

# **Desempate técnico**

Filipe Jaeger Zabala

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

Programa: Estatística  
Orientador: Prof. Dr. Sérgio Wechsler

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES e CNPq.

São Paulo, 03 de março de 2010.

## **Desempate técnico**

Este exemplar corresponde à redação  
da dissertação defendida por  
Filipe Jaeger Zabala.

Banca Examinadora:

- Prof. Dr. Sérgio Wechsler - IME-USP.
- Prof. Dr. Carlos Alberto de Bragança Pereira - IME-USP.
- Prof. Dr. Hélio dos Santos Migon - IM-UFRJ.

## Agradecimentos

À minha mãe Lia Helena e à minha avó Beatriz, por todo amor, carinho, dedicação e paciência. Ao meu pai Ricardo, pelas lições de humildade. Às tias Maria Inês, Sandra e Denise por serem meio tias e meio mães. Aos meus avós Nara e Ronê, pelo carinho, apoio e confiança. À minha tia Susie pelos conselhos e longas conversas. À priminha Júlia, por ser tão amada e talentosa. Aos meus primos Matheus, Vinícius (Fino), Pedro e Guilherme, por serem os irmãos que eu não tive. Às minhas primas Fernanda e Paula, por serem tão roqueiras. Ao meu primo e companheiro musical Leonardo, pela confiança depositada em nosso projeto ao longo desses 10 anos. Ao grande amigo e guitarrista Victor Nichele, por acreditar em nosso projeto e investir nele. À Letícia, pelo companheirismo e carinho. À minha família de São Paulo, os Pontes Batista, em especial ao seu Antônio, dona Raimunda, Luana e Júnior. Aos meus amigos Fabiano Feijó, Fernando Sola e Diego Torgão, pelas risadas ao longo de todos esses anos. Ao meu orientador, Sérgio Wechsler, pelos ensinamentos e leveza para realizar este trabalho. Ao meu inestimável amigo Eduardo Yoshio Nakano, que contribuiu muito para a elaboração da conjectura apresentada neste trabalho, posteriormente provada pelo brilhante Luis Gustavo Esteves. Ao sábio professor Carlinhos Pereira, pelo incentivo e contribuição ao nosso trabalho. Ao professor Hélio Migon, por aceitar participar da banca e fazer uso de alguns resultados aqui obtidos.



## Resumo

Neste trabalho faz-se uma comparação entre as inferências bayesiana e frequentista para a estimação de proporções em um processo eleitoral brasileiro. Foi feito um contraponto entre as probabilidades a posteriori bayesianas e os intervalos/regiões de confiança frequentistas. Mostrou-se que os intervalos de confiança apresentados pelos institutos de pesquisa brasileiros não possuem a confiança que afirmam ter, levando o autor a desenvolver uma metodologia alternativa, denominada neste trabalho como *frequentista correta*. De posse das abordagens bayesiana, frequentista correta e dos institutos, encontrou-se um teorema – em um cenário com dois candidatos – que relaciona intervalos de confiança frequentistas com probabilidades a posteriori bayesianas. Para três candidatos são apresentados exemplos que contradizem a noção de empate técnico definida pelos institutos, largamente veiculada nos meios de comunicação.

**Palavras-chave:** inferência bayesiana, eleições, proporção, pesquisa eleitoral.



## **Abstract**

In this paper was made comparisons between the bayesian and frequentist inferences for the estimation of the proportions in a brazilian electoral process. It was made a comparison between the bayesian posterior probabilities and frequentist confidence intervals/regions. It was shown that the confidence intervals submitted by brazilian research institutes lack the confidence that they claim to have, taking the author to develop an alternative approach, called in this work correct frequentist. With bayesian, correct frequentist and institutes approaches, we derived a theorem that relates – in a cenario with two candidates – frequentist confidence intervals with bayesian posterior probabilities. For three candidates are given examples that contradict the “tie” notion defined by the brazilian research institutes, widely disseminated in mass media.

**Keywords:** bayesian inference, electoral competition, proportion, electoral survey.





# Sumário

<b>Lista de Abreviaturas</b>	<b>xi</b>
<b>Lista de Símbolos</b>	<b>xiii</b>
<b>Lista de Figuras</b>	<b>xv</b>
<b>Lista de Tabelas</b>	<b>xvii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Considerações Preliminares . . . . .	1
1.2 Objetivos . . . . .	1
1.3 Organização do Trabalho . . . . .	1
<b>2 Inferência Bayesiana</b>	<b>3</b>
2.1 Fórmula de Bayes . . . . .	3
2.2 Permutabilidade . . . . .	4
2.3 Princípio da verossimilhança . . . . .	5
2.4 Estatísticas suficientes . . . . .	7
2.5 Informação <i>a priori</i> . . . . .	9
2.5.1 Distribuições <i>a priori</i> conjugadas . . . . .	9
2.5.2 Distribuições <i>a priori</i> não-informativas . . . . .	11
2.6 Informação <i>a posteriori</i> . . . . .	13
2.6.1 Resumindo a inferência <i>a posteriori</i> . . . . .	13

2.7	Métodos de Monte Carlo . . . . .	14
2.7.1	Monte Carlo Ordinário . . . . .	14
<b>3</b>	<b>Inferência para Proporção</b>	<b>17</b>
3.1	Independência . . . . .	17
3.2	Institutos . . . . .	18
3.2.1	Estimação por ponto . . . . .	18
3.2.2	Estimação por intervalo de confiança . . . . .	19
3.2.3	Dois Candidatos . . . . .	19
3.2.4	Três Candidatos . . . . .	21
3.2.5	$k$ Candidatos . . . . .	22
3.3	Inferência Frequentista Correta . . . . .	23
3.3.1	Estimação por ponto . . . . .	23
3.3.2	Estimação por intervalo e regiões de confiança . . . . .	23
3.3.3	Dois Candidatos . . . . .	23
3.3.4	$k$ Candidatos . . . . .	27
3.4	Bayesiana . . . . .	29
3.4.1	$k$ Candidatos . . . . .	29
<b>4</b>	<b>Comparações</b>	<b>31</b>
4.1	Dois Candidatos . . . . .	31
4.2	Três Candidatos . . . . .	38
<b>5</b>	<b>Conclusões</b>	<b>43</b>
5.1	Considerações Finais . . . . .	43
5.2	Sugestões para Pesquisas Futuras . . . . .	43
<b>A</b>	<b>Artigos 33, 34 e 35 da Lei N°. 9504/97</b>	<b>45</b>

*SUMÁRIO*

ix

**Referências Bibliográficas**

**47**



## Lista de Abreviaturas

va	Variável aleatória ( <u>Random variable</u> ).
vas	Variáveis aleatórias ( <u>Random variables</u> ).
iid	Independente identicamente distribuída ( <u>Independent identically distributed</u> ).
fdp	Função densidade de probabilidade ( <u>Probability density function</u> ).
fp	Função de probabilidade ( <u>Probability function</u> ).
Eq.	Equação ( <u>Equation</u> ).
pp	Probabilidade <i>aposteriori</i> (Posterior probability).



## Lista de Símbolos

$\Omega$	Espaço fundamental.
$\mathcal{A}$	Família (álgebra ou $\sigma$ -álgebra) dos acontecimentos dotados de probabilidade.
$P$	Medida de probabilidade.
$\doteq$	Aproximadamente igual.
$\propto$	Proporcional a.
$\sim$	Tem distribuição.
$\dot{\sim}$	Tem distribuição aproximadamente.
$h(\theta)$	Distribuição <i>a priori</i> de $\theta$ .
$h(\theta x)$	Distribuição <i>a posteriori</i> de $\theta$ dado $x$ .
$L(\theta x) = f(x \theta)$	(Função de) Verossimilhança.
$n$	Tamanho da amostra.
$\alpha$	Nível de significância.
$z$	Quantil da normal padrão, equivalente a $z_{1-\frac{\alpha}{2}}$ .
$p_{A_i}$	Proporção de votos para o candidato $A_i$ na população.
$\hat{p}_{A_i}$	Proporção de votos para o candidato $A_i$ na amostra.
.	Separador decimal.
,	Separador de milhar.
☺	Fim de exemplo.
□	Fim de demonstração.





## Lista de Figuras

3.1	Sobreposição dos ICs para a proporção de votos para os candidatos $A_1$ e $A_2$ . . . . .	21
3.2	Sobreposição dos ICs para a proporção de votos para os candidatos $A_1$ , $A_2$ e $A_3$ . . . . .	22
3.3	IC 95% para a diferença de proporção de votos para os candidatos $A_1$ e $A_2$ , com a correção $\frac{N-n}{N-1}$ . 27	
3.4	IC 95% para a diferença de proporção de votos para os candidatos $A_1$ e $A_2$ , sem a correção $\frac{N-n}{N-1}$ . 27	
3.5	Simplex gerado pela restrição $p_{A_1} + p_{A_2} + p_{A_3} = 1$ , com regiões representando os possíveis eventos em um cenário eleitoral com três candidatos. . . . .	29
4.1	Proporções $\hat{p}_{A_1}$ vs probabilidades <i>a posteriori</i> para $\alpha = 0.05$ e $n = 50, 100, 500, 1000$ . . . . .	33
4.2	Proporções $\hat{p}_{A_1}$ vs probabilidades <i>a posteriori</i> para $n = 100, 000$ e $\alpha = 0.01, 0.05, 0.10, 0.20$ . . . . .	34
4.3	Elipses de confiança 95% com $n = 50$ e $n = 100$ . . . . .	39
4.4	Elipses de confiança 95% com $n = 500$ e $n = 1000$ . . . . .	40



## Lista de Tabelas

4.1	Possibilidades em um cenário com três candidatos e as probabilidades <i>a posteriori</i> de cada resultado. . . . .	41
-----	---	----



# Capítulo 1

## Introdução

### 1.1 Considerações Preliminares

A cada dois anos o povo brasileiro participa das eleições, um direito conquistado após anos de repressão política e militar e que asseguraram um Estado democrático, ao menos em países onde o povo tem informação e expressa sua opinião através do voto. Devido às incertezas que circundam o resultado eleitoral, faz-se quase que obrigatória a presença de avaliações a respeito dessas incertezas. A isto damos o nome de *pesquisa eleitoral*, que pode ser dividida em dois grandes momentos: o delineamento amostral e a análise dos dados obtidos através deste delineamento. O foco deste trabalho está na segunda parte, onde assume-se que a amostra foi “bem selecionada”. O autor não entrará no mérito da qualidade da seleção visto que este tópico renderia outra dissertação. Para maiores detalhes a respeito de amostragem, sugere-se [Bolfarine and Bussab \[2005\]](#) e [Cochran \[1977\]](#).

### 1.2 Objetivos

O objetivo deste trabalho é ressaltar e analisar as diferenças entre as abordagens frequentista e bayesiana para a estimação de proporções em um cenário eleitoral brasileiro. Para isto foram estruturadas metodologias apresentadas ao longo dos Capítulos [2](#) e [3](#), que embasam a discussão apresentada no Capítulo [4](#).

### 1.3 Organização do Trabalho

No Capítulo [2](#), são apresentados os conceitos fundamentais de Inferência Bayesiana. No Capítulo [3](#) faz-se a estruturação e comparação de três metodologias de inferência para proporção: i) institutos, ii) frequentista correta e iii) bayesiana. No Capítulo [4](#) são feitas comparações entre as três metodologias descritas no Capítulo [3](#). Finalmente, no Capítulo [5](#) discutimos algumas conclusões obtidas neste trabalho. Analisamos as vantagens e desvantagens dos métodos propostos, fazendo um apanhado geral do que foi discutido nos capítulos anteriores, sugerindo alguns tópicos para pesquisas futuras.



## Capítulo 2

### Inferência Bayesiana

Inferência estatística é o importante ramo da Estatística que tem por objetivo fazer afirmações sobre quantidades desconhecidas (parâmetros). Tal tipo de afirmação deve sempre vir acompanhada de uma medida de precisão sobre sua veracidade. Para realizar este trabalho o estatístico coleta informações de dois tipos: experimentais (as amostras) e culturais (aquelas que obtém através da literatura e do conhecimento tácito). As duas principais escolas de inferência são a Inferência Bayesiana e a Inferência Frequentista (ou Clássica).

*Inferência Bayesiana* é um tipo de inferência estatística que descreve as incertezas sobre quantidades desconhecidas de forma probabilística. Incertezas são modificadas periodicamente após observações de novos dados ou resultados. A operação que calibra a medida das incertezas é conhecida como operação bayesiana e é baseada na Fórmula de Bayes.

#### 2.1 Fórmula de Bayes

A *Fórmula de Bayes* é um mecanismo formal utilizado para combinar a opinião do pesquisador *a priori* com a informação dos dados, obtida através da (função de) verossimilhança (Paulino et al. [2003]). Podemos pensar intuitivamente na Fórmula de Bayes como uma ferramenta que nos informa a probabilidade da causa dada a consequência. Muitas vezes as pessoas tratam erroneamente a Inferência Bayesiana como a simples aplicação desta fórmula.

Seja o espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ , onde

- $\Omega$  é um espaço fundamental não vazio (que nas aplicações coincide, em geral, com o espaço amostral) com elementos  $\omega$ ,  $\omega \in \Omega$ , chamados eventos elementares e subconjuntos  $A$ ,  $A \subseteq \Omega$ , designados eventos;
- $\mathcal{A}$  é a família (álgebra ou  $\sigma$ -álgebra) dos eventos dotados de probabilidade;
- $P$  é a medida de probabilidade definida para os eventos  $A \subseteq \Omega$ ,  $A \subseteq \mathcal{A}$ , em que  $P(A)$  é a probabilidade do evento  $A$ .

Seja uma partição finita (ou infinita) de  $\Omega$ ,

$$A_1, A_2, \dots, A_m, P(A_i) > 0, A_i \cap A_j = \emptyset, i \neq j, \cup_i A_i = \Omega.$$

Dado outro evento  $B$  qualquer, com  $P(B) > 0$ , é fácil verificar a decomposição de  $B$  na união de conjuntos disjuntos

$$B = \cup_i (A_i \cap B).$$

Consequentemente à aditividade da função  $P$  e à definição de probabilidade condicional, temos

$$P(B) = \sum_i P(A_i \cap B) = \sum_i P(B|A_i)P(A_i).$$

Finalmente, notando

$$P(A_i \cap B) = P(B|A_i)P(A_i) = P(A_i|B)P(B)$$

e resolvendo em relação a  $P(A_i|B)$ , chega-se à Fórmula de Bayes

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)}. \quad (2.1)$$

## 2.2 Permutabilidade

Introduzido por [de Finetti \[1974\]](#), o conceito de *permutabilidade* (*exchangeability*) é de grande importância na Inferência Bayesiana. Esta ideia surgiu quando de Finetti percebeu uma incongruência em afirmar que eventos  $(A_1, A_2, \dots, A_n)$  observados em sequência – em particular variáveis aleatórias  $(X_1, X_2, \dots, X_n)$  – são independentes com probabilidade constante mas desconhecida ([Paulino et al. \[2003\]](#)).

Formalmente, dizemos que um conjunto finito de variáveis aleatórias  $(X_1, X_2, \dots, X_n)$  é permutável quando

$$f(x_1, x_2, \dots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}), \quad (2.2)$$

qualquer que seja a permutação de  $\{\pi(1), \pi(2), \dots, \pi(n)\}$  do conjunto  $\{1, 2, \dots, n\}$ . Um conjunto infinito de variáveis aleatórias  $X_i, i = 1, 2, \dots$  é dito permutável se qualquer subconjunto for permutável no sentido de [\(2.2\)](#).



Finetti mostrou que todas as sequências binárias permutáveis são misturas de sequências de Bernoulli. Hewitt and Savage [1955] generalizaram o resultado de Finetti para qualquer sequência de variáveis.

Intuitivamente, permutabilidade poderia ser considerada da seguinte maneira: se uma pessoa observa uma sequência de experimentos aleatórios mas não acredita que a ordem destes experimentos é relevante para a obtenção de informação sobre os parâmetros, então ela considera esta sequência permutável. A suposição de permutabilidade é mais fraca que a suposição de independência, flexibilizando a solução de problemas.

