\theta} \in \Theta} L_{\mathbf{x}}(\boldsymbol{\theta}) = \bar{\mathbf{x}}$,
- $\boldsymbol{\theta}^* = \arg \sup_{\boldsymbol{\theta} \in \Theta_0} L_{\mathbf{x}}(\boldsymbol{\theta}) = \left(\frac{1}{2}, \frac{x_2}{2(x_2 + x_3)}, \frac{x_3}{2(x_2 + x_3)} \right)$ e
- $RV(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L_{\mathbf{x}}(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} L_{\mathbf{x}}(\boldsymbol{\theta})} = \left(\frac{n}{2} \right)^n \left(\frac{1}{x_1} \right)^{x_1} \left(\frac{1}{n - x_1} \right)^{n - x_1}$.

No caso geral em que $m(\mathbf{x}) = \max(x_1, x_2, x_3) = m$ temos

$$RV(\mathbf{x}) = \left(\frac{n}{2} \right)^n \left(\frac{1}{m} \right)^m \left(\frac{1}{n - m} \right)^{n - m}$$

Considerando a variável aleatória $M(\mathbf{X}) = \max(X_1, X_2, X_3)$, temos, de forma análoga ao Resultado 3.5, que, para cada $\mathbf{x} \in \mathcal{X}$, a região $C_{\mathbf{x}}$ é dada por

$$C_{\mathbf{x}} = \left\{ \mathbf{z} \in \mathcal{X} : RV(\mathbf{z}) \leq RV(\mathbf{x}) \right\} = \left\{ \mathbf{z} \in \mathcal{X} : M(\mathbf{z}) \geq m(\mathbf{x}) \right\}$$

e, assim, o p -value $p_2(\mathbf{x})$ é

$$\begin{aligned} p_2(\mathbf{x}) &= \sup_{\boldsymbol{\theta} \in \Theta_0} P(\mathbf{X} \in C_{\mathbf{x}} | \boldsymbol{\theta}) = \sup_{\boldsymbol{\theta} \in \Theta_0} P(M(\mathbf{X}) \geq m(\mathbf{x}) | \boldsymbol{\theta}) = \sup_{\boldsymbol{\theta} \in \Theta_0} \sum_{k=1}^3 \sum_{j=m(\mathbf{x})}^n \binom{n}{j} \theta_k^j (1 - \theta_k)^{n-j} \\ &= 2P(\text{Bin}(n, 1/2) \geq m(\mathbf{x})) \quad (\text{ver Apêndice A.1}) \end{aligned}$$

□

Observe que, nesse caso, o p -value $p_2(\mathbf{x})$ é duas vezes o valor do p -value $p_1(\mathbf{x})$ para o caso anterior, onde queria-se testar se um candidato específico atingiria ou não a maioria absoluta dos votos válidos.

Dessa forma, suponha que uma pesquisa de intenção de voto foi conduzida para testar a ocorrência ou não de segundo turno em uma eleição com três candidatos, digamos A, B e C. Além disso, deseja-se verificar se o candidato A, que liderava as pesquisas, vencerá a eleição no primeiro turno. Assim, suponha que foi fixado que 1500 eleitores seriam entrevistados, de modo que seria observado o candidato de preferência de cada um deles. Realizada a pesquisa, o vetor de observações obtido foi

$\mathbf{x} = (784, 715, 1)$. Suponha também que foi fixado o nível de significância usual de 0,05. O p -value para testar a hipótese de não ocorrência de segundo turno, obtido por meio do Resultado 3.6, é $p_2(\mathbf{x}) = 0,0836$. Dessa forma, não há evidências para concluir que eleição acabará no primeiro turno. Entretanto, quando testa-se a hipótese do candidato A vencer no primeiro turno, baseado no mesmo vetor $\mathbf{x}$, obtém-se um p -value igual a $p_1(\mathbf{x}) = 0,0418$ (por meio do Resultado 3.5). Conclui-se então que há evidências de que o candidato A vencerá a eleição no primeiro turno, o que é, aparentemente, uma contradição.

Do ponto de vista matemático, isso também é algo incoerente. No primeiro caso, onde o interesse era verificar a ocorrência ou não de segundo turno, as hipóteses testadas foram $H_1 : \theta \in \Theta_0$ contra $A_1 : \theta \in \Theta_1$, com $\Theta_0 = \left\{ \theta \in \Theta : \bigcap_{k=1}^3 (\theta_k \leq 1/2) \right\}$ e $\Theta_1 = \Theta_0^c$, enquanto que no segundo caso, testou-se $H_2 : \theta \in \Theta'_0$ contra $A_2 : \theta \in \Theta'_1$, onde $\Theta'_0 = \{ \theta \in \Theta : \theta_1 \leq 1/2 \}$. Note que $\Theta'_0 \subseteq \Theta_0$, ou seja, a hipótese H_1 que não foi rejeitada é mais restritiva que a hipótese H_2 que foi rejeitada, apontando a ausência de “monotonicidade” dos testes da razão de verossimilhanças nesse cenário.

Essa “ilógica” dos testes clássicos já foi apontada anteriormente por outros autores. Por exemplo, em Royall [22], é apresentado o seguinte exemplo: considere que $X|\theta \sim \text{Bin}(n, \theta)$ e deseja-se testar $H : \theta = 0,5$ contra $A : \theta \neq 0,5$. Quando observa-se um valor $x > n/2$, deve-se rejeitar a hipótese H a um nível α se $P(X \geq x | \theta = 0,5) \leq \alpha/2$. Se, por outro lado, deseja-se testar a hipótese $H' : \theta \leq 0,5$ contra $A' : \theta > 0,5$, a observação x justifica a rejeição de H' se $P(X \geq x | \theta = 0,5) \leq \alpha$. Dessa forma, um valor x tal que $\alpha/2 < P(X \geq x | \theta = 0,5) \leq \alpha$ representa uma “forte evidência” para rejeitar a hipótese composta $\theta = 0,5$ ou $\theta < 0,5$ mas não justifica a rejeição da hipótese simples $\theta = 0,5$. Conclui-se então, ao nível de significância α , que $\theta = 0,5$ e $\theta < 0,5$ são ambas falsas mas não conclui-se que $\theta = 0,5$ é falsa isoladamente. Utilizando a linguagem da lógica, a conclusão seria “nem A nem B” mas não poderia ser “não A”. Alternativamente, quando rejeita-se $\theta = 0,5$, conclui-se que $\theta < 0,5$ ou $\theta > 0,5$, então qualquer evidência que justificasse a conclusão mais forte que $\theta > 0,5$ deveria levar-nos a mesma conclusão. Ou seja, a evidência que justifica a conclusão “A é verdadeiro”, certamente deveria justificar a conclusão mais fraca “ou A ou B é verdadeiro”.

Por fim, será apresentado o p -value para testar a ocorrência ou não de segundo turno sob o modelo multinomial negativo. Seja $\mathbf{Y}|\boldsymbol{\theta} \sim \text{MultNeg}(y_3, \boldsymbol{\theta})$ e $\mathbf{y} = (y_1, y_2)$ o valor observado de $\mathbf{Y}$ e considere o problema de testar as hipóteses $H : \bigcap_{k=1}^3 \{ \theta_k \leq 1/2 \}$ contra $A : \bigcup_{k=1}^3 \{ \theta_k > 1/2 \}$.

De forma análoga ao Resultado 3.5, temos que $p(\mathbf{y}) = 1$ quando $m(\mathbf{y}) \leq \frac{s(\mathbf{y})}{2}$, onde $m(\mathbf{y}) = \max(y_1, y_2, y_3)$ e $s(\mathbf{y}) = \sum_{i=1}^3 y_i$, pois

$$\sup_{\theta \in \Theta_0} L_{\mathbf{y}}(\theta) = \sup_{\theta \in \Theta} L_{\mathbf{y}}(\theta) = L_{\mathbf{y}}(\hat{\theta}) \Rightarrow RV(\mathbf{y}) = 1 \Rightarrow C_y = \mathcal{Y} \Rightarrow p(\mathbf{y}) = 1$$

Considere agora o caso em que $m(\mathbf{y}) > s(\mathbf{y})/2$. Suponha, sem perda de generalidade, que $\max(y_1, y_2, y_3) = y_1$. Assim, como no desenvolvimento do Resultado 3.6, temos que

- $\hat{\theta} = \arg \sup_{\theta \in \Theta} L_{\mathbf{y}}(\theta) = \bar{\mathbf{y}}$ e
- $\theta^* = \arg \sup_{\theta \in \Theta_0} L_{\mathbf{y}}(\theta) = \left(\frac{1}{2}, \frac{y_2}{2(y_2 + y_3)}, \frac{y_3}{2(y_2 + y_3)} \right)$

Deste modo, no caso geral em que $m(\mathbf{y}) = \max(y_1, y_2, y_3)$, temos

$$RV(\mathbf{y}) = \frac{\sup_{\theta \in \Theta_0} L_{\mathbf{y}}(\theta)}{\sup_{\theta \in \Theta} L_{\mathbf{y}}(\theta)} = \left(\frac{s(\mathbf{y})}{2m(\mathbf{y})} \right)^{m(\mathbf{y})} \left(\frac{s(\mathbf{y})}{2(s(\mathbf{y}) - m(\mathbf{y}))} \right)^{s(\mathbf{y}) - m(\mathbf{y})}$$

Assim, a região C_y é dada por

$$\begin{aligned} C_y &= \left\{ \mathbf{u} \in \mathcal{Y} : RV(\mathbf{u}) \leq RV(\mathbf{y}) \right\} = \\ &= \left\{ \mathbf{u} \in \mathcal{Y} : \left(\frac{s(\mathbf{u})}{2m(\mathbf{u})} \right)^{m(\mathbf{u})} \left(\frac{s(\mathbf{u})}{2(s(\mathbf{u}) - m(\mathbf{u}))} \right)^{s(\mathbf{u}) - m(\mathbf{u})} \leq \left(\frac{s(\mathbf{y})}{2m(\mathbf{y})} \right)^{m(\mathbf{y})} \left(\frac{s(\mathbf{y})}{2(s(\mathbf{y}) - m(\mathbf{y}))} \right)^{s(\mathbf{y}) - m(\mathbf{y})} \right\} \\ &= \left\{ \mathbf{u} \in \mathcal{Y} : \left(\frac{s(\mathbf{u})}{2m(\mathbf{u})} \right)^{m(\mathbf{u})} \left(\frac{s(\mathbf{u})}{2(s(\mathbf{u}) - m(\mathbf{u}))} \right)^{s(\mathbf{u}) - m(\mathbf{u})} \leq RV(\mathbf{y}) \right\} \end{aligned}$$

A Figura 3.3 a seguir apresenta a região C_y para $\mathbf{y} = (10, 1)$ com $y_3 = 2$.