A principal consequência da permutabilidade é o Teorema da Representação de Finetti, de interpretação bastante difícil e sucintamente apresentado a seguir. Se  $X_i$ ,  $i = 1, 2, \dots$  é uma sucessão infinita de variáveis aleatórias binárias permutáveis, então para qualquer inteiro  $n$  existe uma função de distribuição  $H$  tal que

$$f(x_1, x_2, \dots, x_n) = \int_0^1 \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} dH(\theta), \quad (2.3)$$

onde

$$H(\theta) = \lim_{n \rightarrow \infty} P\left(\frac{S_n}{n} \leq \theta\right), \quad (2.4)$$

com

$$S_n = X_1 + X_2 + \dots + X_n, \quad \theta = \lim_{n \rightarrow \infty} \frac{S_n}{n}. \quad (2.5)$$

## 2.3 Princípio da verossimilhança

Usando a fórmula de Bayes (2.1) com um modelo de probabilidade escolhido, significa que o dado  $x$  afeta a inferência a posteriori somente através da função  $f(x|\theta)$ , que quando considerada como função de  $\theta$  para um  $x$  fixado é chamada (*função de*) *verossimilhança*. Desta maneira a Inferência Bayesiana obedece ao chamado *princípio da verossimilhança*, que declara que para uma certa informação amostral, quaisquer dois modelos de probabilidade  $f(x|\theta)$  que tenham a mesma verossimilhança fornecem a mesma inferência para  $\theta$  (Gelman et al. [2004]). Segundo Bernardo and Smith [1994], este princípio é uma consequência direta da fórmula de Bayes, e não algo que seja imposto ou que os bayesianos desejem obedecer (Paulino et al. [2003]).

**Exemplo 1.** (adaptado de Lindley and Phillips [1976]): Considere uma sucessão de lançamentos de uma moeda, independentes e condicionados por  $\theta$ , que designa a probabilidade de sair “cara”. Suponha que obtenha-se o seguinte resultado ou amostra,

$$x = \{C, \bar{C}, C, C, \bar{C}, \bar{C}, C, \bar{C}, \bar{C}, \bar{C}\},$$

onde  $C$  designa “cara” e  $\bar{C}$  “coroa”. Este resultado poderia ser obtido de diversos processos experimentais

ou regras de parada, como

- lançar a moeda com o número de lançamentos – no caso 10 – fixado anteriormente;
- lançar a moeda até aparecerem 6 coroas;
- lançar a moeda até aparecerem 3 coroas consecutivas;
- lançar a moeda até o jogador ficar saturado, tendo a saturação ocorrido no 10º lançamento.

Em qualquer dos quatro casos a função de verossimilhança é proporcional a  $\theta^4(1 - \theta)^6$ .

Assim, adotando-se o princípio da verossimilhança, toda a informação que  $x$  pode dar sobre  $\theta$  encontra-se nesta expressão. Saber qual dos quatro processos experimentais foi utilizado (cada um com seu particular espaço amostral) ou saber qual foi a regra de parada adotada nada tem a acrescentar. Note que a possibilidade de o experimentador parar, por seu arbítrio, ao considerar o resultado  $x$  satisfatório, não altera em nada o que foi dito.

😊

**Exemplo 2.** (adaptado de [Lindley and Phillips \[1976\]](#)): Suponha agora que queiramos testar a hipótese  $H_0 : \theta = 1/2$  contra  $H_1 : \theta > 1/2$ . São contemplados dois processos experimentais:

- $E_1$ : lançar a moeda 12 vezes;
- $E_2$ : lançar a moeda até que apareçam 3 caras;

Admita que o resultado observado nas duas experiências foi  $x = 9$  coroas (portanto  $12 - 9 = 3$  caras), que é uma particular realização da variável aleatória  $X$ , que designa o número total de coroas dos experimentos  $E_1$  e  $E_2$ . Para um frequentista o nível crítico (ou valor- $p$  ou nível descritivo amostral, ou seja, a probabilidade de obter o valor  $X = 9$  ou um valor  $X > 9$  ainda mais desfavorável para a hipótese) da hipótese  $H_0 : \theta = 1/2$  decorrente da observação difere nos dois casos.

No caso  $E_1$ ,  $X$  tem distribuição binomial –  $X \sim \text{Bin}(12, \theta)$  – cujo nível crítico é

$$P_1 = P(X \geq 9 | \theta = \frac{1}{2}) = \binom{12}{9} \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^3 + \binom{12}{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^2 + \binom{12}{11} \left(\frac{1}{2}\right)^{11} \left(\frac{1}{2}\right)^1 + \binom{12}{12} \left(\frac{1}{2}\right)^{12} \left(\frac{1}{2}\right)^0 \doteq 0.0730 = 7.30\%.$$

No caso  $E_2$ ,  $X$  tem distribuição binomial negativa –  $X \sim \text{BN}(3, 1 - \theta)$  – que tem nível crítico

$$P_2 = P(X \geq 9 | \theta = \frac{1}{2}) = \binom{11}{9} \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^2 + \binom{12}{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^2 + \binom{13}{11} \left(\frac{1}{2}\right)^{11} \left(\frac{1}{2}\right)^2 + \dots \doteq 0.0327 = 3.27\%.$$

Logo, se for adotado um limiar de significância de 5%,  $H_0$  é rejeitada no caso de  $E_2$  mas o contrário ocorre com  $E_1$ . Segundo o princípio da verossimilhança esta análise não é correta, pois as conclusões tiradas nos dois casos devem ser idênticas, visto que em qualquer deles a função de verossimilhança é proporcional a  $\theta^9(1 - \theta)^3$ . De fato, as verossimilhanças são

- $E_1 : L_1(\theta|x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x},$

onde, com  $n = 12$  e  $x = 9$ ,  $L_1(\theta|9) = \binom{12}{9} \theta^9 (1 - \theta)^3 = 220 \theta^9 (1 - \theta)^3 \propto \theta^9 (1 - \theta)^3;$

- $E_2 : L_2(\theta|x) = \binom{k+x-1}{x} \theta^x (1 - \theta)^k,$

onde, com  $k = 3$  e  $x = 9$ ,  $L_2(\theta|9) = \binom{11}{9} \theta^9 (1 - \theta)^3 = 55 \theta^9 (1 - \theta)^3 \propto \theta^9 (1 - \theta)^3.$

☺

## 2.4 Estatísticas suficientes

Qualquer função real  $t(x_1, \dots, x_n)$  de observações em um espaço amostral  $\Omega$  é chamada *estatística* (DeGroot and Schervish [2002]). Sejam as informações amostrais  $x_i$ ,  $i = 1, \dots, n$  e os parâmetros  $\theta$  multidimensionais. A classe  $\mathcal{F}$  é uma *família exponencial* se todos os membros têm a forma

$$f(x_i|\theta) = p(x_i)g(\theta)\exp\{\phi(\theta)^T u(x_i)\}.$$

Os fatores  $\phi(\theta)$  e  $u(x_i)$  são, em geral, vetores de mesma dimensão de  $\theta$ . O vetor  $\phi(\theta)$  é chamado “parâmetro natural” da família  $\mathcal{F}$ . A verossimilhança correspondente à sequência  $x = (x_1, \dots, x_n)$  de observações iid é

$$f(x|\theta) = \prod_{i=1}^n p(x_i)g(\theta)^n \exp\left\{\phi(\theta)^T \sum_{i=1}^n u(x_i)\right\}.$$

Para todo  $n$  e  $x$ , esta expressão tem a forma fixa (como função de  $\theta$ ) dada por

$$f(x|\theta) \propto g(\theta)^n \exp\{\phi(\theta)^T t(x)\}, \text{ onde } t(x) = \sum_{i=1}^n u(x_i).$$

A quantidade  $t(x)$  é chamada *estatística suficiente* para  $\theta$  pois a verossimilhança para  $\theta$  depende dos dados  $x$  somente através do valor  $t(x)$  (Gelman et al. [2004]). Isto significa que o conhecimento contido em  $t(x)$  resume toda a informação dos dados amostrais  $(x_1, \dots, x_n)$ , tornando o conhecimento dos valores individuais da amostra irrelevantes para a estimação de  $\theta$ .

No Exemplo 1 mostrou-se que para estimar  $\theta$  baseado no princípio da verossimilhança, basta saber:

- o número de lançamentos e de faces cara ou
- o número de lançamentos e de faces coroa ou
- o número de faces cara e coroa.

Fazendo

$$y_i = \begin{cases} 1 & \text{se der cara} \\ 0 & \text{se der coroa} \end{cases} \quad \text{e} \quad t(y) = \sum_{i=1}^n y_i,$$

temos que uma estatística suficiente para o Exemplo 1 é

$$t(y) = \sum_{i=1}^{10} y_i = 1 + 0 + 1 + 1 + 0 + 0 + 1 + 0 + 0 + 0 = 4.$$

No caso, temos que a função de verossimilhança é proporcional a

$$\theta^{t(y)}(1 - \theta)^{n-t(y)} = \theta^4(1 - \theta)^6.$$

**Teorema 1** (Critério da fatoração). *Seja  $x_1, \dots, x_n$  uma amostra aleatória de uma distribuição contínua ou discreta cuja fdp ou fp é  $f(x|\theta)$ , onde o valor  $\theta$  é desconhecido e pertence a um dado espaço paramétrico  $\Omega$ . Uma estatística  $t(x_1, \dots, x_n)$  é uma estatística suficiente para  $\theta$  se e somente se para todos os valores de  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  e todos os valores de  $\theta \in \Omega$  a fdp conjunta ou a fp conjunta  $f_n(x|\theta)$  de  $x_1, \dots, x_n$  pode ser fatorada como*

$$f_n(x|\theta) = u(x)v[r(x), \theta]. \quad (2.6)$$

*Aqui, as funções  $u$  e  $v$  são não negativas; a função  $u$  pode depender de  $x$  mas não depende de  $\theta$  e a função  $v$  dependerá do valor observado  $x$  somente através do valor da estatística  $r(x)$ .*

*Demonstração.* Vide [DeGroot and Schervish \[2002\]](#), página 372. □

Para qualquer valor de  $x$  cuja  $f_n(x|\theta) = 0$  para todos os valores de  $\theta \in \Omega$ , o valor da função  $u(x)$  na Eq. (2.6) pode ser escolhido para ser 0. Assim, quando o critério da fatoração estiver sendo aplicado, é suficiente verificar que a fatoração da forma dada na Eq. (2.6) é satisfeita para cada valor de  $x$  tal que  $f_n(x|\theta) > 0$  para pelo menos um valor de  $\theta \in \Omega$ .

## 2.5 Informação *a priori*

A Inferência Bayesiana tem entre seus fundamentos o fato de possuir um mecanismo formal para agregar informação externa ao experimento, chamada informação *a priori*. Esta informação está em poder de um especialista do problema, que pode ser um estatístico, pesquisador ou qualquer pessoa interessada em fazer afirmações sobre entidades desconhecidas de maneira racional. Boa parte desta informação é de caráter subjetivo, e pode estar radicada em fontes objetivas (dados históricos, literatura) ou em conclusões pessoais tiradas pelo especialista (experiência, conhecimento tácito).

A quantificação probabilística destas crenças *a priori*, especificando a *distribuição a priori*, recorre a conceitos de probabilidade estranhos à Inferência Frequentista. Um dos pontos centrais do revés entre as escolas bayesiana e clássica é o fato de os bayesianos aceitarem incluir informações subjetivas ao abordar um problema, fato rejeitado pelos frequentistas. Lindley [1990] sugere que se um clássico diz a um bayesiano “*where did you get that prior?*”, o bayesiano deve responder “*where did you get that sample space?*” (Paulino et al. [2003]), mostrando de certa maneira que os frequentistas também utilizam informações subjetivas. Segundo Gelman et al. [2004], “Todos os métodos estatísticos que utilizam probabilidade são subjetivos no sentido de dependerem de idealizações matemáticas do mundo”.

### 2.5.1 Distribuições *a priori* conjugadas

Os dois procedimentos básicos utilizados em Inferência Bayesiana são otimização e integração. O problema é que quando as funções são um pouco mais complexas, realizar estas duas tarefas pode se tornar um trabalho infatigável. Para contornar em parte estes problemas, escolhem-se *prioris conjugadas*, ou seja, distribuições *a priori* convenientes para utilizar com amostras obtidas de outras distribuições, que levam a uma *posteriori* da mesma família da *priori*, mas com parâmetros atualizados pelos dados sem a necessidade de cálculos.

#### Famílias conjugadas

A propriedade de uma distribuição *posteriori* seguir a mesma forma paramétrica da distribuição *priori* é chamada *conjugação*. A família conjugada é matematicamente conveniente no sentido de a distribuição *posteriori* seguir uma forma paramétrica conhecida. Obviamente se a informação disponível contradisser a família paramétrica conjugada, será necessário usar uma distribuição *priori* mais realística, ainda que inconveniente (Gelman et al. [2004]).

Formalmente, diz-se que se  $\mathcal{F}$  é uma classe de distribuições amostrais  $f(x|\theta)$ , então a classe  $\mathcal{H}$  é *conjugada* de (ou para)  $\mathcal{F}$  se

$$h(\theta) \in \mathcal{H} \Rightarrow h(\theta)f(x|\theta) \propto h(\theta|x) \in \mathcal{H}.$$

Nestas condições diz-se também que  $\mathcal{H}$  é uma *fechada sob amostragem* de (ou para)  $\mathcal{F}$ .

Esta definição é formalmente vaga pois se escolhe-se  $\mathcal{H}$  como a classe de todas as distribuições, então  $\mathcal{H}$  é sempre conjugada, não importando qual classe de distribuições amostrais seja utilizada. Estamos mais interessados nas famílias de *prioris* conjugadas *naturais*, que são obtidas considerando  $\mathcal{H}$  como o conjunto de todas as densidades tendo a mesma forma funcional da verossimilhança.

Neste contexto mais restrito, de modo a utilizar o conceito de conjugação na prática, temos que  $\mathcal{H} = \{h(\theta|a) : a \in \mathcal{A}\}$ , onde  $\mathcal{A}$  denota o conjunto de valores para os hiperparâmetros (parâmetros da distribuição *a priori*). Assim,  $\mathcal{H}$  é fechada sob amostragem de  $\mathcal{F}$  se  $L(\theta|x) \equiv f(x|\theta)$ , para cada  $x$ , é proporcional a um membro de  $\mathcal{H}$  e  $\mathcal{H}$  é fechada em relação a produtos, ou seja, para todo  $a_0, a_1 \in \mathcal{A}$ , existe  $a_2 \in \mathcal{A}$  tal que

$$h(\theta|a_0)h(\theta|a_1) \propto h(\theta|a_2).$$

A família  $\mathcal{H}$  diz-se então *conjugada natural* de  $\mathcal{F}$  (Paulino et al. [2003]).

### Conjugação na família exponencial

A subclasse mais importante da família de distribuições admitindo estatísticas suficientes de dimensionalidade fixa é a família exponencial Paulino et al. [2003].

**Definição 1** (Distribuição Normal Padrão). *A função definida por*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (2.7)$$

*é chamada função densidade normal padrão. Sua integral*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt \quad (2.8)$$

*é a função distribuição normal padrão (Feller [1968]).*

**Teorema 2** (Dirichlet conjugada da Multinomial). *Seja  $\mathbf{X}_1, \dots, \mathbf{X}_n$  uma amostra aleatória de uma distribuição multinomial de tamanho  $n$  onde  $\mathbf{X}_i = (X_{i1}, \dots, X_{ik})'$ ,  $i = 1, \dots, n$ , e cujo vetor de parâmetros  $(\theta_1, \dots, \theta_k)$  é desconhecido ( $0 < \theta_j < 1$ ,  $j = 1, \dots, k$  e  $\sum_{j=1}^k \theta_j = 1$ ). Suponha que a distribuição a priori do vetor  $(\theta_1, \dots, \theta_k)$  é uma Dirichlet com vetor de parâmetros  $(d_1, \dots, d_k)$  ( $d_j > 0$ ). Então a distribuição a posteriori do vetor  $(\theta_1, \dots, \theta_k)$  dado que  $\mathbf{X}_i = \mathbf{x}_i$  ( $i = 1, \dots, n$ ) é uma Dirichlet com vetor de parâmetros  $(d_1 + \sum_{i=1}^n x_{i1}, \dots, d_k + \sum_{i=1}^n x_{ik})$ .*

*Demonstração.* Vide DeGroot [1970], página 174. □

### 2.5.2 Distribuições *a priori* não-informativas

Em diversos casos não tem-se uma opinião objetiva ou subjetiva sobre um parâmetro ou vetor de parâmetros, situação conhecida como “ignorância *a priori*” (Paulino et al. [2003]). Distribuições que conseguem representar a ignorância em algum sentido, são chamadas de *prioris vagas, difusas* ou *não-informativas*. Poderia-se dizer que o pesquisador ou o estatístico “deixam os dados falarem por si mesmos”, buscando a situação em que a *posteriori* seja afetada minimamente por informações externas aos dados (Gelman et al. [2004]).