Figura 3.3.

Figura 3.3: Região C_y no modelo Multinomial Negativo.

Dessa forma, o p -value é dado por

$$p(\mathbf{y}) = \sup_{\theta \in \Theta_0} P(Y \in C_y | \theta) = \theta_3^{y_3} \sum_{(i,j) \in C_y} \binom{i+j+y_3-1}{i, j, y_3-1} \theta_1^i \theta_2^j \quad (3.3)$$

O tratamento analítico da soma em (3.3) é bastante complicado, o que torna a obtenção do valor exato do p -value inviável nesse caso. Alternativamente, foi desenvolvida uma função no software R [18] (ver Apêndice B) para calcular a $P(Y \in C_y | \theta)$ por simulação, para valores de $\mathbf{y} \in \mathcal{Y}$ e de $\boldsymbol{\theta} \in \Theta_0$ fixados. Contudo, a obtenção do $\sup_{\Theta_0} P(Y \in C_y | \boldsymbol{\theta})$ é comprometida em decorrência da imprecisão do cálculo de $P(Y \in C_y | \boldsymbol{\theta})$ para valores de θ_3 menores que o ϵ de máquina (o ϵ de máquina é o menor valor tal que $1 + \epsilon \neq 1$ e, para as máquinas utilizadas, o ϵ era em torno de 2.2×10^{-16}).

Para ilustrar esse problema, considere que a observação $\mathbf{x} = (10, 1, 2)$ foi obtida da distribuição

$Mult(13, \theta)$. Nesse caso, o p -value obtido por meio do Resultado 3.6 é 0,092. Considere agora que a observação $\mathbf{y} = (10, 1)$ foi obtida da distribuição $MultNeg(2, \theta)$, gerando uma função de verossimilhança proporcional ao caso multinomial citado anteriormente. Nesse caso, o $\sup_{\Theta_0} P(Y \in C_y | \theta)$ aparentemente ocorre quando θ_3 se aproxima de zero (ver Figura 3.4) e parece valer algo em torno de 0,0437 em um ponto da forma $\theta^o = (0, 5; 0, 5 - \epsilon; \epsilon)$. Contudo, não podemos garantir que esse é o valor procurado. Deste modo, a avaliação de $P(Y \in C_y | \theta)$ em Θ_0 é bastante limitada mesmo via simulação.

Figura 3.4.

Figura 3.4: $P(Y \in C_y | \theta)$ para o modelo multinomial, obtida por simulação

Apenas para ilustrar a violação ao princípio da verossimilhança pelo teste da razão de verossimilhanças, vamos considerar uma modificação nos modelos multinomial e multinomial negativo para contornar o problema apresentado anteriormente. Consideremos agora os experimentos descritos a seguir.

- Observar o candidato preferido de n eleitores, com $n \in \mathbb{N} \setminus \{0\}$ fixado, como no início da seção. Contudo, vamos impor que o terceiro candidato deve ter ao menos um voto ($x_3 \geq 1$). Nesse caso, o espaço amostral é $\mathcal{X}' = \{(x_1, x_2) \in \mathbb{N}^2 : x_1 + x_2 \leq n - 1\}$ e a distribuição do vetor $\mathbf{X}$

com as intenções de voto, condicionado ao conhecimento de $\boldsymbol{\theta}$, tem distribuição *multinomial truncada* dada por

$$f(\mathbf{x}|\boldsymbol{\theta}) = \binom{n}{x_1, x_2, x_3} \frac{\theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}}{1 - (1 - \theta_3)^n}, \quad (3.4)$$

tal que $\theta_i \geq 0$, $\forall i \in \{1, 2, 3\}$, $\theta_1 + \theta_2 \leq 1$ e $\theta_3 = 1 - \theta_1 - \theta_2$.

- Observar o candidato preferido dos eleitores até que o terceiro candidato obtenha y_3 votos, com $y_3 \in \mathbb{N} \setminus \{0\}$ fixado, como feito anteriormente. Entretanto, agora vamos supor que $y_3 = 1$ e impor que a soma dos votos dos dois primeiros candidatos até a ocorrência de um voto para o candidato 3 deve ser no máximo t , com $t \in \mathbb{N}$ fixado. Nesse caso, o espaço amostral é $\mathcal{Y}' = \{(y_1, y_2) \in \mathbb{N}^2 : y_1 + y_2 \leq t\}$ e a distribuição do vetor $\mathbf{Y}$ com as intenções de voto, condicionado ao conhecimento de $\boldsymbol{\theta}$, tem distribuição *multinomial negativa truncada* dada por

$$f(\mathbf{y}|\boldsymbol{\theta}) = \binom{y_1 + y_2}{y_1} \frac{\theta_1^{y_1} \theta_2^{y_2} \theta_3}{1 - (1 - \theta_3)^{t+1}}, \quad (3.5)$$

tal que $\theta_i \geq 0$, $\forall i \in \{1, 2, 3\}$, $\theta_1 + \theta_2 \leq 1$ e $\theta_3 = 1 - \theta_1 - \theta_2$.

Devemos salientar que reescrevendo (3.4) e (3.5) de forma alternativa, os modelos Multinomial Truncado e Multinomial Negativo Truncado estão bem definidos para $\theta_3 = 0$.

Suponha que, no primeiro caso, foi fixado $n = 5$ e foi observado o ponto amostral $\mathbf{x} = (4, 0, 1)$. A função de verossimilhança nesse caso é $L_x(\boldsymbol{\theta}) = \frac{5 \theta_1^4 \theta_3}{1 - (1 - \theta_3)^5}$. Suponha que, no segundo experimento, foi fixado $y_3 = 1$ e $t = 4$ e foi observado o ponto amostral $\mathbf{y} = (4, 0)$, de modo que a função de verossimilhança é $L_y(\boldsymbol{\theta}) = \frac{\theta_1^4 \theta_3}{1 - (1 - \theta_3)^5} \propto L_x(\boldsymbol{\theta})$.

O cálculo exato da razão de verossimilhanças para alguns pontos é apresentada no Apêndice A.2. Para os pontos onde a razão de verossimilhanças não foi calculada analiticamente, é possível obter um valor aproximado para a mesma utilizando a função desenvolvida em R [18] e apresentada no Apêndice B.2. As Tabelas 3.1 e 3.2 apresentam, respectivamente, os valores da razão de verossimilhanças para os modelos multinomial truncado e multinomial negativo truncado. Os valores em negrito indicam, em cada caso, os pontos “mais desfavoráveis” à hipótese nula que o correspondente ponto observado sob a ordenação do espaço amostral por meio da razão de verossimilhanças.

	0	1	2	3	4
0	1/31	0,393	0,923	10/31	5/31
1	0,393	0,922	1	16/27	
2	0,923	1	1		
3	10/31	16/27			
4	5/31				

Tabela 3.1: RV para o modelo multinomial truncado.

	0	1	2	3	4
0	32/62	0,989	20/31	10/31	5/31
1	0,989	1	27/32	16/27	
2	20/31	27/32	1		
3	10/31	16/27			
4	5/31				

Tabela 3.2: RV para o modelo multinomial negativo truncado.

Com base na Tabela 3.2, o p -value $p_2(4, 0)$ para o caso da multinomial negativa truncada é dado por:

$$p_2(4, 0) = \sup_{\theta \in \Theta_0} P(\mathbf{Y} \in \{(4, 0), (0, 4)\} | \boldsymbol{\theta}) = \sup_{\theta \in \Theta_0} \left(\frac{\theta_1^4 \theta_3 + \theta_2^4 \theta_3}{1 - (1 - \theta_3)^5} \right)$$

Vamos mostrar que $\sup_{\Theta_0} P(\mathbf{Y} \in \{(4, 0), (0, 4)\} | \boldsymbol{\theta})$ se dá em $(0, 1/2, 1/2)$.

Seja $\theta_3 \in [0; 1/2]$ fixado e $g_{\theta_3}(\theta_1) = \theta_1^4 + (1 - \theta_1 - \theta_3)^4$. Note que, fixado θ_3 , maximizar $\frac{\theta_1^4 \theta_3 + \theta_2^4 \theta_3}{1 - (1 - \theta_3)^5}$ é equivalente a maximizar g_{θ_3} . Assim, para $\theta_1 \in (0, 1/2)$, temos

$$g'_{\theta_3}(\theta_1) = 4 \left(\theta_1^3 - (1 - \theta_1 - \theta_3)^4 \right) > 0 \iff \theta_1 > \frac{1 - \theta_3}{2}$$

Como $\frac{1 - \theta_3}{2} < 1/2$, segue que g_{θ_3} é máximo quando $\theta_1 = 0$ ou $\theta_1 = 1/2$. Comparando g_{θ_3} nesses pontos, temos

$$g_{\theta_3}(0) > g_{\theta_3}(1/2) \iff (1 - \theta_3)^4 > \left(\frac{1}{2} \right)^4 + \left(\frac{1}{2} - \theta_3 \right)^4 \iff 16\theta_3^3 - 36\theta_3^2 + 28\theta_3 - 7 < 0$$

Seja $h(\theta_3) = 16\theta_3^3 - 36\theta_3^2 + 28\theta_3 - 7$. Temos que $h(0) = -7$ e $h(1/2) = 0$. Vamos mostrar que h é crescente e, portanto, $g_{\theta_3}(0) > g_{\theta_3}(1/2)$, $\forall \theta_3 \in (0, 1/2)$.

$h'(\theta_3) = 4 [12\theta_2^2 - 18\theta_3 + 7] > 0$. Logo, h é crescente, $h(\theta_1) < 0$ e, portanto, $g_{\theta_3}(0) > g_{\theta_3}(1/2)$

$$\Rightarrow \arg \sup g_{\theta_3}(\theta_1) = 0.$$

Dessa forma, considere $f(\theta_2) = \frac{\theta_2^4(1-\theta_2)}{1-\theta_2^5}$ e note que $f(0) = 0$ e $f(1_-) = 1/5$.

Vamos mostrar que f é crescente no intervalo $(0, 1)$.

$$f'(\theta_2) = \frac{(4\theta_2^3 - 5\theta_2^4)(1 - \theta_2^5) + 5\theta_2^4(\theta_2^4 - \theta_2^5)}{[1 - \theta_2^5]^2} = \frac{\theta_2^3}{[1 - \theta_2^5]^2} (4 - 5\theta_2 + \theta_2^5)$$

$$f'(\theta_2) > 0 \iff u(\theta_2) = 4 - 5\theta_2 + \theta_2^5 > 0$$

Mas $u(0) = 4$, $u(1) = 0$ e $u'(\theta_2) = -5(1 - \theta_2^4) < 0$, $\forall \theta_2 \in (0, 1)$. Portanto, $u > 0$ e f é crescente no intervalo $(0, 1)$, de forma que $\sup_{\Theta_0} P(\mathbf{Y} \in \{(4, 0), (0, 4)\} | \boldsymbol{\theta})$ se dá em $(0, 1/2, 1/2)$.

Assim, o p -value para esse caso é $p_2(4, 0) = \frac{1}{31} \approx 0,032$.

Com base na Tabela 3.1, o p -value $p_1(4, 0, 1)$ para o caso da multinomial truncada é dado por:

$$p_1(4, 0, 1) = \sup_{\boldsymbol{\theta} \in \Theta_0} P(\mathbf{X} \in \{(0, 0, 5), (4, 0, 1), (0, 4, 1)\} | \boldsymbol{\theta}) = \sup_{\boldsymbol{\theta} \in \Theta_0} \left(\frac{\theta_3^5 + 5\theta_1^4\theta_3 + 5\theta_2^4\theta_3}{1 - (1 - \theta_3)^5} \right),$$

$$\text{de modo que } p_1(4, 0, 1) \geq \left(\frac{\theta_3^5 + 5\theta_1^4\theta_3 + 5\theta_2^4\theta_3}{1 - (1 - \theta_3)^5} \right) \Big|_{(0, \frac{1}{2}, \frac{1}{2})} = \frac{6(\frac{1}{2})^5}{1 - (\frac{1}{2})^5} = \frac{6}{31} \approx 0,194$$

Nesse exemplo é possível verificar que o p -value viola o princípio da verossimilhança. Se fixarmos um nível de significância de 0,05, concluiríamos, sob modelo multinomial negativo truncado, que não haverá segundo turno, enquanto que sob o modelo multinomial, a conclusão seria que nenhum dos candidatos alcançou a maioria absoluta dos votos válidos e, portanto, haverá segundo turno. Devemos ressaltar que o p -value é bem diferente nestes dois casos, de modo que $p_1(4, 0, 1) \geq 6 p_2(4, 0)$.

3.3 Exemplo: Abordagem Bayesiana

Nessa seção serão apresentados os testes de hipóteses bayesianos para os problemas propostos anteriormente. Para tanto, considere a distribuição Dirichlet, definida a seguir.

Definição 3.3. *Seja $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ um vetor aleatório cujas componentes assumem valores reais não negativos tal que $\sum_{i=1}^k \theta_i \leq 1$. Diz-se que $\boldsymbol{\theta}$ tem **distribuição Dirichlet** com parâmetros $\mathbf{d} = (d_1, d_2, \dots, d_k; d_{k+1})$ (denota-se $\boldsymbol{\theta} \sim \text{Dir}(\mathbf{d})$) se a função de densidade de probabilidade de $\boldsymbol{\theta}$ é dada por:*

$$f(\boldsymbol{\theta}|\mathbf{d}) = \frac{\Gamma\left(\sum_{i=1}^{k+1} d_i\right)}{\prod_{i=1}^{k+1} \Gamma(d_i)} \prod_{i=1}^{k+1} \theta_i^{d_i-1}, \quad (3.6)$$

tal que $d_i \in \mathbb{R}_+$, $\forall i \in \{1, \dots, k+1\}$, $\theta_i \geq 0$, $\forall i \in \{1, \dots, k\}$, $\sum_{i=1}^k \theta_i \leq 1$ e $\theta_{k+1} = 1 - \sum_{i=1}^k \theta_i$.

Em particular, quando $k = 1$, temos a distribuição Beta($d_1; d_2$).

Resultado 3.7. *Considere o vetor aleatório $\boldsymbol{\theta}$ com distribuição Dirichlet, como na Definição 3.3. Então $\theta_i \sim \text{Beta}\left(d_i, \sum_{j \neq i} d_j\right)$, $i \in \{1, 2, \dots, k+1\}$.*

A distribuição Dirichlet é bastante maleável e acomoda diversos tipos de opiniões sobre um vetor de proporções $(\theta_1, \theta_2, \dots, \theta_k)$ tal que $\theta_i \geq 0$, $i = 1, \dots, k$, $\sum_{i=1}^k \theta_i \leq 1$, além de ser de fácil tratamento analítico. Em particular, quando $d_i = 1$, $\forall i \in \{1, 2, \dots, k+1\}$, temos a distribuição uniforme k -variada definida no conjunto de interesse $\left\{(\theta_1, \theta_2, \dots, \theta_k) : \theta_i \geq 0, i = 1, \dots, k, \sum_{i=1}^k \theta_i \leq 1\right\}$. Dessa forma, o modelo Dirichlet será adotado como distribuição a priori para $\boldsymbol{\theta}$ ao longo dessa seção. Além disso, a utilização da distribuição Dirichlet facilita a obtenção da distribuição posterior nos exemplos discutidos ao longo desse capítulo, fato enunciado no Resultado 3.8 a seguir.

Resultado 3.8. *A distribuição Dirichlet é conjugada das distribuições multinomial e multinomial negativa. Isto é, considere que a priori $\boldsymbol{\theta} \sim \text{Dir}(\mathbf{d})$, com $\mathbf{d} = (d_1, d_2, \dots, d_k; d_{k+1})$, e observe $\mathbf{X} = (X_1, X_2, \dots, X_k)$ que, condicionado ao conhecimento de $\boldsymbol{\theta}$, tem distribuição $\text{Mult}(n, \boldsymbol{\theta})$ (ou $\text{MultNeg}(x_{k+1}, \boldsymbol{\theta})$). Então, temos que a distribuição a posterior de $\boldsymbol{\theta}|\mathbf{X} = \mathbf{x}$ é $\text{Dir}(\mathbf{d} + \mathbf{x})$.*

Demonstração. Suponha que $\mathbf{X}|\boldsymbol{\theta} \sim \text{Multinomial}(n, \boldsymbol{\theta})$ e $\boldsymbol{\theta} \sim \text{Dirichlet}(\mathbf{d})$. Então

$$\begin{aligned}
L_{\mathbf{x}}(\boldsymbol{\theta}) &\propto \prod_{i=1}^{k+1} \theta_i^{x_i} \\
f(\boldsymbol{\theta}) &= \frac{\Gamma\left(\sum_{i=1}^{k+1} d_i\right)}{\prod_{i=1}^{k+1} \Gamma(d_i)} \prod_{i=1}^{k+1} \theta_i^{d_i-1} \propto \prod_{i=1}^{k+1} \theta_i^{d_i-1} \\
f(\boldsymbol{\theta}|\mathbf{x}) &= \frac{L_{\mathbf{x}}(\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int_{\Theta} L_{\mathbf{x}}(\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto L_{\mathbf{x}}(\boldsymbol{\theta})f(\boldsymbol{\theta}) \propto \prod_{i=1}^{k+1} \theta_i^{d_i+x_i-1} \propto \frac{\Gamma\left(\sum_{i=1}^{k+1} d_i + x_i\right)}{\prod_{i=1}^{k+1} \Gamma(d_i + x_i)} \prod_{i=1}^{k+1} \theta_i^{d_i+x_i-1} \\
&\Rightarrow \boldsymbol{\theta}|\mathbf{X} = \mathbf{x} \sim \text{Dir}(\mathbf{d} + \mathbf{x})
\end{aligned}$$

A demonstração é análoga para a distribuição Multinomial Negativa.