#### Distribuições próprias e impróprias

Uma distribuição de probabilidade é dita *própria* quando não depende dos dados e sua integral sobre o espaço paramétrico  $\Theta$  é 1. Se ela integra qualquer valor finito positivo, é chamada *densidade não normalizada*, podendo ser renormalizada – multiplicada por uma constante apropriada – para integrar 1. Se uma distribuição de probabilidade integra infinito, ela é chamada de *imprópria*.

*Prioris* impróprias podem resultar (não necessariamente) em *posteriores* próprias. Porém, *posteriores* obtidas desta forma devem ser interpretadas com muito cuidado, devendo-se verificar se a *posteriori* tem integral finita e uma forma coerente (Gelman et al. [2004]). O problema ocorre quando a distribuição *a posteriori* do parâmetro de interesse é imprópria, inviabilizando naturalmente a realização de inferências (Paulino et al. [2003]).

Procurar por uma distribuição *a priori* que é sempre vaga é um engano: se a verossimilhança é realmente dominante em algum dado problema, então a escolha entre uma gama de *prioris* não informativas não deve fazer diferença. Estabelecendo uma particular especificação como a distribuição *priori* de referência, parece encorajar seu uso automático, mas possivelmente inapropriado (Gelman et al. [2004]).

#### Método de Bayes-Laplace

O primeiro argumento utilizado para gerar distribuições não informativas foi o Princípio da Razão Insuficiente de Bayes e Laplace. De acordo com este princípio, na ausência de razão suficiente para privilegiar alguma possibilidade em detrimento das outras decorrente da escassez de informação *a priori*, deve-se adotar a equiprobabilidade. A consideração deste argumento como objetivista não é totalmente pacífica, já que é questionável falar-se de “ausência de razão suficiente” sem envolvimento de juízos subjetivos.

No caso em que  $\Theta$  é finito, digamos  $\Theta = \{\theta_1, \dots, \theta_k\}$ , a distribuição não informativa gerada por este argumento é a distribuição Uniforme Discreta

$$h(\theta) = \frac{1}{k}, \theta \in \Theta.$$

Quando  $\Theta$  é infinito enumerável, é sabido que não existe nenhuma distribuição de probabilidade seguindo

os axiomas de Kolmogorov compatível com a equiprobabilidade de todos os valores possíveis de  $\theta$ . Distribuições nestes casos são impróprias.

No caso de  $\Theta$  infinito não enumerável, o Princípio da Razão Insuficiente é interpretado de modo a conduzir à distribuição Uniforme Contínua, que é imprópria se  $\Theta$  não é limitado (Paulino et al. [2003]).

Em muitos problemas não existe uma escolha clara para uma *priori* vaga, visto que uma densidade que é uniforme em uma parametrização não será em outra. Esta é a dificuldade com o Princípio da Razão Insuficiente: em qual escala o princípio deve ser aplicado? Por exemplo, parece ser razoável definirmos uma *priori* uniforme para a média  $\theta$  e a densidade  $h(\sigma^2) \propto 1/\sigma^2$  para  $\sigma^2$  em uma  $N(\theta, \sigma^2)$ . Porém, se definirmos  $\phi = \log \sigma^2$ , então a *priori* de  $\phi$  é

$$h(\phi) = h(\sigma^2) \left| \frac{d\sigma^2}{d\phi} \right| \propto \frac{1}{\sigma^2} \sigma^2 = 1,$$

ou seja, uniforme em  $\phi = \log \sigma^2$ . Com distribuições discretas existe uma dificuldade análoga em decidir como subdividir os resultados em segmentos de igual probabilidade (Gelman et al. [2004]).

### Método de Jeffrey

Jeffrey introduziu uma abordagem para definir distribuições *a priori* não informativas considerando transformações biunívocas do parâmetro  $\phi = g(\theta)$ . Mesmo transformando-se as variáveis, a *priori*  $h(\theta)$  é equivalente, de modo que representa as mesmas crenças que a *priori*  $h(\phi)$ , dada por

$$h(\phi) = h(\theta) \left| \frac{d\theta}{d\phi} \right| = h(\theta) |g'(\theta)|^{-1}. \quad (2.9)$$

O princípio geral de Jeffrey diz que qualquer regra para determinar a *priori*  $h(\theta)$  deve ter resultado equivalente se aplicado no parâmetro transformado, ou seja,  $h(\phi)$  calculado determinando-se  $h(\theta)$  e aplicando-se (2.9) deve levar à mesma distribuição obtida determinando-se  $h(\phi)$  diretamente usando o modelo transformado  $h(x, \phi) = h(x|\phi)h(\phi)$ .

Este princípio leva a uma definição de *priori* não informativa invariante sob transformações biunívocas como  $h(\theta) \propto [I(\theta)]^{1/2}$ , onde  $I(\theta)$  é a *informação de Fisher* para  $\theta$ , dada por

$$I(\theta) = E \left[ \left( \frac{d \log f(x|\theta)}{d\theta} \right)^2 \middle| \theta \right] = -E \left[ \frac{d^2 \log f(x|\theta)}{d\theta^2} \middle| \theta \right]. \quad (2.10)$$

O princípio de Jeffrey pode ser estendido para o caso multiparamétrico, mas os resultados são mais controversos. Abordagens simples baseadas em assumir *prioris* não informativas independentes para os componentes do vetor de parâmetros  $\theta$  podem dar resultados diferentes quando obtidos pelo princípio de Jeffrey. Quando o número de parâmetros cresce, é mais viável abandonar *prioris* não informativas puras e considerar modelos



hierárquicos (Gelman et al. [2004]).

## 2.6 Informação a posteriori

O processo de Inferência Bayesiana consiste em atualizar a distribuição *a priori*  $h(\theta)$  para uma distribuição *a posteriori*  $h(\theta|x)$ , e é natural esperar que algumas relações naturais possam se manter entre estas duas distribuições. Por exemplo, pode-se esperar que, devido à incorporação dos dados pela *posteriori*, esta será menos variável que a *priori*. Esta noção é formalizada na expressão 2.12.

$$E(\theta) = E(E(\theta|X)) \quad (2.11)$$

e

$$Var(\theta) = E(Var(\theta|X)) + Var(E(\theta|X)). \quad (2.12)$$

O resultado expreso pela equação (2.11) indica que a média (*mean*) *a priori* de  $\theta$  é a média (*average*) de todas as possíveis médias *a posteriori* sobre a distribuição dos dados possíveis. A fórmula em (2.12) é interessante pois mostra que a variância da *posteriori* é menor em média do que a variância da *priori*, por uma quantidade que depende das variações nas médias *a posteriori* sobre a distribuição de todos os dados possíveis. Quanto maior for a variação nos dados, maior será o potencial para reduzir nossa incerteza com relação a  $\theta$ .

É claro que as relações de média e variância descrevem apenas valores esperados, e em situações específicas que a variância da *posteriori* pode ser similar ou até maior que a variância da *priori*, ainda que isto possa ser um indicativo de conflito ou inconsistência entre o modelo amostral e a distribuição *a priori*. Esta é uma característica muito geral da Inferência Bayesiana: a *posteriori* é centrada em um ponto que representa um meio termo entre a informação *a priori* e os dados, e este meio termo é dominado pelos dados à medida que o tamanho da amostra aumenta (Gelman et al. [2004]).

### 2.6.1 Resumindo a inferência a posteriori

A distribuição de probabilidade *a posteriori* contém toda a informação até o presente momento sobre  $\theta$ . Uma vantagem da abordagem Bayesiana é a flexibilidade com a qual as inferências *a posteriori* podem ser resumidas, mesmo após transformações complicadas.

As medidas resumo de locação mais utilizadas são a média, mediana e moda(s) da *posteriori*. A variação é comumente resumida pelo desvio padrão, intervalo interquartilico e outros quantis. A moda é importante em estratégias computacionais em problemas mais complexos pois é frequentemente mais fácil de ser calculada se comparada à média ou à mediana. Quando uma distribuição *posteriori* tem forma fechada, medidas resumo

tais como média, mediana e desvio padrão são utilizadas (Gelman et al. [2004]).

Em um problema de decisão definido por um espaço amostral  $\Omega$ , um espaço de decisão  $D$  e uma função perda  $L$ , o *risco de Bayes* é definido como o maior limite inferior para os riscos  $\rho(P, d)$  para todas as decisões  $d \in D$ , para qualquer distribuição  $P$  do parâmetro  $W$ . Qualquer decisão  $d^*$  cujo risco seja igual ao risco de Bayes é chamada uma *decisão de Bayes contra a distribuição  $P$* . Se a distribuição de um parâmetro  $W$  é  $P$ , qualquer decisão de Bayes contra  $P$  será uma decisão ótima para o estatístico porque o risco não pode ser menor para qualquer outra decisão (DeGroot [1970]). Isto conecta a Inferência Bayesiana à Teoria da Decisão, de forma que as medidas resumo utilizadas são pontos ótimos de alguma função perda  $L$ .

## 2.7 Métodos de Monte Carlo

Em Inferência Bayesiana há uma constante necessidade em resolver integrais para a obtenção da *posteriori*. Esta tarefa é frequentemente complexa, se não impossível de ser realizada analiticamente. Para isso, utilizam-se os métodos de Monte Carlo, algoritmos criados para calcular aproximadamente o valor de integrais definidas. São uma boa alternativa aos métodos numéricos, principalmente em cenários multidimensionais.

Ao contrário de algoritmos usuais, estes métodos estão baseados em simulação estocástica, ou seja, escolhem pontos aleatoriamente de uma distribuição de probabilidade nos quais o integrando – no caso a *posteriori* – será avaliado. Para executar o método, é necessário gerar números pseudo-aleatórios de uma distribuição Uniforme em  $(0, 1)$ . Este procedimento de geração de pontos  $U(0, 1)$  está implementado em praticamente todos os *softwares* matemáticos e estatísticos, não sendo considerado com detalhes nesse texto. Para maior profundidade, o leitor pode consultar Devroye and Devroye [1986], Ripley [1987] e Gamerman and Lopes [2006].

### 2.7.1 Monte Carlo Ordinário

O algoritmo calcula uma estimativa da integral

$$\int g(\theta)h(\theta|x)d\theta = E[g(\theta)|x], \quad (2.13)$$

onde  $\theta$  e  $x$  podem ser vetores.

Simulando uma amostra aleatória  $\theta_1, \dots, \theta_n$  da densidade *a posteriori*  $h(\theta|x)$ , o método de Monte Carlo Ordinário aproxima a integral (2.13) pela média empírica

$$\hat{E}[g(\theta)|x] = \frac{1}{n} \sum_{i=1}^n g(\theta_i), \quad (2.14)$$

que pela Lei dos Grande Números converge quase certamente para  $E[g(\theta)|x]$ . A precisão desta aproximação

pode ser medida pelo erro padrão (estimado) de Monte Carlo, dado por

$$\frac{1}{\sqrt{n(n-1)}} \left\{ \sum_{i=1}^n \left[ g(\theta_i) - \frac{1}{n} \sum_{i=1}^n g(\theta_i) \right]^2 \right\}^{1/2}. \quad (2.15)$$

Fica claro então que este é um método aproximado, podendo ocorrer situações do tipo  $F(x) < F(y)$ ,  $x > y$ , onde  $F$  é uma função distribuição. A precisão do método fica à critério do analista, que pode aumentar o tamanho  $n$  da amostra aleatória gerada, limitado apenas pela capacidade de processamento de seu computador e de sua habilidade em alocar memória de forma eficiente.

Caso o problema tenha uma dimensão muito grande e este método perca eficiência, talvez seja interessante considerar métodos de Monte Carlo baseados na simulação de Cadeias de Markov, chamados de Monte Carlo Markov Chains ou MCMC ([Paulino et al. \[2003\]](#)).



## Capítulo 3

### Inferência para Proporção

Pode-se dividir as pesquisas eleitorais em dois grandes momentos: a amostragem e a análise dos dados. A questão da amostragem não será abordada neste trabalho, sendo que o foco é a análise dos dados. Para mais detalhes sobre amostragem vide [Bolfarine and Bussab \[2005\]](#) e [Cochran \[1977\]](#).

Este capítulo faz um comparativo entre três métodos: i) institutos, ii) frequentista correta e iii) bayesiana. A criação de uma teoria frequentista diferente daquela utilizada pelos institutos deve-se ao fato de o autor não concordar com o tratamento usual dado pelos institutos, por motivos explicados nas seções subsequentes.

#### 3.1 Independência

Dizer que dois eventos são independentes equivale a afirmar que a ocorrência de um não interfere na probabilidade de ocorrência do outro, ou seja,  $P(A = a|B = b) = P(A = a)$  e  $P(B = b|A = a) = P(B = b)$ . No caso de amostragem, as unidades amostrais são independentes se for realizado um plano amostral *com reposição* ([Bolfarine and Bussab \[2005\]](#)).

**Exemplo 3.** *Suponha que temos uma urna com 3 bolas azuis e 2 bolas brancas. Sejam os eventos  $A$  : retirar a bolas azuis em 3 extrações e  $B$  : retirar  $b$  bolas brancas em 3 extrações.  $A$  e  $B$  serão considerados independentes se, e somente se, as extrações das bolas forem realizadas com reposição.*

*No caso com reposição temos que  $A \sim \text{Bin}(3, 3/5)$  e  $B \sim \text{Bin}(3, 2/5)$ . As probabilidades associadas são*

$$\begin{aligned}P(A = 0) &= P(A = 0|B = b) = \binom{3}{0} \left(\frac{3}{5}\right)^0 \left(\frac{2}{5}\right)^3 = \frac{8}{125} = 0.064 = P(B = 3) = P(B = 3|A = a) \\P(A = 1) &= P(A = 1|B = b) = \binom{3}{1} \left(\frac{3}{5}\right)^1 \left(\frac{2}{5}\right)^2 = \frac{36}{125} = 0.288 = P(B = 2) = P(B = 2|A = a) \\P(A = 2) &= P(A = 2|B = b) = \binom{3}{2} \left(\frac{3}{5}\right)^2 \left(\frac{2}{5}\right)^1 = \frac{54}{125} = 0.432 = P(B = 1) = P(B = 1|A = a) \\P(A = 3) &= P(A = 3|B = b) = \binom{3}{3} \left(\frac{3}{5}\right)^3 \left(\frac{2}{5}\right)^0 = \frac{27}{125} = 0.216 = P(B = 0) = P(B = 0|A = a),\end{aligned}$$

onde  $a \in \{0, 1, 2, 3\}$  e  $b \in \{0, 1, 2, 3\}$ .

*Se fizessemos extrações sem reposição, os resultados seriam diferentes:*

$$\begin{aligned}
P(A = 0) &= \binom{3}{0} \left(\frac{2}{5}\right) \left(\frac{1}{4}\right) \left(\frac{0}{3}\right) = \left(\frac{0}{60}\right) = 0.00 = P(B = 3) \\
P(A = 1) &= \binom{3}{1} \left(\frac{3}{5}\right) \left(\frac{2}{4}\right) \left(\frac{1}{3}\right) = \left(\frac{18}{60}\right) = 0.30 = P(B = 2) \\
P(A = 2) &= \binom{3}{2} \left(\frac{3}{5}\right) \left(\frac{2}{4}\right) \left(\frac{2}{3}\right) = \left(\frac{36}{60}\right) = 0.60 = P(B = 1) \\
P(A = 3) &= \binom{3}{3} \left(\frac{3}{5}\right) \left(\frac{2}{4}\right) \left(\frac{1}{3}\right) = \left(\frac{6}{60}\right) = 0.10 = P(B = 0),
\end{aligned}$$

com as probabilidades condicionais

$$P(A = i|B = j) = 0, \text{ se } i + j \neq 3$$

e

$$P(A = i|B = j) = 1, \text{ se } i + j = 3.$$

☺

Percebe-se então que a noção de independência é muito importante quando calcula-se probabilidades de eventos, mas para grandes populações praticamente não existe diferença entre amostrar com ou sem reposição (Feller [1968]). Este resultado é esperado, pois quanto maior é o número de bolas na urna, menor é a chance de sortearmos uma em particular. Estendendo este raciocínio para um cenário eleitoral com milhares ou milhões de eleitores, fica claro que não é absurdo tratar uma amostra *sem* reposição de maneira independente. Este resultado ficará mais claro com as demonstrações dos Teoremas 3 e 5, apresentadas na Seção 3.3.3.

## 3.2 Institutos

Nesta seção trataremos da abordagem de cunho frequentista em relação à estimação de proporções feita pelos principais institutos de pesquisa no Brasil. Será feita uma discussão baseada em fundamentos de probabilidade aceitos tanto por bayesianos quanto por frequentistas, indicando, quando necessário, os pontos de divergência entre as duas escolas.

### 3.2.1 Estimação por ponto

A estimativa por ponto da proporção populacional dos  $k$  candidatos é obtida através da proporção amostral, que é um estimador não viesado para a proporção populacional (Bolfarine and Bussab [2005]). A estimativa pontual do  $i$ -ésimo candidato é dada por

$$\hat{p}_i = \frac{\text{número de votos para o candidato } i \text{ na amostra}}{\text{número total de votos na amostra}}, \quad (3.1)$$

onde  $i = 1, \dots, k$ .

### 3.2.2 Estimação por intervalo de confiança

Segundo os conceitos da Inferência Frequentista, não basta apenas apresentar uma estimativa pontual como solução para um problema de inferência: é preciso que essa estimativa venha acompanhada de uma medida indicando sua variabilidade. Em problemas unidimensionais esta medida é a variância do estimador. No caso de dois ou mais parâmetros, deve-se considerar também a correlação entre os parâmetros, utilizando-se para isso a matriz de covariâncias estimada.

O fato é que os institutos de pesquisa brasileiros não calculam regiões de confiança para os  $k$  candidatos em um processo eleitoral. Calcula-se um intervalo de confiança *individual* para cada candidato (Nunes [2004]), não levando em consideração a estrutura de covariâncias inerente ao problema de estimação de proporções. Não releva-se o fato de a soma das proporções ser 1, ou seja, em intervalos de confiança que tratam as proporções de forma independente não considera-se que se a proporção de votos do candidato  $i$  aumenta, a dos demais necessariamente deve cair. As subseções 3.2.3, 3.2.4 e 3.2.5 tratam cada particular caso com detalhes.

**Definição 2.** *Empate técnico é a situação na qual não é possível dizer qual candidato está à frente de outro em uma pesquisa eleitoral, segundo os preceitos da Inferência Frequentista.*

Segundo Nunes [2004], “A superposição dos intervalos de confiança dos candidatos determina o empate técnico. É considerado empate técnico quando a diferença entre os candidatos se encontra dentro das margens de erro das pesquisas, ou seja, quando há superposição dos respectivos intervalos de confiança dos candidatos.”. O exemplo fornecido por Nunes [2004] considera dois candidatos.

### 3.2.3 Dois Candidatos

O cenário de dois candidatos é o único que não viola princípios básicos de probabilidade, *quando tratado assintoticamente*. Isto significa dizer que quando sobrepõem-se os intervalos de confiança dos dois candidatos, os valores que levam ao empate técnico são praticamente os mesmos quando fazemos o intervalo para a diferença das proporções populacionais. Veja a subseção 3.3.3, que apresenta a construção do intervalo para a diferença.

#### Sobreposição de intervalos de confiança e empate técnico

Em um cenário eleitoral com dois candidatos (como o segundo turno), os institutos procedem da seguinte forma:

- Calculam-se os intervalos de confiança  $1 - \alpha$  para as proporções  $p_{A_1}$  e  $p_{A_2}$  dos candidatos  $A_1$  e  $A_2$ ,

respectivamente por

$$IC(p_{A_1}, 1 - \alpha) = \hat{p}_{A_1} \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_{A_1}(1 - \hat{p}_{A_1})}{n}} \quad (3.2)$$

e

$$IC(p_{A_2}, 1 - \alpha) = \hat{p}_{A_2} \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_{A_2}(1 - \hat{p}_{A_2})}{n}}, \quad (3.3)$$

onde  $n$ ,  $\hat{p}_{A_1}$  e  $\hat{p}_{A_2}$  são, respectivamente, o tamanho da amostra e a proporção de votos para os candidatos  $A_1$  e  $A_2$  na amostra – após dividir proporcionalmente os votos dos indecisos entre  $A_1$ ,  $A_2$ , brancos e nulos e excluir os votos não válidos (brancos e nulos).  $z_{1-\alpha/2}$  é o quantil da normal padrão que acumula  $1 - \alpha/2$  de probabilidade.

- Se os intervalos se sobrepuserem, é declarado empate técnico, ou seja, que nada pode-se afirmar sobre quem está à frente na pesquisa. Isto ocorre pois, pelo enfoque dos institutos de pesquisa, ainda que um candidato (suponha  $A_1$ ) tenha uma estimativa pontual maior da proporção de votos, a distância para o segundo colocado não seria grande suficiente para afirmar que a diferença é *estatisticamente significativa*.

**Exemplo 4.** Suponha que o número de votos para  $A_1$  em uma amostra seja 217 e para  $A_2$  seja 181. Logo, as proporções estimadas de votos são

$$\hat{p}_{A_1} = \frac{217}{217 + 181} = \frac{217}{398} \doteq 0.5452$$

e

$$\hat{p}_{A_2} = \frac{181}{217 + 181} = \frac{181}{398} \doteq 0.4548.$$

Por (3.2) e (3.3), os intervalos de confiança  $1 - 0.05 = 0.95 = 95\%$ , considerando  $A_1$  e  $A_2$  independentes, são

$$IC(p_{A_1}, 95\%) = 0.5452 \mp 1.96 \sqrt{\frac{0.5452(1 - 0.5452)}{398}} = (0.4963, 0.5941)$$

e

$$IC(p_{A_2}, 95\%) = 0.4548 \mp 1.96 \sqrt{\frac{0.4548(1 - 0.4548)}{398}} = (0.4059, 0.5037).$$

Como os intervalos se sobrepõem, conforme apresentado na Figura 3.1, a interpretação dos institutos é que pode ocorrer  $0.4963 \leq p_{A_1}, p_{A_2} \leq 0.5037$ , não podendo-se portanto afirmar quem está à frente.



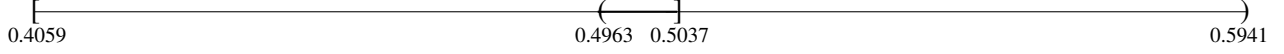


Figura 3.1: Sobreposição dos ICs para a proporção de votos para os candidatos  $A_1$  e  $A_2$ .

### 3.2.4 Três Candidatos

Em um cenário com três candidatos já percebe-se os efeitos da desconsideração com a estrutura de co-variâncias e com a omissão de a soma das proporções ser 1.

#### Sobreposição de intervalos de confiança e empate técnico

Em um cenário eleitoral com três candidatos, os institutos procedem da seguinte forma:

- Calcula-se o intervalo de confiança  $1 - \alpha$  para as proporções dos candidatos  $A_1$  e  $A_2$  respectivamente por (3.2) e (3.3), e para o candidato  $A_3$  de forma análoga, por

$$IC(p_{A_3}, 1 - \alpha) = \hat{p}_{A_3} \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_{A_3}(1 - \hat{p}_{A_3})}{n}}, \quad (3.4)$$

onde  $n$ ,  $\hat{p}_{A_1}$ ,  $\hat{p}_{A_2}$  e  $\hat{p}_{A_3}$  são, respectivamente, o tamanho da amostra e a proporção de votos para os candidatos  $A_1$ ,  $A_2$  e  $A_3$  na amostra – após dividir proporcionalmente os votos dos indecisos entre  $A_1$ ,  $A_2$ ,  $A_3$ , brancos e nulos e excluir os votos não válidos (brancos e nulos).  $z_{1-\alpha/2}$  é o quantil da normal padrão que acumula  $1 - \alpha/2$  de probabilidade.

- O empate técnico agora é mais complicado se comparado ao caso de dois candidatos, visto que existem mais situações possíveis. Porém, a idéia de sobreposição se mantém.

**Exemplo 5.** Suponha que o número de votos para  $A_1$  em uma amostra seja 217, para  $A_2$  seja 181 e para  $A_3$  seja 170. Logo, as proporções estimadas de votos são

$$\hat{p}_{A_1} = \frac{217}{217 + 181 + 170} = \frac{217}{568} \doteq 0.3820,$$

$$\hat{p}_{A_2} = \frac{181}{217 + 181 + 170} = \frac{181}{568} \doteq 0.3187$$

e

$$\hat{p}_{A_3} = \frac{170}{217 + 181 + 170} = \frac{170}{568} \doteq 0.2993$$

Por (3.2), (3.3) e (3.4), os intervalos de confiança  $1 - 0.05 = 0.95 = 95\%$ , considerando  $A_1$ ,  $A_2$  e  $A_3$

independentes, são

$$IC(p_{A_1}, 95\%) = 0.3820 \pm 1.96 \sqrt{\frac{0.3820(1 - 0.3820)}{568}} = (0.3421, 0.4220), \quad ( \quad )$$

$$IC(p_{A_2}, 95\%) = 0.3187 \pm 1.96 \sqrt{\frac{0.3187(1 - 0.3187)}{568}} = (0.2803, 0.3570) \quad [ \quad ]$$

e

$$IC(p_{A_3}, 95\%) = 0.2993 \pm 1.96 \sqrt{\frac{0.2993(1 - 0.2993)}{568}} = (0.2616, 0.3370). \quad \{ \quad \}$$

Como os intervalos se sobrepõem, conforme apresentado na Figura 3.2, a interpretação dos institutos é que pode ocorrer  $0.2803 \leq p_{A_2}, p_{A_3} \leq 0.3370$  e  $0.3421 \leq p_{A_1}, p_{A_2} \leq 0.33570$ , não podendo-se portanto afirmar quem está à frente.

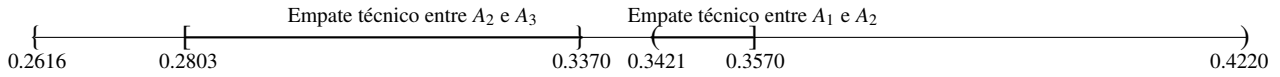


Figura 3.2: Sobreposição dos ICs para a proporção de votos para os candidatos  $A_1$ ,  $A_2$  e  $A_3$

☺

Este é um exemplo escolhido propositalmente, para mostrar como podem ocorrer situações confusas neste tipo de estimação. Temos que os intervalos de confiança de  $A_1$  e  $A_2$  se sobrepõem, indicando empate técnico entre estes dois candidatos. Na mesma condição estão  $A_2$  e  $A_3$ . Assim, surge uma conclusão um tanto controversa:  $A_1$  está tecnicamente empatado com  $A_2$ ,  $A_2$  está tecnicamente empatado com  $A_3$ , mas  $A_1$  e  $A_3$  não estão empatados. Ou estão?

### 3.2.5 $k$ Candidatos

Na situação geral de  $k$  candidatos, utilizada pelos principais institutos de pesquisa do Brasil e largamente divulgada pela mídia, temos um cenário ainda mais complexo.

#### Sobreposição de intervalos de confiança e empate técnico

Em um cenário eleitoral geral com  $k$  candidatos, os institutos procedem da seguinte forma:

- Calcula-se o intervalo de confiança  $1 - \alpha$  para as proporções de cada um dos candidatos  $A_i$ ,  $i = 1, \dots, k$

por

$$IC(p_{A_i}, 1 - \alpha) = \hat{p}_{A_i} \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_{A_i}(1 - \hat{p}_{A_i})}{n}}, \quad (3.5)$$

onde  $n$ ,  $\hat{p}_{A_i}$  são, respectivamente, o tamanho da amostra e a proporção de votos para o  $i$ -ésimo candidato  $A_i$  na amostra – após dividir proporcionalmente os votos dos indecisos entre  $A_1, \dots, A_k$ , brancos e nulos e excluir os votos não válidos (brancos e nulos).  $z_{1-\alpha/2}$  é o quantil da normal padrão que acumula  $1 - \alpha/2$  de probabilidade.

- Novamente a idéia de sobreposição se mantém, apesar de poderem ocorrer situações muito sinuosas.

Baseando-se no Exemplo 5, não é difícil imaginar uma situação na qual exista uma sucessão de empates técnicos, onde é extremamente complexo – se não impossível – dizer quem está tecnicamente empatado com quem.

### 3.3 Inferência Frequentista Correta

Em contraponto ao que foi apresentado na Seção 3.2, será feita agora uma abordagem que considera a dependência entre as proporções amostrais em cenários eleitorais. Desta forma construiu-se uma teoria frequentista que não viola os axiomas de probabilidade nem os preceitos adotados por esta escola.

#### 3.3.1 Estimação por ponto

A estimação pontual das proporções de votos é tratada aqui da mesma forma que na Subseção 3.2.1.

#### 3.3.2 Estimação por intervalo e regiões de confiança

Neste enfoque será considerada a estrutura de covariâncias existente entre as proporções amostrais em cenários eleitorais. O caso de dois candidatos será tratado de forma diferente do cenário com três ou mais concorrentes. Nos casos de dois e três candidatos pode-se visualizar as estimativas, o que não ocorre nos casos de dimensão maior.

#### 3.3.3 Dois Candidatos

Quando deseja-se extremo rigor ao estimar proporções de votos com dois candidatos, deve-se tratar o problema com uma distribuição hipergeométrica. Porém surgem complicações nos cálculos, visto que envolvem combinações de números muito grandes. Mas o principal problema deste enfoque é encontrar um *pivô* ou *quantidade pivotal*, ou seja, uma transformação que nos leve a uma distribuição que não dependa dos parâmetros populacionais.

Sejam as quantidades

$N_{A_1}$ : total de votos para o candidato  $A_1$  na população;

$n_{A_1}$ : total de votos para o candidato  $A_1$  na amostra;

$N_{A_2}$ : total de votos para o candidato  $A_2$  na população;

$n_{A_2}$ : total de votos para o candidato  $A_2$  na amostra;

$p_{A_1}$ : proporção de votos para o candidato  $A_1$  na população;

$\hat{p}_{A_1}$ : proporção de votos para o candidato  $A_1$  na amostra;

$p_{A_2}$ : proporção de votos para o candidato  $A_2$  na população;

$\hat{p}_{A_2}$ : proporção de votos para o candidato  $A_2$  na amostra;

$N = N_{A_1} + N_{A_2}$ : total de votos na população;

$n = n_{A_1} + n_{A_2}$ : total de votos na amostra.

**Definição 3** (Hipergeométrica). *Seja uma população com  $N$  elementos, sendo  $N_{A_1}$  do primeiro tipo e  $N_{A_2} = N - N_{A_1}$  do segundo tipo. Um grupo de  $n$  elementos é escolhido aleatoriamente. A probabilidade  $H(n_{A_1}; n, N_{A_1}, N)$  de que o grupo sorteado tenha exatamente  $n_{A_1}$  elementos do tipo  $A_1$  é*

$$H(n_{A_1}; n, N_{A_1}, N) = \frac{\binom{N_{A_1}}{n_{A_1}} \binom{N - N_{A_1}}{n - n_{A_1}}}{\binom{N}{n}}. \quad (3.6)$$

O sistema de probabilidades definido desta forma é chamado distribuição hipergeométrica (Feller [1968]).