□

Vamos assumir como distribuição a priori para as proporções de votos o modelo *Dirichlet*($\mathbf{d}$), onde $\mathbf{d} = (d_1, \dots, d_k; d_{k+1})$. Pelo Resultado 3.8, após observar-se $X = \mathbf{x}$, a distribuição a posteriori será *Dirichlet*($\mathbf{d} + \mathbf{x}$), independente da escolha do primeiro ou do segundo esquema amostral. Como já foi dito anteriormente, se $d_i = 1, \forall i \in \{1, \dots, k+1\}$, temos a distribuição uniforme definida no conjunto de interesse. Note que nesse particular caso, a distribuição posterior é proporcional à função de verossimilhança gerada pelo ponto $\mathbf{x}$.

A seguir, será apresentado o procedimento bayesiano para testar se o j -ésimo candidato vencerá ou não a eleição no primeiro turno e, na sequência, o teste para a ocorrência ou não de segundo turno.

Resultado 3.9. Considere que, a priori, $\boldsymbol{\theta} \sim \text{Dir}(\mathbf{d})$, onde $\mathbf{d} = (d_1, d_2, d_3)$, e seja $\mathbf{X}|\boldsymbol{\theta} \sim \text{Mult}(n, \boldsymbol{\theta})$ e $\mathbf{x} = (x_1, x_2, x_3)$ o valor observado de $\mathbf{X}$, de forma que $\boldsymbol{\theta}|\mathbf{X} = \mathbf{x} \sim \text{Dir}(\mathbf{d} + \mathbf{x})$. O procedimento de bayes para testar $H: \theta_j \leq 1/2$ contra $A: \theta_j > 1/2$, com relação a função de perda apresentada na Tabela 2.2, consiste em rejeitar H se, e somente se,

$$q_1(\mathbf{x}) = 1 - P(\theta_j > 1/2|\mathbf{x}) = 1 - P(\text{Beta}(d_j + x_j, s_{-j}) > 1/2) < \frac{a_1 - b_1}{(a_0 + a_1) - (b_0 + b_1)},$$

onde $s_{-j} = \sum_{i \neq j} (d_i + x_i)$.

Demonstração. Basta considerar os Resultados 3.8 e 3.7 e a Proposição 2.1. □

Resultado 3.10. Seja $\mathbf{X}|\boldsymbol{\theta} \sim \text{Mult}(n, \boldsymbol{\theta})$ e $\mathbf{x} = (x_1, x_2, x_3)$ o valor observado de $\mathbf{X}$. Considere que, a priori, $\boldsymbol{\theta} \sim \text{Dir}(\mathbf{d})$, onde $\mathbf{d} = (d_1, d_2, d_3)$, de forma que $\boldsymbol{\theta}|\mathbf{X} = \mathbf{x} \sim \text{Dir}(\mathbf{d} + \mathbf{x})$. O procedimento de bayes para testar $H: \bigcap_{k=1}^3 \{\theta_k \leq 1/2\}$ contra $A: \bigcup_{k=1}^3 \{\theta_k > 1/2\}$, com relação a função de perda apresentada na Tabela 2.2, é rejeitar H se, e somente se,

$$q_2(\mathbf{x}) = 1 - P\left(\bigcup_{k=1}^3 \{\theta_k > 1/2\} | \mathbf{x}\right) = 1 - \sum_{k=1}^3 P(\text{Beta}(d_k + x_k, s_{-k}) > 1/2) < \frac{a'_1 - b'_1}{(a'_0 + a'_1) - (b'_0 + b'_1)},$$

onde $s_{-k} = \sum_{j \neq k} (d_j + x_j)$.

Demonstração. Primeiramente, note que

$$P(\boldsymbol{\theta} \in \Theta_1 | \mathbf{x}) = P\left(\bigcup_{k=1}^3 \{\theta_k > 1/2\} | \mathbf{x}\right) = \sum_{k=1}^3 P(\{\theta_k > 1/2\} | \mathbf{x}).$$

Agora, basta considerar os Resultados 3.8 e 3.7 e a Proposição 2.1. □

Na seção anterior, apresentamos o seguinte exemplo para verificar a ausência da propriedade da “monotonicidade” nos testes clássicos: uma pesquisa de intenção de voto foi conduzida para avaliar o desempenho de três candidatos, digamos A, B e C. Foi fixado que 1500 eleitores seriam entrevistados, de modo que seria observado o candidato de preferência de cada um deles. Realizada a pesquisa, o vetor de observações obtido foi $\mathbf{x} = (784, 715, 1)$. O p -value para testar a hipótese de não ocorrência de segundo turno, obtido por meio do Resultado 3.6, foi $p_2(\mathbf{x}) = 0,0836$. Dessa forma, a um nível de significância usual de 0,05, não há evidências para concluir que eleição acabará no primeiro turno. Contudo, ao testar a hipótese do candidato A vencer no primeiro turno, baseado no mesmo vetor $\mathbf{x}$, obtém-se um p -value igual a $p_1(\mathbf{x}) = 0,0418$ (por meio do Resultado 3.5). Conclui-se então que há evidências de que o candidato A vencerá a eleição no primeiro turno, o que é, aparentemente, uma contradição. Isso se deve ao fato de que o p -value clássico, calculado com base na estatística da razão de verossimilhanças, não apresenta a propriedade da “monotonicidade”, fato que pode nos levar a conclusões contraditórias (ou “ilógicas”).

Considerando esse mesmo exemplo, vamos apresentar os testes sob a abordagem bayesiana. Suponha que, a priori, $\boldsymbol{\theta} \sim \text{Dir}(1, 1; 1)$, de modo que $\boldsymbol{\theta}|\mathbf{x} \sim \text{Dir}(785, 716; 2)$. Para testar a hipótese de não ocorrência de segundo turno, temos, por meio do Resultado 3.10, que $q_2(\mathbf{x}) = 0,0042$. Analogamente, utilizando o Resultado 3.9, obtemos $q_1(\mathbf{x}) = 0,0395$ para testar a hipótese do candidato A vencer no primeiro turno. Assim, diferentemente do p -value clássico apresentado anteriormente, o teste bayesiano apresenta a propriedade da “monotonicidade”, de modo que $q_2(\mathbf{x}) \leq q_1(\mathbf{x})$. Esse fato se verifica para todo $\mathbf{x} \in \mathcal{X}$ (basta olhar os Resultados 3.9 e 3.10). Além disso, considerando pares de funções de perda bastante razoáveis (por exemplo, que sejam proporcionais à distância de ponto à hipótese), a monotonicidade é verificada, em geral, pelos testes bayesianos baseados na probabilidade a posteriori.

Como já foi dito, os procedimentos bayesianos se baseiam somente na distribuição posterior para $\boldsymbol{\theta}$, e esta, por sua vez, depende dos dados somente através do núcleo (parte que depende de $\boldsymbol{\theta}$) da função de verossimilhança. Deste modo, não há violação do princípio da verossimilhança. No particular caso dos testes apresentados, dada uma função de verossimilhança $L(\mathbf{x}|\boldsymbol{\theta}) \propto \prod_j \theta_j^{x_j}$, não importa se a observação $\mathbf{x}$ provém do modelo $\text{Mult}(n, \boldsymbol{\theta})$, com $n = x_1 + x_2 + x_3$, ou do modelo $\text{MultNeg}(x_3, \boldsymbol{\theta})$ (veja, por exemplo, Resultado 3.8), ou seja, a “regra de parada” não é importante nesse caso (Barnett [1]; Berger e Wolpert [3]).

Capítulo 4

Conclusões

4.1 Considerações Finais

Neste trabalho foi realizado um estudo comparativo de testes de hipóteses sob as abordagens clássica e bayesiana. Foram desenvolvidos o *p-value* clássico, determinado a partir do teste da razão de verossimilhanças, e, sob a abordagem bayesiana, a probabilidade a posteriori das hipóteses de interesse. Em particular, estudou-se esses procedimentos de teste em relação a duas características: a propriedade da “monotonicidade” e o princípio da verossimilhança.

Os testes clássicos da razão de verossimilhanças (em particular, o *p-value*) foram desenvolvidos em um caso pouco ilustrado na literatura, onde a hipótese (nula) é composta e não precisa (Berger e Delampady [2]). Além disso, não se fez uso de resultados assintóticos, usualmente utilizados em problemas onde a obtenção do teste exato não é trivial. A escolha da estatística da razão de verossimilhanças para a ordenação do espaço paramétrico (Casella e Berger [4]; Dempster [7]) foi feita pela sua importância nos procedimentos clássicos de teste de hipóteses e pela inexistência, no cenário considerado, de testes uniformemente mais poderosos. Do mesmo modo, a escolha do procedimento de teste utilizado sob a abordagem bayesiana foi feita por se tratar de um método bastante difundido (DeGroot [6]; Raiffa e Schlaifer [19]) e coerente com a teoria das decisões.

Nota-se que, em geral, os métodos clássicos não apresentam a propriedade da “monotonicidade”. Uma implicação direta desse resultado é que as conclusões de dois testes realizados com base na mesma observação podem ser contraditórias, como as apresentadas na página 26 do Capítulo 3.

Por meio da Proposição 2.1, é fácil verificar que os testes bayesianos apresentam a propriedade da “monotonicidade” pois baseiam-se somente na distribuição posterior para θ . Suponha que há

interesse em realizar dois testes de hipóteses: no primeiro, o objetivo é testar $H_1 : \boldsymbol{\theta} \in \Theta_0$ contra $A_1 : \boldsymbol{\theta} \in \Theta_1 = \Theta_0^c$, e, no segundo, deseja-se testar $H_2 : \boldsymbol{\theta} \in \Theta'_0$ contra $A_1 : \boldsymbol{\theta} \in \Theta'_1$, de modo que $\Theta_0 \subseteq \Theta'_0 \subseteq \Theta$. O teste bayesiano baseia-se no cálculo da probabilidade posterior do subconjunto do espaço paramétrico que define a hipótese nula e, assim, temos que $P(\boldsymbol{\theta} \in \Theta_0 | \mathbf{x}) \leq P(\boldsymbol{\theta} \in \Theta'_0 | \mathbf{x})$. Para pares de funções de perda bastante razoáveis, a propriedade da “monotonicidade” é satisfeita pelos testes bayesianos baseados na probabilidade a posteriori. Para casos em que a hipótese nula é precisa (com probabilidade posterior igual a zero), existem testes bayesianos alternativos, por exemplo o Teste de Significância Totalmente Bayesiano (*Full Bayesian Significance Test* ou FBST) (Pereira e Stern [16]). Conforme descrito por Stern [25], esse teste também apresenta a propriedade da “monotonicidade”.

Destaca-se também nesse trabalho que a inferência clássica viola o princípio da verossimilhança, algo bem conhecido na literatura (Barnett [1]; Berger e Wolpert [3]; Inoue [9]). Alguns estatísticos clássicos não concordam com o princípio da verossimilhança ou baseiam-se em outros princípios (como o princípio da confiança), outros simplesmente ignoram a existência do mesmo. Em Berger e Wolpert [3], incoerências do princípio da confiança são apresentadas, bem como um estudo detalhado do princípio da verossimilhança e as implicações lógicas de sua violação.

Os procedimentos bayesianos, em geral, não violam o princípio da verossimilhança pois se baseiam somente na distribuição posterior para $\boldsymbol{\theta}$, e esta, por sua vez, depende dos dados somente através do núcleo da função de verossimilhança. Este resultado também é amplamente descrito na literatura estatística.

4.2 Sugestões para Pesquisas Futuras

- Estudar funções de perda “coerentes”, isto é, que tornam os procedimentos bayesianos concordes com a propriedade da “monotonicidade”.
- Estudar testes de hipóteses “conjuntamente coerentes”.

Apêndice A

Demonstrações Adicionais

A.1 *p-value* da Multinomial

Nessa seção, iremos mostrar que o *p-value* para testar se a hipótese $H : \bigcap_{i=1}^3 \left\{ \theta_i \leq \frac{1}{2} \right\}$ contra $A : \bigcup_{i=1}^3 \left\{ \theta_i > \frac{1}{2} \right\}$ sob o modelo multinomial proposto, é dado por

$$p_2(\mathbf{x}) = \sup_{\theta \in \Theta_0} P \left(\bigcup_{i=1}^3 X_i \geq m | \boldsymbol{\theta} \right) = 2P(\text{Bin}(n, 1/2) \geq m),$$

onde $m = m(\mathbf{x}) = \max(x_1, x_2, x_3)$, se $m > n/2$ (o caso $m \leq n/2$ já foi detalhado na demonstração do Resultado 3.5).

Estamos interessados em obter o supremo de

$$\begin{aligned} f(\boldsymbol{\theta}) &= P \left(\bigcup_{i=1}^3 \{X_i \geq m\} | \boldsymbol{\theta} \right) \\ &= \sum_{j=m}^n \binom{n}{j} \left[\theta_1^j (1 - \theta_1)^{n-j} + \theta_2^j (1 - \theta_2)^{n-j} + (1 - \theta_1 - \theta_2)^j (\theta_1 + \theta_2)^{n-j} \right], \end{aligned}$$

com $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta_0 = \{ \boldsymbol{\theta} \in \mathbb{R}^2 / 0 \leq \theta_i \leq \frac{1}{2}, \forall i \in \{1, 2\}, \theta_1 + \theta_2 \leq 1 \}$ e $m \in \{n/2 + 1, \dots, n\}$.

Fixando $\theta_1^o \in (0, 1/2]$, definimos então $h_{\theta_1^o}(\theta_2) = f(\theta_1^o, \theta_2)$. Vamos mostrar na sequência que

$\forall j \in \{m, m+1, \dots, n\}$, $g_{\theta_1^o}^{(j)}(\theta_2) = (\theta_2)^j(1-\theta_2)^{n-j} + (1-\theta_1^o-\theta_2)^j(\theta_1^o+\theta_2)^{n-j}$ atinge valor máximo em $\theta_2 = 1/2 - \theta_1^o$ ou $\theta_2 = 1/2$. Isso implicará que $h_{\theta_1^o}$ também atingirá máximo nesses pontos.

Lema A.1.1. Fixado θ_1^o em $(0, 1/2]$, a função $g_{\theta_1^o}^{(j)}(\theta_2) = (\theta_2)^j(1-\theta_2)^{n-j} + (1-\theta_1^o-\theta_2)^j(\theta_1^o+\theta_2)^{n-j}$ é simétrica em torno de $\left(\frac{1-\theta_1^o}{2}\right)$.

Demonstração. Temos que, $\forall \epsilon > 0$,

$$\begin{aligned} & g_{\theta_1^o}^{(j)}\left(\frac{1-\theta_1^o}{2} - \epsilon\right) \\ &= \left(\frac{1-\theta_1^o}{2} - \epsilon\right)^j \left(1 - \left(\frac{1-\theta_1^o}{2} - \epsilon\right)\right)^{n-j} + \left(1 - \theta_1^o - \left(\frac{1-\theta_1^o}{2} - \epsilon\right)\right)^j \left(\theta_1^o + \frac{1-\theta_1^o}{2} - \epsilon\right)^{n-j} \\ &= \left(\frac{1-\theta_1^o}{2} + \epsilon\right)^j \left(\frac{1+\theta_1^o}{2} - \epsilon\right)^{n-j} + \left(\frac{1-\theta_1^o}{2} - \epsilon\right)^j \left(\frac{1+\theta_1^o}{2} + \epsilon\right)^{n-j} \\ &= \left(\frac{1-\theta_1^o}{2} + \epsilon\right)^j \left(1 - \left(\frac{1-\theta_1^o}{2} + \epsilon\right)\right)^{n-j} + \left(1 - \theta_1^o - \left(\frac{1-\theta_1^o}{2} + \epsilon\right)\right)^j \left(\theta_1^o + \frac{1-\theta_1^o}{2} + \epsilon\right)^{n-j} \\ &= g_{\theta_1^o}^{(j)}\left(\frac{1-\theta_1^o}{2} + \epsilon\right) \end{aligned}$$

□

Teorema A.1.1. $h_{\theta_1^o}(\theta_2)$ assume seu valor máximo sob o conjunto Θ_0 quando $\theta_2 = 1/2 - \theta_1^o$ ou $\theta_2 = 1/2$ e esses valores (de máximo) são iguais.

Demonstração. Primeiramente, notemos que,

$$h_{\theta_1^o}(\theta_2) = \sum_{j=m}^n \binom{n}{j} \left[(\theta_1^o)^j(1-\theta_1^o)^{n-j} + (\theta_2)^j(1-\theta_2)^{n-j} + (1-\theta_1^o-\theta_2)^j(\theta_1^o+\theta_2)^{n-j} \right] = \sum_{j=m}^n \binom{n}{j} g_{\theta_1^o}^{(j)}(\theta_2)$$

Vamos verificar, na sequência, que $g_{\theta_1^o}^{(j)}(\theta_2)$ é máxima sob o conjunto Θ_0 quando $\theta_2 = 1/2 - \theta_1^o$ ou $\theta_2 = 1/2$, qualquer que seja o valor de $j \in \{m, m+1, \dots, n\}$. Dessa forma, o máximo de

$h_{\theta_1^o}(\theta_2)$ também é atingido nesses pontos. (A igualdade mencionada no Teorema pode ser facilmente verificada substituindo-se os supostos valores de θ_2 que maximizam a função na expressão de $g_{\theta_1^o}^{(j)}(\theta_2)$).

Primeiramente, mostraremos que $g_{\theta_1^o}^{(j)}(\theta_2)$ é decrescente em $\left(\frac{1}{2} - \theta_1^o, \frac{1-\theta_1^o}{2}\right)$ (e, portanto, crescente em $\left(\frac{1-\theta_1^o}{2}, \frac{1}{2}\right)$, de acordo com o Lema A.1.1), para todo $j \in \{m, \dots, n\}$.

Para isso, basta mostrarmos que $g'_{\theta_1^o}(\theta_2) < 0$ em $\left(\frac{1}{2} - \theta_1^o, \frac{1-\theta_1^o}{2}\right)$.