**Definição 4** (Binomial). *Seja  $B(n_{A_1}; n, p)$  a probabilidade que  $n$  ensaios de Bernoulli<sup>1</sup> com probabilidade  $p$  para sucesso e  $1 - p$  para fracasso resultem em  $n_{A_1}$  sucessos e  $n - n_{A_1}$  fracassos. Então,*

$$B(n_{A_1}; n, p) = \binom{n}{n_{A_1}} p^{n_{A_1}} (1 - p)^{n - n_{A_1}}. \quad (3.7)$$

O sistema de probabilidades definido desta forma é chamado distribuição binomial (Feller [1968]).

Iniciou-se o estudo supondo-se que a probabilidade de encontrarem-se exatamente  $n_{A_1}$  elementos em uma amostra aleatória de  $n$  elementos segue uma distribuição  $H(N, n, N_{A_1})$ . Após algumas contas simples, chegou-se à distribuição de probabilidade das diferenças dos totais de uma distribuição hipergeométrica. Sua expressão é

$$P(n_{A_1} - n_{A_2} = x) = \frac{\binom{N_{A_1}}{\frac{n+x}{2}} \binom{N - N_{A_1}}{\frac{n-x}{2}}}{\binom{N}{n}}.$$

---

<sup>1</sup>Ensaio repetido e independente são chamados de *Ensaio de Bernoulli* se existem apenas dois possíveis resultados para cada ensaio e as probabilidades se mantêm inalteradas em todos os ensaios. É comum denotar as probabilidades por  $p$  e  $1 - p$ , e referir-se ao resultado com probabilidade  $p$  como *sucesso* e ao outro como *fracasso*. (Feller [1968])

Não encontrou-se um pivô que levasse a uma hipergeométrica que não dependesse de  $A_1$  e  $A_2$ . E mesmo que fosse encontrado, existiria o problema de calcular as probabilidades quando  $N$  e  $n$  fossem muito grandes. Inicialmente aproximou-se a hipergeométrica pela binomial, de acordo com o resultado do Teorema 3.

**Teorema 3** (Aproximação da Hipergeométrica pela Binomial). *Uma população de  $N$  elementos é dividida em elementos do tipo  $A_1$  e  $A_2$  na proporção  $p_1 : p_2$ , onde  $p_1 + p_2 = 1$ . Uma amostra de tamanho  $n$  é retirada sem reposição. A probabilidade de que ela contenha exatamente  $n_{A_1}$  elementos do tipo  $A_1$  é dada por uma distribuição hipergeométrica, definida em 3.6. Quando  $N \rightarrow \infty$  esta probabilidade se aproxima de  $B(n_{A_1}; n, p)$ .*

*Demonstração.* Vide Irony and Pereira [1994], página 192. □

**Teorema 4** (Teorema Central do Limite). *Seja  $\{X_k\}$  uma sequência de variáveis aleatórias mutuamente independentes com uma distribuição em comum. Suponha que  $\mu = E(X_k)$  e  $\sigma^2 = \text{Var}(X_k)$  existam e seja  $S_n = X_1 + \dots + X_n$ . Então, para cada  $\beta$  fixado,*

$$P\left(\frac{S_n - n\mu}{\sigma \sqrt{n}} < \beta\right) \rightarrow \Phi(\beta), \quad (3.8)$$

onde  $\Phi(\beta)$  é a distribuição normal padrão definida em 2.8 (Feller [1968]).

*Demonstração.* Vide Feller [1968], página 244. □

**Corolário 1** (Aproximação da Binomial pela Normal). *Seja  $\{X_k\}$  uma sequência de variáveis aleatórias mutuamente independentes com distribuição binomial de parâmetros  $n$  e  $p$ . Seja  $S_n = X_1 + \dots + X_n$ . Então, para cada  $\beta$  fixado,*

$$P\left(\frac{S_n - np}{\sqrt{np(1-p)}} < \beta\right) \rightarrow \Phi(\beta), \quad (3.9)$$

onde  $\Phi(\beta)$  é a distribuição normal padrão definida em 2.8 (Feller [1968]).

Deseja-se encontrar um intervalo de confiança  $1 - \alpha$  para a diferença das proporções populacionais  $p_{A_1} - p_{A_2}$ .

$$P\left(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\therefore P\left(z_{\frac{\alpha}{2}} < \frac{n_{A_1} - n \frac{N_{A_1}}{N}}{\sqrt{\left(\frac{N-n}{N-1}\right) \frac{n N_{A_1} N_{A_2}}{N^2}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\therefore P\left(\frac{N}{n}\left[n_{A_1} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{N-n}{N-1}\frac{nN_{A_1}N_{A_2}}{N^2}}\right] < N_{A_1} < \frac{N}{n}\left[n_{A_1} - z_{\frac{\alpha}{2}}\sqrt{\frac{N-n}{N-1}\frac{nN_{A_1}N_{A_2}}{N^2}}\right]\right) = 1 - \alpha$$

Este é um intervalo de confiança  $1 - \alpha$  para  $N_{A_1}$ , o total de votos para o candidato  $A_1$  na população. Para obtermos um IC  $1 - \alpha$  para  $N_{A_1} - N_{A_2}$ , temos que

$$N_{A_1} - N_{A_2} = N_{A_1} - (N - N_{A_1}) = 2N_{A_1} - N,$$

ou seja, a menos das constantes 2 e  $N$ , um IC  $1 - \alpha$  para a diferença dos totais de votos  $N_{A_1} - N_{A_2}$  pode ser obtido fazendo-se

$$P\left(\frac{2N}{n}\left[n_{A_1} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{N-n}{N-1}\frac{nN_{A_1}N_{A_2}}{N^2}}\right] - N < 2N_{A_1} - N < \frac{2N}{n}\left[n_{A_1} - z_{\frac{\alpha}{2}}\sqrt{\frac{N-n}{N-1}\frac{nN_{A_1}N_{A_2}}{N^2}}\right] - N\right) = 1 - \alpha. \quad (3.10)$$

Finalmente, um IC  $1 - \alpha$  para a diferença das proporções de votos para  $A_1$  e  $A_2$  pode ser obtido dividindo-se o intervalo do qual calcula-se a probabilidade em 3.10 por  $N$ , resultando em

$$P\left(\frac{2}{n}\left[n_{A_1} - z_{1-\frac{\alpha}{2}}\sqrt{\text{Var}(n_{A_1})}\right] - 1 < 2p_{A_1} - 1 < \frac{2}{n}\left[n_{A_1} - z_{\frac{\alpha}{2}}\sqrt{\text{Var}(n_{A_1})}\right] - 1\right) = 1 - \alpha,$$

onde  $\text{Var}(n_{A_1}) = \frac{N-n}{N-1}np_{A_1}p_{A_2}$  e  $p_{A_1} = \frac{N_{A_1}}{N}$ . Estimando-se  $p_{A_1}$  e  $p_{A_2}$  respectivamente por

$$\hat{p}_{A_1} = \frac{n_{A_1}}{n} \text{ e } \hat{p}_{A_2} = \frac{n_{A_2}}{n},$$

temos um IC  $1 - \alpha$  estimado para  $p_{A_1} - p_{A_2}$  tal que

$$IC(p_{A_1} - p_{A_2}, 1 - \alpha) = \frac{2}{n}\left[n\hat{p}_{A_1} \mp z_{1-\frac{\alpha}{2}}\sqrt{\hat{\text{Var}}(n_{A_1})}\right] - 1 \quad (3.11)$$

onde  $\hat{\text{Var}}(n_{A_1}) = \frac{N-n}{N-1}n\hat{p}_{A_1}\hat{p}_{A_2}$ .

Porém, para  $N$  suficientemente grande,  $\frac{N-n}{N-1}$  é muito próximo de 1, o que permite aproximar a variância estimada de  $n_{A_1}$  por  $\hat{\text{Var}}(n_{A_1}) = n\hat{p}_{A_1}\hat{p}_{A_2}$ .

**Exemplo 6.** Considere os dados do Exemplo 4 e suponha que  $N = 10000$ . Neste caso

$$\hat{p}_{A_1} = \frac{217}{217 + 181} = \frac{217}{398} \text{ e } \hat{p}_{A_2} = \frac{181}{217 + 181} = \frac{181}{398}.$$

Temos ainda que as variâncias, respectivamente com e sem o fator de correção, são

$$\hat{Var}(n_{A_1}) = \frac{10000 - 398}{10000 - 1} \times 398 \times \frac{217}{398} \times \frac{181}{398} \doteq 94.77$$

$$\hat{Var}(n_{A_1}) = 398 \times \frac{217}{398} \times \frac{181}{398} \doteq 98.69.$$

Portanto, um IC  $1 - 0.05 = 0.95 = 95\%$  para  $p_{A_1} - p_{A_2}$  com o fator de correção é tal que

$$IC(p_{A_1} - p_{A_2}, 95\%) = \left( \frac{2}{398} \left[ 217 - 1.96 \sqrt{94.77} \right] - 1, \frac{2}{398} \left[ 217 + 1.96 \sqrt{94.77} \right] - 1 \right) \doteq (-0.0054, 0.1863).$$

Por sua vez, um IC  $1 - 0.05 = 0.95 = 95\%$  para  $p_{A_1} - p_{A_2}$  sem o fator de correção é tal que

$$IC(p_{A_1} - p_{A_2}, 95\%) = \left( \frac{2}{398} \left[ 217 - 1.96 \sqrt{98.69} \right] - 1, \frac{2}{398} \left[ 217 + 1.96 \sqrt{98.69} \right] - 1 \right) \doteq (-0.0074, 0.1883).$$

Como o zero pertence aos intervalos calculados, conclui-se que os candidatos  $A_1$  e  $A_2$  estão tecnicamente empatados. Nas Figuras 3.3 e 3.4 estão apresentados os gráficos desta solução.

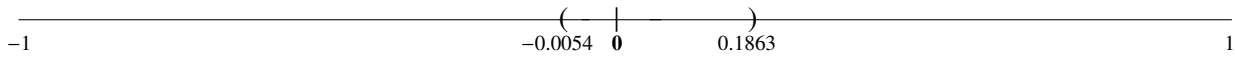


Figura 3.3: IC 95% para a diferença de proporção de votos para os candidatos  $A_1$  e  $A_2$ , com a correção  $\frac{N-n}{N-1}$ .

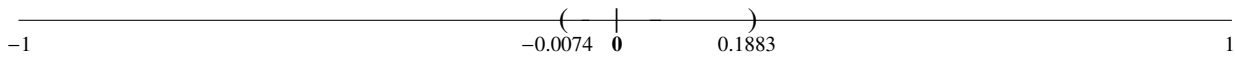


Figura 3.4: IC 95% para a diferença de proporção de votos para os candidatos  $A_1$  e  $A_2$ , sem a correção  $\frac{N-n}{N-1}$ .

☺

Nota-se que a diferença de amplitude entre os dois intervalos é de apenas 0.004. Esta diferença diminui à medida que aumenta o tamanho da população, que normalmente é maior que os 10 mil do exemplo.

### 3.3.4 k Candidatos

**Definição 5** (Hipergeométrica Generalizada). *Seja uma população com  $N$  elementos, sendo  $N_{A_1}$  do primeiro tipo,  $N_{A_2}$  do segundo tipo, ...,  $N_{A_k}$  do  $k$ -ésimo tipo. Um grupo de  $n$  elementos é escolhido aleatoriamente. A*

probabilidade  $HG(n_{A_1}, \dots, n_{A_k}; n, N_{A_1}, \dots, N_{A_k})$  de que o grupo sorteado tenha exatamente  $n_{A_1}$  elementos do tipo  $A_1, \dots, n_{A_k}$  elementos do tipo  $A_k$  é

$$HG(n_{A_1}, \dots, n_{A_k}; n, N_{A_1}, \dots, N_{A_k}) = \frac{\binom{N_{A_1}}{n_{A_1}} \dots \binom{N_{A_k}}{n_{A_k}}}{\binom{N}{n}}. \quad (3.12)$$

O sistema de probabilidades definido desta forma é chamado distribuição hipergeométrica generalizada (Feller [1968]).

**Definição 6** (Multinomial). Seja  $Mult(n_{A_1}, \dots, n_{A_k}; n, p_1, \dots, p_k)$  a probabilidade de saírem exatamente  $n_{A_1}$  observações do evento  $E_1$ ,  $n_{A_2}$  observações do evento  $E_2$ , ...,  $n_{A_k}$  observações do evento  $E_k$  em  $n$  ensaios com probabilidade  $p_1$  para a realização do evento  $E_1$ ,  $p_2$  para a realização do evento  $E_2$ , ...,  $p_k$  para a realização do evento  $E_k$ . Para  $k = 2$  temos ensaios de Bernoulli. Então,

$$Mult(n_{A_1}, \dots, n_{A_k}; n, p_1, \dots, p_k) = \frac{n!}{n_{A_1}! n_{A_2}! \dots n_{A_k}!} p_1^{n_{A_1}} p_2^{n_{A_2}} \dots p_k^{n_{A_k}}, \quad (3.13)$$

com  $p_1 + \dots + p_k = 1$ ,  $p_i \geq 0$  e  $n_{A_1} + \dots + n_{A_k} = n$ .

O sistema de probabilidades definido desta forma é chamado distribuição multinomial (Feller [1968]).

**Teorema 5** (Aproximação da Hipergeométrica Generalizada pela Multinomial). Uma população de  $N$  elementos é dividida em elementos do tipo  $A_1, \dots, A_k$  na proporção  $p_1 : p_2 : \dots : p_k$ , onde  $p_1 + \dots + p_k = 1$ . Uma amostra de tamanho  $n$  é retirada sem reposição. A probabilidade de que ela contenha exatamente  $n_{A_1}$  elementos do tipo  $A_1$ , ...,  $n_{A_k}$  elementos do tipo  $A_k$  é dada por uma distribuição hipergeométrica generalizada, definida em 3.12. Quando  $N \rightarrow \infty$ , esta probabilidade se aproxima de  $Mult(n_{A_1}, \dots, n_{A_k}; n, p_1, \dots, p_k)$ .

*Demonstração.* Vide Johnson et al. [1997], página 172. □

**Teorema 6** (Teorema Central do Limite Multivariado). Se  $\bar{\mathbf{y}}$  é o vetor de médias de uma amostra aleatória  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$  de uma população com vetor de médias  $\mu$  e matriz de covariâncias  $\Sigma$ , então quando  $n \rightarrow \infty$ , a distribuição de  $\sqrt{n}(\bar{\mathbf{y}} - \mu)$  aproxima-se de  $N_p(\mathbf{0}, \Sigma)$  (Rencher [2002]).

**Corolário 2** (Aproximação da Multinomial pela Normal Multivariada). Se  $\bar{\mathbf{y}}$  é o vetor de médias de uma amostra aleatória  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$  de uma população multinomial definida em 3.13, então quando  $n \rightarrow \infty$ , a distribuição de  $\sqrt{n}(\bar{\mathbf{y}} - \mu)$  aproxima-se de  $N_p(\mathbf{0}, \Sigma)$ .

Aproximando-se a distribuição do vetor  $\hat{p}_1, \dots, \hat{p}_{1-k}$  por uma normal multivariada, pode-se calcular um intervalo ou região de confiança  $1 - \alpha$ . Os intervalos são utilizados no caso em que só temos um parâmetro sendo estimado (caso uniparamétrico), e as regiões quando temos dois ou mais parâmetros (caso multiparamétrico).



Para a interpretação destes intervalos (ou regiões) não pode-se falar em probabilidade de a estimativa estar contida no intervalo (ou região) calculado, visto que a Inferência Frequentista não admite tratar parâmetros populacionais como variáveis aleatórias. Assim, a interpretação mais conhecida é que *se calculássemos  $n$  intervalos (ou regiões) de confiança,  $(1-\alpha)\times 100\%$  conteriam o verdadeiro parâmetro (ou vetor de parâmetros)*.