Mas,

$$\begin{aligned} g'_{\theta_1^o}(\theta_2) = j & \left[\theta_2^{j-1}(1-\theta_2)^{n-j-1} - (1-\theta_1^o-\theta_2)^{j-1}(\theta_1^o+\theta_2)^{n-j-1} \right] \\ & + n \left[-\theta_2^j(1-\theta_2)^{n-j-1} + (1-\theta_1^o-\theta_2)^j(\theta_1^o+\theta_2)^{n-j-1} \right] \end{aligned}$$

Assim,

$$g'_{\theta_1^o}(\theta_2) < 0 \iff \frac{n}{j} < \frac{\theta_2^{j-1}(1-\theta_2)^{n-j-1} - (1-\theta_1^o-\theta_2)^{j-1}(\theta_1^o+\theta_2)^{n-j-1}}{\theta_2^j(1-\theta_2)^{n-j-1} - (1-\theta_1^o-\theta_2)^j(\theta_1^o+\theta_2)^{n-j-1}} = \phi(j)$$

Como $n/j < 2$, basta mostrar que $\phi(j) > 2$. Mas $\theta_2 < 1 - \theta_1^o - \theta_2$, então, de fato,

$$\begin{aligned} \phi(j) & > \frac{\theta_2^{j-1}(1-\theta_2)^{n-j-1} - (1-\theta_1^o-\theta_2)^{j-1}(\theta_1^o+\theta_2)^{n-j-1}}{(1-\theta_2-\theta_1^o)\theta_2^{j-1}(1-\theta_2)^{n-j-1} - (1-\theta_2-\theta_1^o)(1-\theta_1^o-\theta_2)^{j-1}(\theta_1^o+\theta_2)^{n-j-1}} \\ & = \frac{1}{1-\theta_1^o-\theta_2} > 2 \end{aligned}$$

□

Corolário A.1.1. Para que $\theta = (\theta_1, \theta_2, \theta_3)$ seja ponto de máximo global da função $f(\theta_1, \theta_2)$ sob o conjunto Θ_0 , é necessário e suficiente que dois dos θ_i 's valham $1/2$ e o outro, conseqüentemente, valha zero.

Demonstração. $\forall (\theta_1, \theta_2) \in \Theta_0$, com $\theta_1 > 0$, temos, pelo Teorema A.1.1, que

$$f(\theta_1, \theta_2) = h_{\theta_1}(\theta_2) \leq h_{\theta_1}\left(\frac{1}{2}\right) = h_{\theta_1}\left(\frac{1}{2} - \theta_1\right).$$

Assim, $\forall (\theta_1, \theta_2) \in \Theta_0$, com $\theta_1 > 0$,

$$f(\theta_1, \theta_2) \leq h_{\theta_1}\left(\frac{1}{2}\right) = f\left(\theta_1, \frac{1}{2}\right) = f\left(\frac{1}{2}, \theta_1\right) = h_{\frac{1}{2}}(\theta_1) \leq h_{\frac{1}{2}}\left(\frac{1}{2}\right) = f\left(\frac{1}{2}, \frac{1}{2}\right).$$

Dessa forma, $\forall (\theta_1, \theta_2) \in \Theta_0$, com $\theta_1 > 0$, $f(\theta_1, \theta_2) \leq f\left(\frac{1}{2}, \frac{1}{2}\right)$.

Como $f\left(0, \frac{1}{2}\right) = f\left(\frac{1}{2}, \frac{1}{2}\right)$, segue que, $\forall (\theta_1, \theta_2) \in \Theta_0$,

$$f(\theta_1, \theta_2) \leq f\left(\frac{1}{2}, \frac{1}{2}\right) = f\left(\frac{1}{2}, 0\right) = f\left(0, \frac{1}{2}\right).$$

Portanto, esses são os pontos de máximo de f .

□

Essa maximização já foi apresentada anteriormente por Shifeng e Guoying [24], utilizando o argumento que a função f é Schur-convexa simétrica.

Corolário A.1.2. *O p -value $p_2(\mathbf{x})$ no ponto $\mathbf{x} \in \mathcal{X}$ tal que $m(\mathbf{x}) = \max(x_1, x_2, x_3) > n/2$ para testar a hipótese $H : \boldsymbol{\theta} \in \Theta_0$ contra $A : \boldsymbol{\theta} \in \Theta_0^c$, com Θ_0 especificado anteriormente, é dado por*

$$p_2(\mathbf{x}) = 2P(\text{Bin}(n, 1/2) \geq m(\mathbf{x})).$$

Demonstração. Pelo Corolário A.1.1, para calcular $p_2(\mathbf{x})$, devemos tomar algum vetor $\boldsymbol{\theta}$ com duas componentes valendo 1/2 e a outra valendo 0. Assim, tomando $\boldsymbol{\theta} = (1/2, 1/2, 0)$, segue que, o p -value $p_2(\mathbf{x})$ é dado por

$$p_2(\mathbf{x}) = \sup_{\boldsymbol{\theta} \in \Theta_0} P\left(\bigcup_{i=1}^3 X_i \geq m(\mathbf{x}) | \boldsymbol{\theta}\right) = \sup_{\boldsymbol{\theta} \in \Theta_0} f(\theta_1, \theta_2)$$

$$= f\left(\frac{1}{2}, \frac{1}{2}, 0\right) = 2 \sum_{j=m(\mathbf{x})}^n \binom{n}{j} \left(\frac{1}{2}\right)^j \left(1 - \frac{1}{2}\right)^{n-j} = 2P(\text{Bin}(n, 1/2) \geq m(\mathbf{x})).$$

□

A.2 RV nos Modelos Multinomial e Multinomial Negativo Truncados

Nessa seção, serão calculadas as razões de verossimilhanças para alguns pontos do espaço amostral sob o modelo multinomial negativo truncado. Analogamente, os cálculos são desenvolvidos para o modelo multinomial truncado.

Considere X com distribuição multinomial truncada (com distribuição de probabilidade descrita pela equação (3.4) da seção 3.2) com parâmetros $n = 5$ e θ e Y com distribuição multinomial negativa truncada (com distribuição de probabilidade descrita pela equação (3.5) da seção 3.2) com parâmetros $x_3 = 1$, $t = 4$ e θ .

Inicialmente, considere o lema a seguir.

Lema A.2.1. *A função $g(\theta) = \frac{\theta}{1-(1-\theta)^5}$ é crescente no intervalo $(0, 1)$.*

Demonstração. Note que $g(0_+) = 1/5$ e $g(1) = 1$. Vamos mostrar que $g'(\theta)$ é positiva $\forall \theta \in (0, 1)$ (e, portanto, g é crescente nesse intervalo).

$$g'(\theta) = \frac{1 - (1-\theta)^5 + 5\theta(1-\theta)^4}{[1 - (1-\theta)^5]^2} = \frac{1 + (1-\theta)^4(6\theta - 1)}{[1 - (1-\theta)^5]^2} \geq 0 \iff h(\theta) = (1-\theta)^4(6\theta - 1) \geq -1$$

Mas $h(0) = -1$, $h(1) = 0$ e $h'(\theta) = 10(1-\theta)^3(1-3\theta) > 0 \iff \theta < 1/3$, com $h(1/3) > 0$, e portanto, $h(\theta) \geq -1$, $\forall \theta \in [0, 1]$ e, consequentemente, g é crescente em $(0, 1)$.

□

Primeiramente vamos calcular a razão de verossimilhanças para o modelo Multinomial Negativo Truncado. Para cada ponto $(a, b) \in \mathcal{Y}'$, a função de verossimilhanças é dada por

$$L_{(a,b)}(\theta) = \frac{\theta_1^a \theta_2^b \theta_3}{1 - (1-\theta_3)^5} \quad , \quad a \geq 0 \quad , \quad b \geq 0 \quad , \quad a + b \leq 4 \quad .$$

- *Caso 1:* $a = b = 0$

Nesse caso, temos pelo Lema A.2.1 que a função de verossimilhanças é crescente em $(0, 1)$ e, portanto,

$$RV(0, 0) = \frac{\sup_{\theta \in \Theta_0} L_{(0,0)}(\theta)}{\sup_{\theta \in \Theta} L_{(0,0)}(\theta)} = \frac{L_{(0,0)}(0, 1/2, 1/2)}{L_{(0,0)}(0, 0, 1)} = \frac{L_{(0,0)}(1/2, 0, 1/2)}{L_{(0,0)}(0, 0, 1)} = \frac{16}{31}$$

- *Caso 2:* $a > 1$ e $b = 0$ (o mesmo argumento vale para $b > 1$ e $a = 0$)

Fixando $\theta_1 \in (0, 1)$ temos, pelo Lema A.2.1, que $\frac{\theta_3}{1-(1-\theta_3)^5}$ é crescente e, portanto, vamos tomar $\theta_3 = 1 - \theta_1$ ($\theta_2 = 0$). Assim, precisamos maximizar a função

$$f(\theta_1) = \frac{\theta_1^a (1 - \theta_1)}{1 - \theta_1^5}.$$

$$f'(\theta_1) = \frac{(a\theta_1^{a-1} - (a+1)\theta_1^a)(1 - \theta_1^5) + 5\theta_1^4\theta_1^a(1 - \theta_1)}{(1 - \theta_1^5)^2} > 0$$

$$\iff 5\theta_1^{a+4} - 5\theta_1^{a+5} + a\theta_1^{a-1} - (a+1)\theta_1^a - a\theta_1^{a+4} + (a+1)\theta_1^{a+5} > 0$$

$$\iff \theta_1^{a-1} \left[(a-4)\theta_1^6 + (5-a)\theta_1^5 - (a+1)\theta_1 + a \right] > 0$$

$$\iff g(\theta_1) = (a-4)\theta_1^6 + (5-a)\theta_1^5 - (a+1)\theta_1 + a > 0,$$

com $g(0) = a$ e $g(1) = 0$. Vamos mostrar que g é estritamente decrescente.

$$g'(\theta_1) = 6(a-4)\theta_1^5 + 5(5-a)\theta_1^4 - (a+1), \quad \text{tal que } g'(0) = -(a+1) \text{ e } g'(1) = 0$$

$$g''(\theta_1) = \theta_1^3 \left[20(5-a) - 30(4-a)\theta_1 \right] > 0 \iff \theta_1 < \frac{2(5-a)}{3(4-a)}, \text{ se } a < 4.$$

Nesse caso, $1 < a < 4$, $\frac{2(5-a)}{3(4-a)} \geq 1$ e, portanto, $g''(\theta_1) > 0$ e $g'(\theta_1)$ é crescente (se $a = 4$, $g''(\theta_1) = 20(5-a)\theta_1^3 > 0$ e a conclusão é a mesma). Logo, $g'(\theta_1) < 0$ e, portanto, $g(\theta_1)$ é decrescente, donde $g(\theta_1) > 0$, $\forall \theta_1 \in (0, 1)$. Assim, $f'(\theta_1) > 0$ e f é crescente em $(0, 1)$.

$$\begin{aligned} &\Rightarrow \sup_{\theta \in \Theta} L_{(a,0)}(\boldsymbol{\theta}) \quad \text{se dá no ponto } (1, 0, 0), \quad a \in \{2, 3, 4\} \\ &\Rightarrow \sup_{\theta \in \Theta_0} L_{(a,0)}(\boldsymbol{\theta}) \quad \text{se dá no ponto } \left(\frac{1}{2}, 0, \frac{1}{2}\right), \quad a \in \{2, 3, 4\} \end{aligned}$$

Dessa forma, temos

$$RV(a, 0) = RV(0, a) = \frac{\sup_{\theta \in \Theta_0} L_{(a,0)}(\boldsymbol{\theta})}{\sup_{\theta \in \Theta} L_{(a,0)}(\boldsymbol{\theta})} = \frac{\frac{(\frac{1}{2})^{a+1}}{1 - (\frac{1}{2})^5}}{\frac{1}{5}} = \frac{5}{31} 2^{4-a} \quad (\text{A.1})$$

- *Caso 3:* $2 \leq a + b \leq 4$ (com $a \neq 0$ e $b \neq 0$)

Considere $\theta_3 \in [0, 1]$ fixado e $g_{\theta_3}(\theta_1) = \theta_1^a (1 - \theta_1 - \theta_3)^b$.

$$\begin{aligned} g'_{\theta_3}(\theta_1) &= a\theta_1^{a-1} (1 - \theta_1 - \theta_3)^b - b\theta_1^a (1 - \theta_1 - \theta_3)^{b-1} \\ &= \theta_1^{a-1} (1 - \theta_1 - \theta_3)^{b-1} \left[a(1 - \theta_3) - (a + b)\theta_1 \right] \\ g'_{\theta_3}(\theta_1) > 0 &\iff (a + b)\theta_1 < a(1 - \theta_3) \iff \theta_1 < \frac{a}{a + b} (1 - \theta_3) \end{aligned}$$

Dessa forma, $g_{\theta_3}(\theta_1)$ é máxima em $\frac{a}{a+b} (1 - \theta_3)$, para cada $\theta_3 \in [0, 1]$ fixado, e, portanto, basta maximizar em θ_3 a seguinte função

$$\frac{\left(\frac{a}{a+b} (1 - \theta_3)\right)^a \left(\frac{b}{a+b} (1 - \theta_3)\right)^b \theta_3}{1 - (1 - \theta_3)^5} \propto \frac{\theta_3 (1 - \theta_3)^{a+b}}{1 - (1 - \theta_3)^5} = f(\theta_3),$$

com $1 < a + b < 5$. Temos que $f(0_-) = 1/5$ e $f(1) = 0$.

$$f'(\theta_3) = \frac{\left[1 - (1 - \theta_3)^5\right] \left[(1 - \theta_3)^{a+b} - (a + b)\theta_3 (1 - \theta_3)^{a+b-1}\right] - 5\theta_3 (1 - \theta_3)^{a+b} (1 - \theta_3)^4}{\left[1 - (1 - \theta_3)^5\right]^2}$$

$$= \frac{(1 - \theta_3)^{a+b-1}}{\left[1 - (1 - \theta_3)^5\right]^2} \left\{ [1 - \theta_3 (a + b + 1)] - (1 - \theta_3)^5 [1 - \theta_3 (a + b - 4)] \right\}$$

Para estudar o comportamento da função f , vamos estudar o sinal de sua derivada ou, equivalentemente, estudar a função

$$u(\theta_3) = 1 - \theta_3 (a + b + 1) - (1 - \theta_3)^5 [1 - \theta_3 (a + b - 4)],$$

onde $u(0) = 0$, $u(1) = -(a + b)$ e

$$u'(\theta_3) = -(a + b + 1) + (1 - \theta_3)^4 [(a + b + 1) - 6\theta_3 (a + b - 4)].$$

Assim, considere também

$$v(\theta_3) = (1 - \theta_3)^4 [(a + b + 1) - 6\theta_3 (a + b - 4)],$$

de modo que $v(0) = a + b + 1$, $v(1) = 0$ e

$$v'(\theta_3) = (1 - \theta_3)^3 [-4(a + b + 1) - 6(a + b - 4) + 30(a + b - 4)\theta_3].$$

Temos, para $a + b = 4$, que $v'(\theta_3) = -4(1 - \theta_3)^3 (a + b + 1) < 0$. Para $a + b < 4$, temos que

$$v'(\theta_3) < 0 \iff 30(a + b - 4)\theta_3 < 4(a + b + 1) + 6(a + b - 4) \iff \theta_3 > \frac{2 - a - b}{3(4 - a - b)}$$

Como $a + b \geq 2$, segue que $v'(\theta_3) < 0$, $\forall \theta_3 > 0$.

Assim, temos que,

$$a + b \geq 2 \Rightarrow v'(\theta_3) \leq 0 \Rightarrow v \text{ é decrescente} \Rightarrow v(\theta_3) \leq a + b + 1 \Rightarrow u'(\theta_3) \leq 0$$

$$\Rightarrow u \text{ é decrescente} \Rightarrow u(\theta_3) \leq 0 \Rightarrow f'(\theta_3) \leq 0 \Rightarrow f \text{ é decrescente}, \forall \theta_3 \in (0, 1)$$

$$\begin{aligned} &\Rightarrow \sup_{\theta \in \Theta} L_{(a,b)}(\theta) \quad \text{se dá no ponto } \left(\frac{a}{a+b}, \frac{b}{a+b}, 0 \right), \quad 2 \leq a+b \leq 4 \\ &\Rightarrow \sup_{\theta \in \Theta_0} L_{(a,b)}(\theta) \quad \text{se dá no ponto } \left(\frac{1}{2}, \frac{1}{2}, 0 \right), \quad 2 \leq a+b \leq 4 \end{aligned}$$

Dessa forma, temos

$$RV(a,b) = RV(b,a) = \frac{\sup_{\theta \in \Theta_0} L_{(a,b)}(\theta)}{\sup_{\theta \in \Theta} L_{(a,b)}(\theta)} = \frac{\frac{\left(\frac{1}{2}\right)^{a+b}}{5}}{\frac{\left(\frac{a}{a+b}\right)^a \left(\frac{b}{a+b}\right)^b}{5}} = \left(\frac{a+b}{2} \right)^{a+b} \frac{1}{a^a b^b} \quad (\text{A.2})$$

- *Caso 4*: $a+b=1$. Nesse caso, os cálculos podem ser feitos de forma análoga ao *Caso 3*, porém é necessário achar as raízes de um polinômio de grau 4. Contudo, não é difícil mostrar que $RV(1,0) = RV(0,1) \geq 8/31 > RV(4,0)$.

Vamos apresentar os cálculos da razão de verossimilhanças (RV) para alguns pontos no modelo multinomial truncado. Note que, para os pontos $(a,b,1)$ com $a+b=4$, temos que a RV é idêntica à RV dos pontos da forma (a,b) com $a+b=4$ no modelo multinomial negativo truncado. Vamos então calcular a RV para o ponto $(0,0,5)$.

$$\begin{aligned} L_{(0,0,5)}(\theta) &= \frac{\theta_3^5}{1 - (1 - \theta_3)^5} = f(\theta_3) \\ f'(\theta_3) &= \frac{5\theta_3^4 [1 - (1 - \theta_3)^5] - 5\theta_3^5 (1 - \theta_3)^4}{[1 - (1 - \theta_3)^5]^2} = \frac{5\theta_3^4}{[1 - (1 - \theta_3)^5]^2} \left\{ 1 - (1 - \theta_3)^4 \right\} > 0 \\ &\Rightarrow f \text{ é crescente, } \forall \theta_3 \in (0,1) \end{aligned}$$

$$\begin{aligned} &\Rightarrow \sup_{\theta \in \Theta} L_{(0,0,5)}(\boldsymbol{\theta}) \quad \text{se dá no ponto } (0, 0, 1) \\ &\Rightarrow \sup_{\theta \in \Theta_0} L_{(0,0,5)}(\boldsymbol{\theta}) \quad \text{se dá no ponto } \left(\frac{1}{2}, 0, \frac{1}{2}\right) \end{aligned}$$

Dessa forma, temos

$$RV(0, 0, 5) = \frac{\sup_{\theta \in \Theta_0} L_{(0,0,5)}(\boldsymbol{\theta})}{\sup_{\theta \in \Theta} L_{(0,0,5)}(\boldsymbol{\theta})} = \frac{\left(\frac{1}{2}\right)^5}{1 - \left(\frac{1}{2}\right)^5} = \frac{1}{31} \quad (\text{A.3})$$

As razões de verossimilhanças para os demais pontos no modelo multinomial truncado podem ser obtidas utilizando a rotina computacional apresentada no Apêndice [B.2](#).