Devido à restrição  $\sum_{i=1}^k p_{A_i} = 1$ , o espaço paramétrico é um *simplex*, i.e., um hiperplano de dimensão menor que o espaço no qual está contido. No caso de três candidatos, podemos visualizar este hiperplano como um triângulo equilátero em um espaço cartesiano de 3 dimensões, conforme as Figuras 4.3 e 4.4. Na Figura 3.5 estão apontadas as regiões do simplex onde ocorrem os possíveis eventos em um cenário eleitoral com três candidatos.

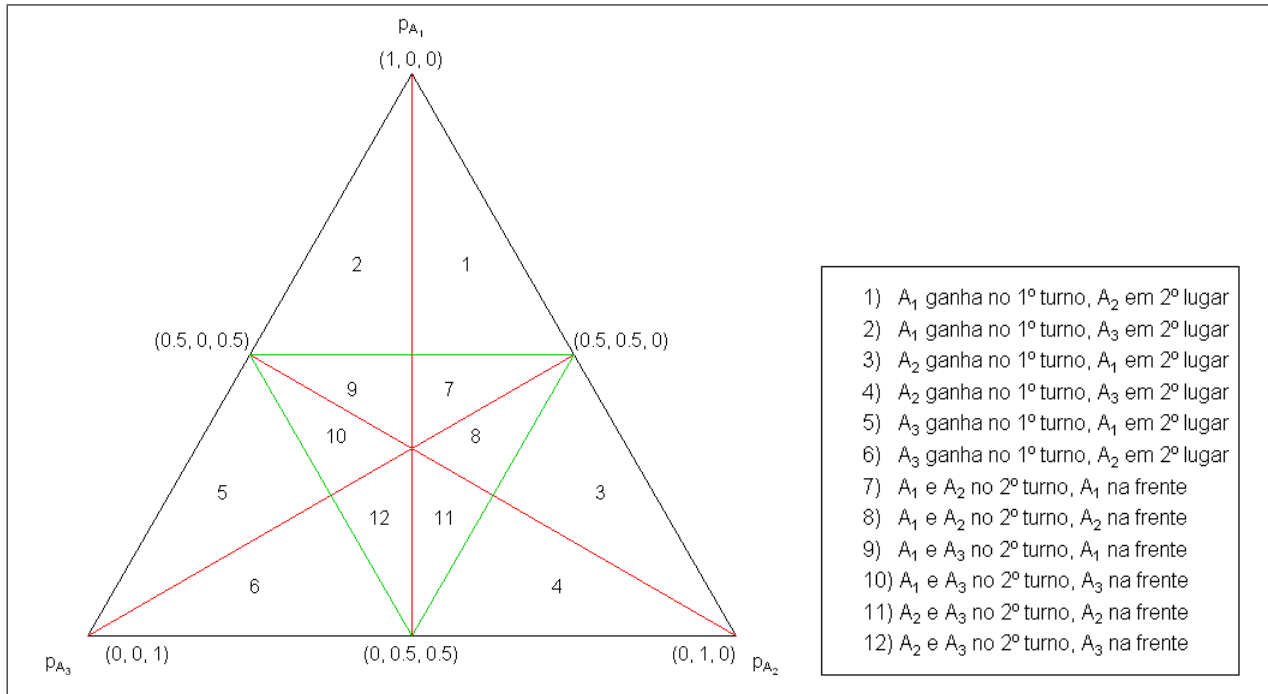


Figura 3.5: Simplex gerado pela restrição  $p_{A_1} + p_{A_2} + p_{A_3} = 1$ , com regiões representando os possíveis eventos em um cenário eleitoral com três candidatos.

### 3.4 Bayesiana

#### 3.4.1 $k$ Candidatos

O problema de estimar proporções é resolvido de forma direta na abordagem Bayesiana, pois é geral para qualquer número de candidatos. Como foi apresentado no Teorema 2, é possível pensar neste problema assumindo-se, por exemplo, uma *priori Dirichlet*(1, 1, ..., 1) de dimensão  $k$ , que com seu núcleo multiplicado pelo núcleo da verossimilhança multinomial de mesma dimensão resulta em uma *posteriori Dirichlet*(1 +

$n_{A_1}, 1 + n_{A_2}, \dots, 1 + n_{A_k}$ ), onde  $n_{A_i}$  é o número de votos atribuídos, após a distribuição dos indecisos, ao  $i$ -ésimo candidato na amostra ( $i = 1, \dots, k$ ).

Com a posteriori conjunta pode-se calcular, por exemplo, a probabilidade de o candidato  $A_1$  vencer a eleição no primeiro turno, i.e., ter a maioria absoluta dos votos válidos na amostra. Para resolver esta integral de  $k$  dimensões, utiliza-se algum recurso numérico. Neste trabalho utilizou-se simulação de Monte Carlo Ordinário para aproximar a integral múltipla da *posteriori* Dirichlet, cujos passos são descritos abaixo.

1. Defina uma priori Dirichlet para o vetor multinomial de votos para os  $k$  candidatos;
2. Gere uma quantidade razoável (na casa das dezenas de milhares) de posteriors Dirichlet de parâmetros (priori para o candidato  $A_1 + n_{A_1}$ , priori para o candidato  $A_2 + n_{A_2}, \dots$ , priori para o candidato  $A_k + n_{A_k}$ ).
3. Some o número de vezes em que a situação que você tem interesse aparece e divida pelo total de Dirichlets geradas.

## Capítulo 4

### Comparações

Neste capítulo são feitas comparações entre as três metodologias descritas no início do Capítulo 3.

Toda a parte computacional e gráfica foi desenvolvida no *software R 2.10.0* (R Development Core Team [2009]), utilizando-se as bibliotecas Murdoch et al. [2007], Adler and Murdoch [2009], Weihs et al. [2005], Yee [2009], Genz et al. [2009]. Para possibilitar a reprodução exata dos resultados pseudo-aleatórios obtidos neste trabalho, definiu-se uma semente através do comando `set.seed(1)`.

**Definição 7.** *Iminência do empate técnico ocorre quando  $k$  candidatos estão empatados tecnicamente por  $k - 1$  pontos segundo a teoria frequentista dos institutos.*

#### 4.1 Dois Candidatos

Nesta seção será apresentada a análise do cenário com dois candidatos, comparando as três metodologias citadas no Capítulo 3. Um teorema que faz a ligação de probabilidades a posteriori com o nível de significância do intervalo de confiança frequentista é apresentado.

Se desejarmos obter o valor de  $\hat{p}_{A_1}$ , para  $n$ ,  $\alpha$  e  $z = z_{1-\frac{\alpha}{2}}$  fixados, tal que os intervalos para  $p_{A_1}$  e  $p_{A_2}$  (dados respectivamente por 3.2 e 3.3) estejam na iminência do empate técnico, devemos ter o intervalo para  $p_{A_1}$  à esquerda do intervalo de  $p_{A_2}$  ou vice-versa, com intersecção apenas nos extremos, ou seja,  $\hat{p}_{A_2} - \varepsilon = \hat{p}_{A_1} + \varepsilon$  ou  $\hat{p}_{A_1} - \varepsilon = \hat{p}_{A_2} + \varepsilon$ , onde  $\varepsilon = z \sqrt{\hat{p}_{A_1} (1 - \hat{p}_{A_1}) / n}$ . Assim temos

Assim temos

$$\begin{aligned}
 \hat{p}_{A_1} - \hat{p}_{A_2} &= 2\varepsilon \\
 \hat{p}_{A_1} - (1 - \hat{p}_{A_1}) &= 2z \sqrt{\frac{\hat{p}_{A_1}(1 - \hat{p}_{A_1})}{n}} \\
 (2\hat{p}_{A_1} - 1)^2 &= \left( 2z \sqrt{\frac{\hat{p}_{A_1}(1 - \hat{p}_{A_1})}{n}} \right)^2 \\
 4n\hat{p}_{A_1}^2 - 4n\hat{p}_{A_1} + n &= 4z^2\hat{p}_{A_1} - 4z^2\hat{p}_{A_1}^2 \\
 4(n + z^2)\hat{p}_{A_1}^2 - 4(n + z^2)\hat{p}_{A_1} + n &= 0
 \end{aligned} \tag{4.1}$$

Por outro lado, se deseja-se obter o valor da proporção amostral de votos  $\hat{p}_{A_1}$  do candidato  $A_1$ , para  $n$ ,  $\alpha$  e  $z = z_{1-\frac{\alpha}{2}}$  fixados, tal que o intervalo para a diferença  $p_{A_1} - p_{A_2}$  (dado por 3.11) esteja na iminência do empate técnico, devemos ter um dos extremos deste intervalo igual a zero. Assim

$$\begin{aligned}
 \frac{2}{n} \left( n\hat{p}_{A_1} - z \sqrt{n\hat{p}_{A_1}(1 - \hat{p}_{A_1})} \right) - 1 &= 0 \\
 (2\hat{p}_{A_1} - 1)^2 &= \left( \frac{2z}{n} \sqrt{n\hat{p}_{A_1}(1 - \hat{p}_{A_1})} \right)^2 \\
 4\hat{p}_{A_1}^2 - 4\hat{p}_{A_1} + 1 &= \frac{4z^2}{n} (\hat{p}_{A_1} - \hat{p}_{A_1}^2) \\
 4\hat{p}_{A_1}^2 - 4\hat{p}_{A_1} + 1 &= \frac{4z^2}{n} \hat{p}_{A_1} - \frac{4z^2}{n} \hat{p}_{A_1}^2 \\
 4 \left( 1 + \frac{z^2}{n} \right) \hat{p}_{A_1}^2 - 4 \left( 1 + \frac{z^2}{n} \right) \hat{p}_{A_1} + 1 &= 0 \\
 4(n + z^2)\hat{p}_{A_1}^2 - 4(n + z^2)\hat{p}_{A_1} + n &= 0
 \end{aligned} \tag{4.2}$$

A solução da equação quadrática 4.2 para  $\hat{p}_{A_1}$  nos retorna duas raízes reais entre 0 e 1, sendo que  $\hat{p}'_{A_1} = 1 - \hat{p}''_{A_1}$ , bastando utilizar a maior das raízes. As soluções das equações 4.1 e 4.2 são equivalentes, tornando as conclusões análogas.

Na Figura 4.1 estão os gráficos das proporções amostrais do candidato  $A_1$  *versus* suas probabilidades de vitória no primeiro turno para  $\alpha = 0.05$  e  $n = 50, 100, 500, 1000$ . O par ordenado apresentado em destaque apresenta na coordenada  $x$  o valor que soluciona 4.1 e 4.2 com sua respectiva probabilidade *a posteriori* ( $pp$ ) na coordenada  $y$ .

Percebe-se claramente que à medida que  $n$  aumenta, as probabilidades de  $\hat{p}_{A_1}$  que definem o intervalo com empate técnico aproximam-se de 0.5, o que é bastante razoável. Outro fato interessante é que a probabilidade

*a posteriori* do ponto que leva os frequentistas a declararem empate técnico parece estar convergindo para o valor 0.975, ou  $1 - 0.05/2$ .

Para reforçar a ideia, estão apresentados na Figura 4.2 os mesmos gráficos da Figura 4.1, mas com  $n = 100,000$  e  $\alpha = 0.01, 0.05, 0.10, 0.20$ . É fácil perceber que para  $\alpha = 0.01$ ,  $pp \doteq 0.995$ , para  $\alpha = 0.05$ ,  $pp \doteq 0.975$ , para  $\alpha = 0.10$ ,  $pp \doteq 0.950$  e para  $\alpha = 0.20$ ,  $pp \doteq 0.900$ .

A intuição por trás destes gráficos é que enquanto os

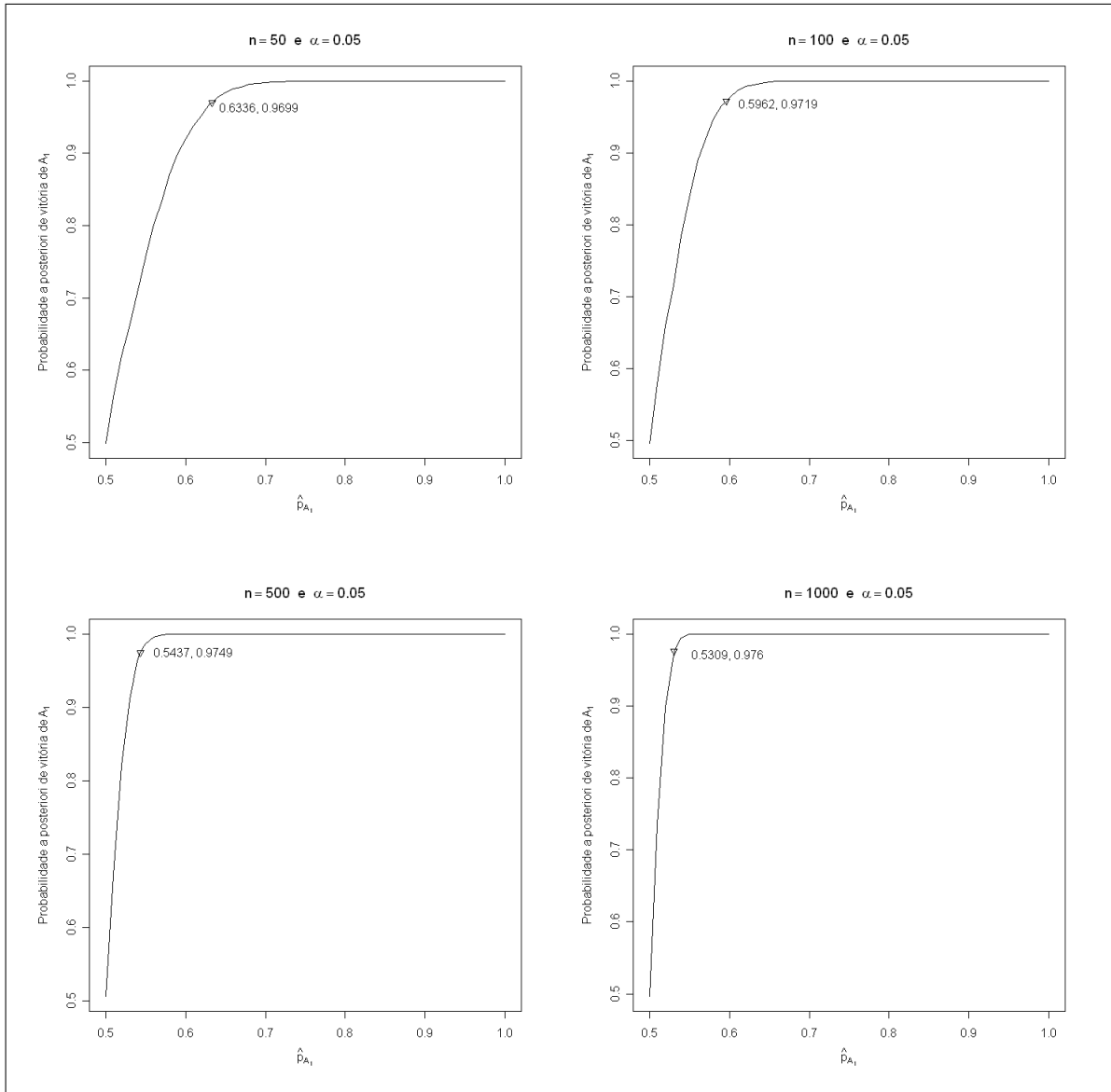


Figura 4.1: Proporções  $\hat{p}_{A_1}$  vs probabilidades *a posteriori* para  $\alpha = 0.05$  e  $n = 50, 100, 500, 1000$ .

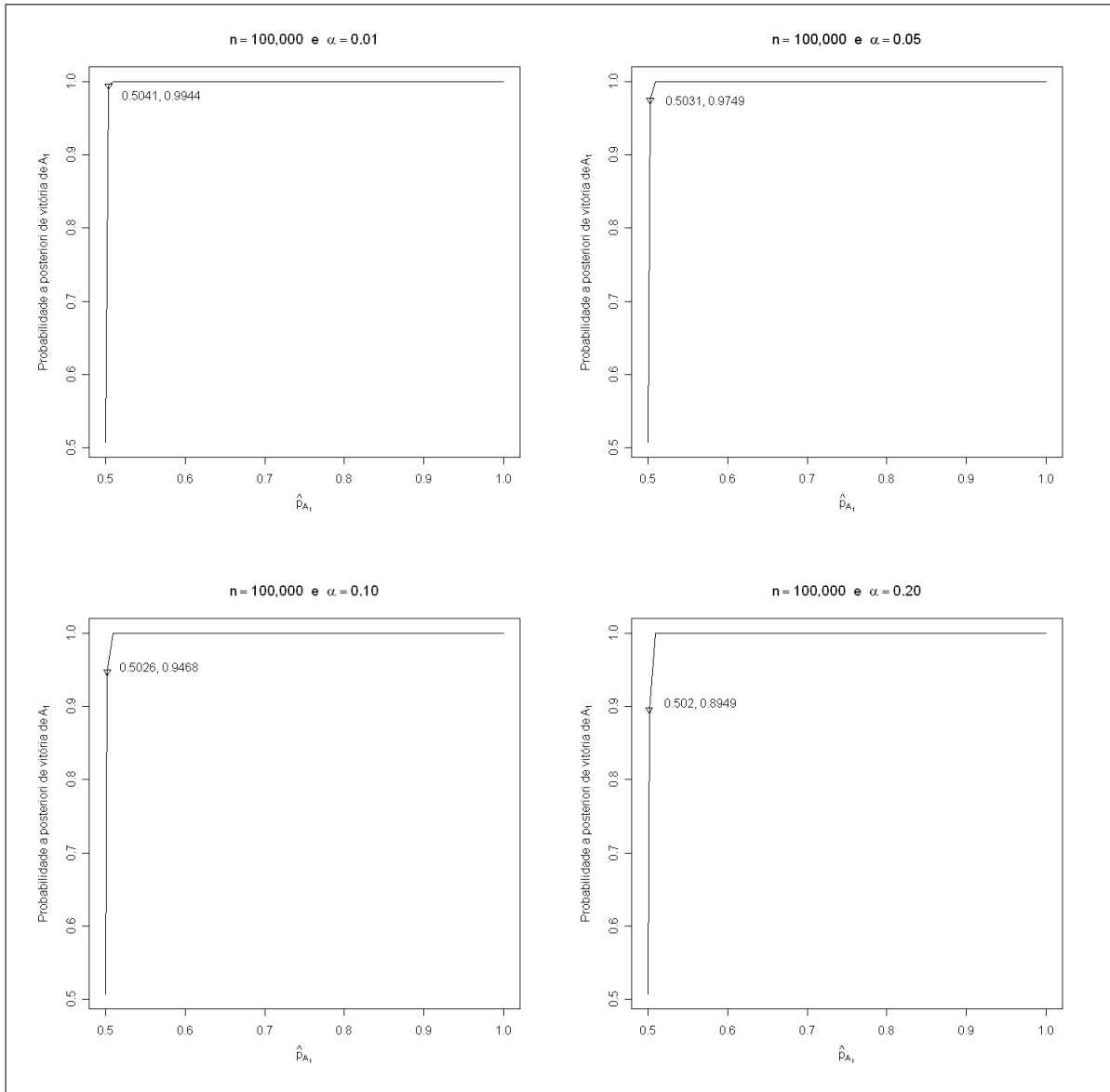


Figura 4.2: Proporções  $\hat{p}_{A_1}$  vs probabilidades *a posteriori* para  $n = 100,000$  e  $\alpha = 0.01, 0.05, 0.10, 0.20$ .

**Teorema 7.** *Sejam as sequências de variáveis aleatórias*

$$a_n = a_0 + \frac{n}{2} + \frac{nz\sqrt{n+z^2}}{2(n+z^2)}$$

e

$$b_n = b_0 + \frac{n}{2} - \frac{nz\sqrt{n+z^2}}{2(n+z^2)},$$

onde  $a_0$  e  $b_0$  são parâmetros de uma  $Beta(a_0, b_0)$ ,  $n$  é o número de observações e  $z = z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ . Então quando  $n \rightarrow \infty$

$$P\left(\theta > \frac{1}{2} \middle| f(n, z) = \frac{n}{2} + \frac{nz\sqrt{n+z^2}}{2(n+z^2)}\right) \rightarrow 1 - \frac{\alpha}{2},$$

onde  $\theta$  é a proporção populacional de alguma característica de interesse.

*Demonstração.* Seja  $\bar{X}_n$  a proporção amostral de alguma característica de interesse em uma amostra de tamanho  $n$  e  $\theta$  a proporção populacional da mesma característica. Pelo Teorema Central do Limite,

$$\frac{\bar{X}_n - \theta}{\sqrt{\theta(1-\theta)/n}} \xrightarrow{D} N(0, 1).$$

Como  $\bar{X}_n \xrightarrow{P} \theta$ , temos que

$$\bar{X}_n(1 - \bar{X}_n) \xrightarrow{P} \theta(1 - \theta).$$

Logo,

$$\frac{\sqrt{\theta(1-\theta)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{P} 1.$$

Assim,

$$\frac{\sqrt{\theta(1-\theta)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \frac{\bar{X}_n - \theta}{\sqrt{\theta(1-\theta)/n}} = \frac{\bar{X}_n - \theta}{\sqrt{\bar{X}_n(1-\bar{X}_n)/n}} \xrightarrow{D} N(0, 1).$$

Logo,

$$\begin{aligned} P\left(z_{\frac{\alpha}{2}} < \frac{\bar{X}_n - \theta}{\sqrt{\bar{X}_n(1-\bar{X}_n)/n}} < z_{1-\frac{\alpha}{2}}\right) &\xrightarrow{n \rightarrow \infty} 1 - \alpha \implies \\ \implies P\left(\bar{X}_n - z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} < \theta < \bar{X}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right) &\xrightarrow{n \rightarrow \infty} 1 - \alpha. \end{aligned} \quad (4.3)$$

O caso de “empate técnico” ou “nada pode-se afirmar sobre o segundo turno” ocorre quando o limite inferior do intervalo de confiança (aproximado) apresentando em (4.3) é igual a  $1/2$ , ou seja,

$$\bar{X} - z \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}} = \frac{1}{2},$$

onde  $\bar{X} = \bar{X}_n$  e  $z = z_{\frac{\alpha}{2}}$  por simplicidade. Calculando  $\bar{X}$  temos

$$\begin{aligned} \left(\bar{X} - \frac{1}{2}\right)^2 &= \left(z \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}\right)^2 \implies \bar{X}^2 \left(1 + \frac{z^2}{n}\right) - \bar{X} \left(1 + \frac{z^2}{n}\right) + \frac{1}{4} = 0 \implies \\ \implies \bar{X} &= \frac{1}{2} + \frac{z \sqrt{n + z^2}}{2(n + z^2)} \\ \implies f(n, z) &= \frac{n}{2} + \frac{nz \sqrt{n + z^2}}{2(n + z^2)}. \end{aligned}$$

Suponha a priori que  $\theta \sim \text{Beta}(a_0, b_0)$ . Assim,

$$\theta \left| f(n, z) = \frac{n}{2} + \frac{nz \sqrt{n + z^2}}{2(n + z^2)} \right. \sim \text{Beta}(a_0 + f(n, z), b_0 + n - f(n, z)).$$

Queremos calcular

$$P\left(\theta > \frac{1}{2} \left| f(n, z) = \frac{n}{2} + \frac{nz \sqrt{n + z^2}}{2(n + z^2)} \right.\right).$$

Sejam

$$a_n = a_0 + \frac{n}{2} + \frac{nz \sqrt{n + z^2}}{2(n + z^2)}$$

e

$$b_n = b_0 + \frac{n}{2} - \frac{nz \sqrt{n + z^2}}{2(n + z^2)}.$$

Sejam ainda  $X_{a_n} \sim \text{Gama}(a_n, 1)$  e  $Y_{b_n} \sim \text{Gama}(b_n, 1)$ , com  $E(X_{a_n}) = a_n/1 = a_n$ ,  $\text{var}(X_1 + \dots + X_{a_n}) = a_n/1^2 = a_n$  e  $E(Y_1 + \dots + Y_{b_n}) = b_n/1 = b_n$ ,  $\text{var}(Y_1 + \dots + Y_{b_n}) = b_n/1^2 = b_n$ . Logo,

$$\frac{X_1 + \dots + X_{a_n}}{X_1 + \dots + X_{a_n} + Y_1 + \dots + Y_{b_n}} \left| f(n, z) = \frac{n}{2} + \frac{nz \sqrt{n + z^2}}{2(n + z^2)} \right. \sim \text{Beta}(a_0 + f(n, z), b_0 + n - f(n, z)).$$



Desta forma pode-se escrever

$$\begin{aligned} P\left(\theta > \frac{1}{2} \middle| f(n, z) = \frac{n}{2} + \frac{nz\sqrt{n+z^2}}{2(n+z^2)}\right) &= P\left(\frac{X_1 + \cdots + X_{a_n}}{X_1 + \cdots + X_{a_n} + Y_1 + \cdots + Y_{b_n}} > \frac{1}{2} \middle| f(n, z) = \frac{n}{2} + \frac{nz\sqrt{n+z^2}}{2(n+z^2)}\right) \\ &= P\left((X_1 + \cdots + X_{a_n}) - (Y_1 + \cdots + Y_{b_n}) > 0 \middle| f(n, z) = \frac{n}{2} + \frac{nz\sqrt{n+z^2}}{2(n+z^2)}\right). \end{aligned}$$

Pelo Teorema Central do Limite,

$$\frac{X_1 + \cdots + X_{a_n} - a_n}{\sqrt{a_n}} \xrightarrow{D} N(0, 1) \quad (4.4)$$

e

$$\frac{Y_1 + \cdots + Y_{b_n} - b_n}{\sqrt{b_n}} \xrightarrow{D} N(0, 1).$$

Como  $\frac{b_n}{a_n} \rightarrow 1$ , temos

$$\frac{\sqrt{b_n}}{\sqrt{a_n}} \frac{Y_1 + \cdots + Y_{b_n} - b_n}{\sqrt{b_n}} \xrightarrow{D} N(0, 1)$$

ou

$$\frac{Y_1 + \cdots + Y_{b_n} - b_n}{\sqrt{a_n}} \xrightarrow{D} N(0, 1). \quad (4.5)$$

Logo, por (4.4) e (4.5)

$$\frac{[X_1 + \cdots + X_{a_n} - a_n] - [Y_1 + \cdots + Y_{b_n} - b_n]}{\sqrt{a_n}} \xrightarrow{D} N(0, 2).$$

Logo,

$$P\left(\frac{[X_1 + \cdots + X_{a_n} - a_n] - [Y_1 + \cdots + Y_{b_n} - b_n]}{\sqrt{2a_n}} > \frac{b_n - a_n}{\sqrt{2a_n}}\right).$$

Tomando  $n \in \{2, 3, \dots\}$ ,

$$\frac{b_n - a_n}{\sqrt{2a_n}} = \frac{(b_0 - a_0) - \frac{nz\sqrt{n+z^2}}{n+z^2}}{\sqrt{2a_0 + \frac{n}{2} + \frac{nz\sqrt{n+z^2}}{2(n+z^2)}}} = \frac{\frac{(b_0 - a_0)\sqrt{n+z^2}}{n} - z}{\sqrt{\frac{n(n+z^2)}{n^2} + \frac{nz\sqrt{n+z^2}}{n^2} + \frac{2a_0(n+z^2)}{n^2}}} \xrightarrow{n \rightarrow \infty} \frac{0 - z}{\sqrt{1 + 0 + 0}} \rightarrow -z$$

Assim,

$$P\left(\theta > \frac{1}{2} \mid f(n, z) = \frac{n}{2} + \frac{nz\sqrt{n+z^2}}{2(n+z^2)}\right) \xrightarrow{n \rightarrow \infty} P(Z > -z) = 1 - \frac{\alpha}{2}.$$

□

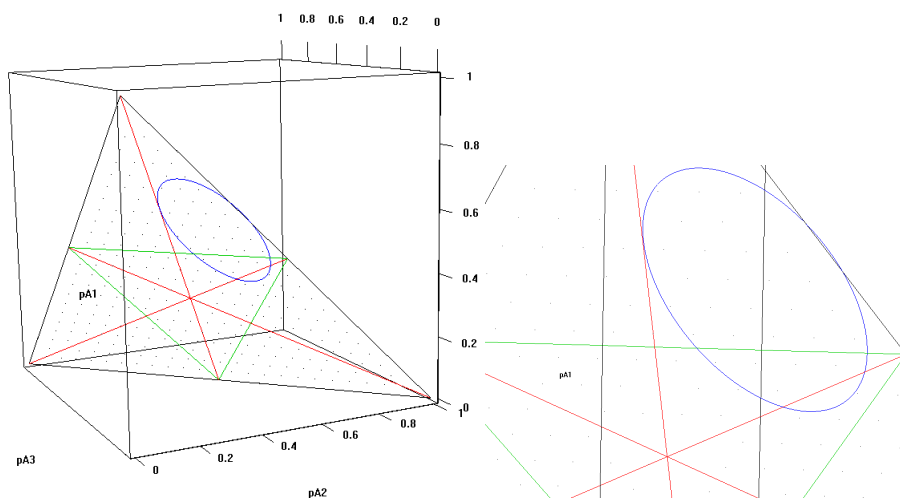
**Corolário 3.** Considere um processo eleitoral brasileiro com dois candidatos,  $A_1$  e  $A_2$ . Seja  $\max(\hat{p}'_{A_1}, \hat{p}''_{A_1}) = \hat{p}_{A_1}$ , onde  $\hat{p}'_{A_1}$  e  $\hat{p}''_{A_1}$  são soluções da equação  $4(n+z^2)\hat{p}_{A_1}^2 - 4(n+z^2)\hat{p}_{A_1} + n = 0$ , tal que  $IC(p_{A_1} - p_{A_2}, 1 - \alpha) = [0, \delta]$ . Quando  $n \rightarrow \infty$  então  $P(p_{A_1} > 1/2 \mid \hat{p}_{A_1}) \rightarrow 1 - \alpha/2$ .

Em outras palavras, este corolário afirma que enquanto os frequentistas estão no limiar entre o empate técnico e a declaração de quem está na frente com base em um intervalo de confiança  $1 - \alpha$ , os bayesianos acreditam que o candidato com mais votos tem probabilidade  $1 - \alpha/2$  de vitória. Por exemplo, o gráfico com  $n = 100,000$  e  $\alpha = 0.05$  na Figura 4.2 indica que se tivermos um cenário com dois candidatos (o segundo turno, por exemplo) com os respectivos  $n$  e  $\alpha$ , e o candidato com mais votos (digamos  $A_1$ ) tiver 50.31% das intenções dos votos, a metodologia frequentista indicará empate técnico, enquanto os bayesianos dirão que  $A_1$  tem aproximadamente  $0.9749 \doteq 1 - 0.05/2$  de probabilidade de vitória.

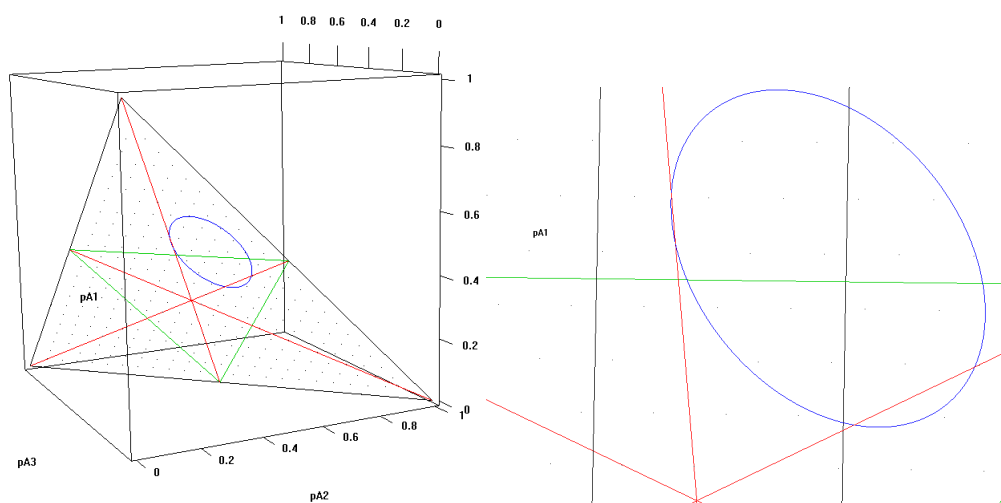
## 4.2 Três Candidatos

Nesta seção será apresentada a análise do cenário com três candidatos, comparando as metodologias citadas no Capítulo 3. Para comparar os três métodos, foi feito o seguinte: encontraram-se as proporções  $\hat{p}_{A_1}$ ,  $\hat{p}_{A_2}$  e  $\hat{p}_{A_3}$  para dados valores de  $n$  e  $\alpha$  que levassem à iminência do empate técnico, i.e., o limite superior do intervalo de confiança para a proporção de votos do candidato com menos votos é igual ao limite inferior do intervalo do candidato que está em segundo lugar na pesquisa; por fim, o limite superior do intervalo do candidato segundo colocado é igual ao limite inferior do intervalo do candidato que está à frente na pesquisa. Com estas proporções foram feitas as análises bayesiana e frequentista correta.