Apêndice B

Rotinas Computacionais

B.1 *p-value* da Multinomial Negativa

Rotina em R [18] para o cálculo de $P(\mathbf{X} \in C|\boldsymbol{\theta})$ via método de Monte Carlo, onde $\mathbf{X} = (X_1, X_2, X_3) \sim MultNeg(x_3, \boldsymbol{\theta})$ e $C = \{\mathbf{y} \in \mathcal{X} : RV(\mathbf{y}) \leq RV(\mathbf{x})\}$.

```
# função pvalue: calcula o p-value para a Mult. Negativa
# p/ o valor obs. x=(x1,x2,x3) e o parâmetro teta=(teta3,teta1)
pvalue = function(x,teta) {

  # função logrv: calcula o log da razão de verossimilhanças para um dado x
  logrv = function(t) {
    m = max(t)
    b = sum(sort(t)[1:(length(t)-1)])
    if(log(m) <= log(b)) { 0 }
    else {
      if(b == 0) { m*log(0.5) }
      else {
        m*(log(1+exp(log(b)-log(m)))-log(2)) + b*(log(1+exp(log(m)-log(b)))-log(2))
      }
    }
  }
}
```

```
# lrv: log da razão de verossimilhanças para o valor x observado
lrv = logrv(x)

# n: quantidade de valores a serem simulados
n = 100000

# simulação da distribuição Mult.Negativa(x[3], teta)
s3 = rnbinom(n, x[3], teta[1]/(1-teta[2]))
s2 = rnbinom(n, x[3]+s3, 1-teta[2])
s = cbind(x[3],s2,s3)

# calcula o pvalue via Monte Carlo
c = apply(s,1,function(y) { if(logrv(y)<=lrv) {1} else {0} })
sum(c)/length(c)
}
```

Exemplo de utilização:

```
x = c(10,1,2)
teta = c(2.22e-16,0.5)
pvalue(x,teta)
```


B.2 RV na Multinomial Truncada

Rotina em R [18] para o cálculo de $RV(\mathbf{a})$, para $\mathbf{a} = (a_1, a_2, a_3) \in \mathcal{X}'$ (ou $\mathbf{a} \in \mathcal{Y}'$, com $a_3 = 1$) para o modelo Multinomial (ou Multinomial Negativo Truncado), definidos na página 31 da Seção 3.2.

```
RV <- function(a){
  if(a[1]+a[2]>4 | sum(a)>5) { cat("ERRO! A soma das coordenadas deve ser menor
    ou igual que 5 \n      e a soma das duas primeiras menor ou igual que 4! \n\n")}

  else{

    if(a[2]>a[1]) {cat("ERRO! Para esse caso, use o ponto (",a[2],a[1],a[3],
      ").\n\n")}

    else {

      f <- function(x){((x[1]^a[1])*((1-x[1]-x[2])^a[2])*(x[2]^a[3])) /
        (1-((1-x[2])^5))}

      # restrição natural: 0 <= TETAi <= 1, i=1,2,3 , e soma(TETAi)=1
      u1 <- c(-1,1,0,-1,0,1)
      dim(u1) <- c(3,2)
      c1 <- c(-1,0,0)
      dim(c1) <- c(3,1)
      # máximo da verossimilhança
      t1 <- constrOptim(c(1/3,1/3), f, grad=NULL, ui=u1, ci=c1,
        control=list(fnscale=-1))

      # restrição de H: 0 <= TETAi <= 1/2, i=1,2,3 , e soma(TETAi)=1
      u0 <- c(1,-1,0,0,0,1,-1)
      dim(u0) <- c(4,2)
      c0 <- c(0,-1/2,0,-1/2)
```

```
dim(c0) <- c(4,1)
# máximo da verossimilhança sob H
t0 <- constrOptim(c(1/3,1/3), f, grad=NULL, ui=u0, ci=c0,
  control=list(fnscale=-1))

# RV(a[1],a[2],a[3])
t0$value/t1$value
}
}
}
```

Exemplo de utilização:

```
a = c(4,0,1)
RV(a)
```

Referências Bibliográficas

- [1] V. Barnett. *Comparative Statistical Inference*. John Wiley & Sons, Inc., Chichester, segunda edição, 1982. 1, 15, 37, 40
- [2] J. O. Berger e M. Delampady. Testing precise hypotheses (with discussion). *Statistical Science*, 2(3):317–335, 1987. 39
- [3] J. O. Berger e R. L. Wolpert. *The Likelihood Principle*, volume 6 of *Monograph Series*. IMS Lecture Notes, Hayward, segunda edição, 1988. 2, 15, 37, 40
- [4] G. Casella e R. L. Berger. *Statistical Inference*. Duxbury, Pacific Grove, segunda edição, 2002. 9, 39
- [5] M. H. DeGroot. *Optimal Statistical Decisions*. MacGraw-Hill, New York, 1970. 2
- [6] M. H. DeGroot. *Probability and Statistics*. Addison-Wesley Publishing Company, Massachusetts, 1975. 2, 39
- [7] A. Dempster. The direct use of likelihood for significance testing. *Statistics and Computing*, 7(4):247–252, 1997. 39
- [8] Ibope. *Pesquisa eleitoral do IBOPE Opinião*. Instituto Brasileiro de Opinião Pública e Estatística. URL http://www.ibope.com.br/calandraWeb/BDarquivos/sobre_pesquisas/pesquisa_eleitoral.html. 2, 18
- [9] L. Y. T. Inoue. *Desenvolvimento e Implicações do Princípio da Verossimilhança*. Dissertação de Mestrado. São Paulo, 1995. 2, 15, 40
- [10] J. E. Mosimann. On the compound negative multinomial distribution and correlations among inversely sampled pollen counts. *Biometrika*, 50(1/2):47–54, 1963. 2, 19
- [11] F. Newport, L. Saad e D. Moore. *How Polls are Conducted?* Gallup, Inc. URL <http://media.gallup.com/PDF/FAQ/HowArePolls.pdf>. 18
- [12] J. M. Nicolau. *Sistema Eleitoral e Reforma Política*. Foglio, Rio de Janeiro, 1993. 1

- [13] I. Olkin e M. Sobel. Integral expressions for tail probabilities of the multinomial and negative multinomial distributions. *Biometrika*, 52(1/2):167–179, 1965. 2
- [14] C. D. Paulino, M. A. Turkman e B. Murteira. *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa, 2003. 6, 12
- [15] C. A. B. Pereira e J. M. Stern. *Inferência Indutiva com Dados Discretos: Uma Visão Genuinamente Bayesiana*. XV Congreso de Matemática Capricórnio, Antofagasta, Chile, 2005. 2, 19
- [16] C. A. B. Pereira e J. M. Stern. Evidence and credibility: Full bayesian significance test for precise hypotheses. *Entropy*, I(4):99–110, 1999. 40
- [17] C. A. B. Pereira e M. A. G. Viana. *Elementos de Inferência Bayesiana*. 5º SINAPE, 1982. 2, 19
- [18] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. URL <http://www.R-project.org>. ISBN 3-900051-07-0. 29, 31, 53, 55
- [19] H. Raiffa e R. Schlaifer. *Applied Statistical Decision Theory*. Harvard University, Boston, 1961. 12, 13, 39
- [20] S. M. Ross. *Stochastic Processes*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, segunda edição, 1996. 11
- [21] G. G. Roussas. *A First Course in Mathematical Statistics*. Addison-Wesley Publishing Company, 1973. 1, 9
- [22] R. Royall. *Statistical Evidence: a Likelihood Paradigm*, volume 71 of *Monograph on Statistics and Applied Probability*. Capman & Hall, London, UK, 1997. 2, 27
- [23] A. Sen. Social choice theory: a re-examination. *Econometrica*, 45(1):53–89, 1977. 2
- [24] X. Shifeng e L. Guoying. Testing for the maximum cell probabilities in multinomial distributions. *Science in China Series A: Mathematics*, 48(7):972–985, 2005. 44
- [25] J. M. Stern. *Frontiers in Artificial Intelligence and its Applications*, volume 101, capítulo Significance tests, Belief Calculi, and Burden of Proof in legal and Scientific Discourse, 139–147. IOS Press, 2003. 40

Índice Remissivo

p-value, [10](#)

Distribuição

Dirichlet, [35](#)

multinomial, [19](#)

multinomial negativa, [20](#)

multinomial negativa truncada, [32](#)

multinomial truncada, [32](#)

Erros tipo I e tipo II, [7](#)

Exemplos, [17](#)

abordagem bayesiana, [35](#)

abordagem clássica, [21](#)

Função de perda, [13](#)

Função poder, [8](#)

Teste

da razão de verossimilhanças, [9](#)

de hipóteses, [7](#)

função poder, *veja* Função poder

tipos de erros, *veja* Erros tipo I e tipo II

uniformemente mais poderoso, [8](#)

Testes de hipóteses bayesianos, [12](#)

Testes de hipóteses clássicos, [7](#)