Pela Tabela 4.1 é possível perceber que à medida que aumentamos  $n$  a situação de segundo turno entre  $A_1$  e  $A_2$  fica melhor definida, com probabilidade em torno de 0.99. Nas Figuras 4.3 e 4.4 estão apresentados os gráficos dos elipsóides de confiança obtidos da análise frequentista correta. Note que as áreas das figuras são muito similares às probabilidades descritas na Tabela 4.1, o que pode sugerir uma relação entre as metodologias também no caso de três candidatos.

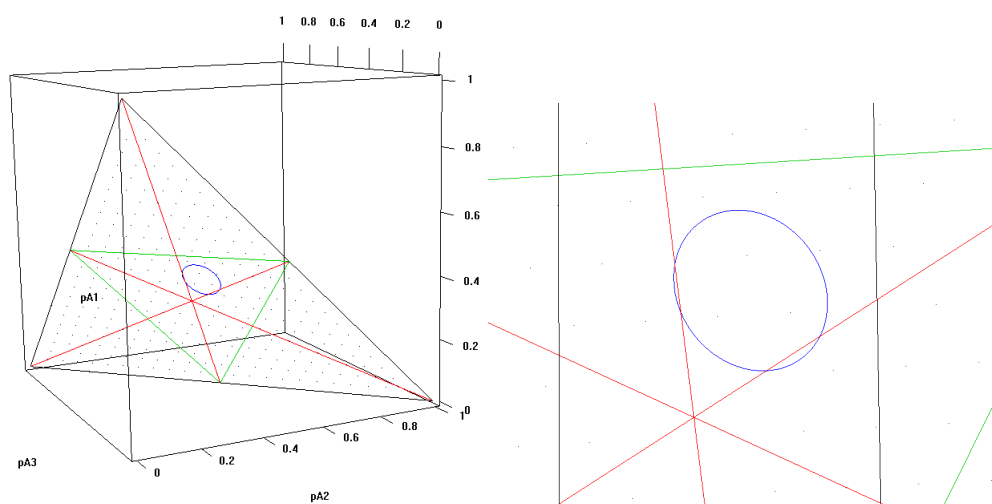


(a)  $n = 50$ ,  $\hat{p}_{A_1} = 0.5813972562$ ,  $\hat{p}_{A_2} = 0.3158114522$  e  $\hat{p}_{A_3} = 0.1027912917$

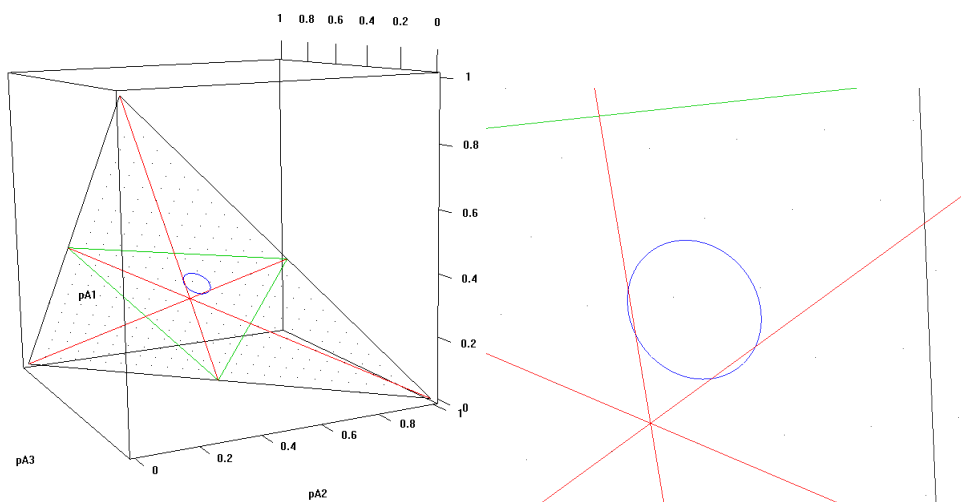


(b)  $n = 100$ ,  $\hat{p}_{A_1} = 0.5144202347$ ,  $\hat{p}_{A_2} = 0.3246860305$  e  $\hat{p}_{A_3} = 0.1608937348$

Figura 4.3: Elipses de confiança 95% com  $n = 50$  e  $n = 100$ .



(a)  $n = 500$ ,  $\hat{p}_{A_1} = 0.4160925601$ ,  $\hat{p}_{A_2} = 0.3316216347$ ,  $\hat{p}_{A_3} = 0.2522858052$



(b)  $n = 1000$ ,  $\hat{p}_{A_1} = 0.3919345050$ ,  $\hat{p}_{A_2} = 0.3324785813$ ,  $\hat{p}_{A_3} = 0.2755869137$

Figura 4.4: Elipses de confiança 95% com  $n = 500$  e  $n = 1000$ .

Situação	Probabilidades a posteriori			
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
$A_1$ vence	0.8327	0.5644	0.0001	0.0000
$A_2$ vence	0.0034	0.0001	0.0000	0.0000
$A_3$ vence	0.0000	0.0000	0.0000	0.0000
$A_1$ e $A_2$ no 2º turno	0.1419	0.4127	0.9780	0.9771
$A_2$ e $A_1$ no 2º turno	0.0200	0.0191	0.0129	0.0128
$A_1$ e $A_3$ no 2º turno	0.0019	0.0037	0.0090	0.0101
$A_3$ e $A_1$ no 2º turno	0.0001	0.0000	0.0000	0.0000
$A_2$ e $A_3$ no 2º turno	0.0000	0.0000	0.0000	0.0000
$A_3$ e $A_2$ no 2º turno	0.0000	0.0000	0.0000	0.0000

Tabela 4.1: Possibilidades em um cenário com três candidatos e as probabilidades *a posteriori* de cada resultado.



## Capítulo 5

### Conclusões

#### 5.1 Considerações Finais

Foi apresentado um comparativo entre as metodologias bayesiana, frequentista correta e a utilizada pelos institutos. No caso de dois candidatos, quando ocorre a iminência do empate técnico definida por 4.1, provou-se que a probabilidade de vitória do candidato com mais votos converge para  $1 - \alpha/2$ , onde  $\alpha$  é a confiança do intervalo frequentista dos institutos ou da abordagem frequentista correta. Isso significa que enquanto os frequentistas declaram empate técnico definido pelo Teorema 7, à medida que o tamanho da amostra cresce os bayesianos acreditam em uma probabilidade de vitória de  $1 - \alpha/2$ .

No caso de três candidatos, fica evidente a diferença das abordagens frequentistas apontadas neste trabalho, uma vez que a própria definição de empate técnico torna-se nebulosa neste cenário. A extensão para quatro ou mais concorrentes é direta.

#### 5.2 Sugestões para Pesquisas Futuras

Sugere-se um estudo mais detalhado para três ou mais candidatos na tentativa de encontrar uma relação mais geral do que a apresentada neste trabalho, ligando elipsóides de confiança – baseados na frequentista correta – às probabilidades à posteriori. Ao observar os gráficos no cenário com três candidatos, nota-se uma proximidade muito grande das áreas dos elipsóides e das probabilidades à posteriori. É importante ressaltar que no caso clássico não há interpretação probabilística, ao contrário do bayesiano.





## **Apêndice A**

### **Artigos 33, 34 e 35 da Lei Nº. 9504/97**

#### **Presidência da República**

##### **Casa Civil**

##### **Subchefia para Assuntos Jurídicos**

LEI Nº 9.504, DE 30 DE SETEMBRO DE 1997. <sup>1</sup>

Estabelece normas para as eleições.

O VICE PRESIDENTE DA REPÚBLICA no exercício do cargo de PRESIDENTE DA REPÚBLICA  
Faço saber que o Congresso Nacional decreta e eu sanciono a seguinte Lei:

##### **Das Pesquisas e Testes Pré-Eleitorais**

Art. 33. As entidades e empresas que realizarem pesquisas de opinião pública relativas às eleições ou aos candidatos, para conhecimento público, são obrigadas, para cada pesquisa, a registrar, junto à Justiça Eleitoral, até cinco dias antes da divulgação, as seguintes informações:

- I - quem contratou a pesquisa;
- II - valor e origem dos recursos despendidos no trabalho;
- III - metodologia e período de realização da pesquisa;
- IV - plano amostral e ponderação quanto a sexo, idade, grau de instrução, nível econômico e área física de realização do trabalho, intervalo de confiança e margem de erro;
- V - sistema interno de controle e verificação, conferência e fiscalização da coleta de dados e do trabalho de campo;

---

<sup>1</sup>Texto retirado de [http://www.planalto.gov.br/ccivil\\_03/leis/19504.htm](http://www.planalto.gov.br/ccivil_03/leis/19504.htm)

VI - questionário completo aplicado ou a ser aplicado;

VII - o nome de quem pagou pela realização do trabalho.

§ 1º As informações relativas às pesquisas serão registradas nos órgãos da Justiça Eleitoral aos quais compete fazer o registro dos candidatos.

§ 2º A Justiça Eleitoral afixará imediatamente, no local de costume, aviso comunicando o registro das informações a que se refere este artigo, colocando-as à disposição dos partidos ou coligações com candidatos ao pleito, os quais a elas terão livre acesso pelo prazo de trinta dias.

§ 3º A divulgação de pesquisa sem o prévio registro das informações de que trata este artigo sujeita os responsáveis a multa no valor de cinquenta mil a cem mil UFIR.

§ 4º A divulgação de pesquisa fraudulenta constitui crime, punível com detenção de seis meses a um ano e multa no valor de cinquenta mil a cem mil UFIR.

Art. 34. (VETADO)

§ 1º Mediante requerimento à Justiça Eleitoral, os partidos poderão ter acesso ao sistema interno de controle, verificação e fiscalização da coleta de dados das entidades que divulgaram pesquisas de opinião relativas às eleições, incluídos os referentes à identificação dos entrevistadores e, por meio de escolha livre e aleatória de planilhas individuais, mapas ou equivalentes, confrontar e conferir os dados publicados, preservada a identidade dos respondentes.

§ 2º O não-cumprimento do disposto neste artigo ou qualquer ato que vise a retardar, impedir ou dificultar a ação fiscalizadora dos partidos constitui crime, punível com detenção, de seis meses a um ano, com a alternativa de prestação de serviços à comunidade pelo mesmo prazo, e multa no valor de dez mil a vinte mil UFIR.

§ 3º A comprovação de irregularidade nos dados publicados sujeita os responsáveis às penas mencionadas no parágrafo anterior, sem prejuízo da obrigatoriedade da veiculação dos dados corretos no mesmo espaço, local, horário, página, caracteres e outros elementos de destaque, de acordo com o veículo usado.

Art. 35. Pelos crimes definidos nos arts. 33, § 4º e 34, §§ 2º e 3º, podem ser responsabilizados penalmente os representantes legais da empresa ou entidade de pesquisa e do órgão veiculador.

Brasília, 30 de setembro de 1997; 176º da Independência e 109º da República.

MARCO ANTONIO DE OLIVEIRA MACIEL

*Iris Rezende*

Este texto não substitui o publicado no D.O.U. de 1º.10.1997

## Referências Bibliográficas

- D. Adler and D. Murdoch. rgl: 3D visualization device system (OpenGL), 2009. URL <http://CRAN.R-project.org/package=rgl>. R package version 0.87. 31
- J.M. Bernardo and A.F.M. Smith. Bayesian Theory. Wiley, New York, 1994. 5
- H. Bolfarine and W.O. Bussab. Elementos de Amostragem. Edgard Blucher, São Paulo, first edition, 2005. 1, 17, 18
- W.G. Cochran. Sampling technique. 1977. 1, 17
- B. de Finetti. Theory of Probability. John Wiley & Sons, Inc., Chichester, first edition, 1974. 4
- M.H. DeGroot. Optimal statistical decisions. McGraw-Hill, New York, 1970. 10, 14
- M.H. DeGroot and M.J. Schervish. Probability and Statistics. Addison Wesley, Boston, MA, Pennsylvania, 2002. 7, 8
- L. Devroye and L. Devroye. Non-uniform random variate generation. Citeseer, 1986. 14
- W. Feller. An Introduction to Probability Theory and Its Applications. John Wiley & Sons, Inc., New York, third edition, 1968. 10, 18, 24, 25, 28
- D. Gamerman and H.F. Lopes. Markov chain Monte Carlo: stochastic simulation for Bayesian inference. Chapman & Hall/CRC, 2006. 14
- A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton, second edition, 2004. 5, 7, 9, 11, 12, 13, 14
- Alan Genz, Frank Bretz, and R port by Torsten Hothorn. mvtnorm: Multivariate Normal and T Distribution, 2009. R package version 0.9-8. 31
- E. Hewitt and L.J. Savage. Symmetric measures on Cartesian products. Transactions of the American Mathematical Society, 80(2):470–501, 1955. 5
- T.Z. Irony and C.A.B. Pereira. Motivation for the use of discrete distributions in quality assurance. Test, 3(2): 181–193, 1994. 25
- N.L. Johnson, S. Kotz, and N. Balakrishnan. Discrete Multivariate Distributions, 1997. 28

- L. Lamport. L<sup>A</sup>T<sub>E</sub>X: A document preparation system / User's guide and reference manual. Addison-Wesley, Massachusetts, second edition, 1994.
- D.V. Lindley. The 1988 Wald memorial lectures: the present position in Bayesian statistics. Statistical Science, 5(1):44–65, 1990. 9
- D.V. Lindley and L.D. Phillips. Inference for a Bernoulli process (a Bayesian view). The American Statistician, 30(3):112–119, 1976. 5, 6
- D. Murdoch, E.D. Chow, and J.M.F. Celayeta. ellipse: Functions for drawing ellipses and ellipse-like confidence regions, 2007. R package version 0.3-5. 31
- M.C. Nunes. Guia de Leitura de Pesquisa Eleitoral - 8. Empate Técnico. IBOPE Inteligência, 2004. URL <http://www.ibope.com.br/calandraWeb/servlet/CalandraRedirect?temp=5&proj=PortalIBOPE&pub=T&db=caldb&comp=Biblioteca&docid=75EC1EDA58720E7D83256EB5006D05AE>. 19
- C.D. Paulino, M.A.A. Turkman, and B. Murteira. Estatística Bayesiana. Fundação Calouste Gulbenkian, Lisboa, 2003. 3, 4, 5, 9, 10, 11, 12, 15
- R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2009. URL <http://www.R-project.org>. ISBN 3-900051-07-0. 31
- A.C. Rencher. Methods of multivariate analysis. John Wiley & Sons, Inc., New York, second edition, 2002. 28
- B.D. Ripley. Stochastic simulation. Wiley New York et al., 1987. 14
- C. Weihs, U. Ligges, K. Luebke, and N. Raabe. klar analyzing german business cycles. In D. Baier, editor, Data Analysis and Decision Support, pages 335–343, Berlin, 2005. Springer-Verlag. 31
- T.W. Yee. VGAM: Vector Generalized Linear and Additive Models, 2009. URL <http://CRAN.R-project.org/package=VGAM>. R package version 0.7-9. 31